

# Patients' online descriptions of their experiences as a measure of healthcare quality

Submitted for the Degree of Doctor of Philosophy

# Felix Edward Clovis Greaves

# October 2013

Imperial College London



# **Abstract**

## **Introduction**

Patients are describing their healthcare experiences online using rating websites. There has been substantial professional opposition to this, but the government in England has promoted the idea as a mechanism to improve healthcare quality. Little is known about the content and effect of healthcare rating and review sites. This thesis aims to look at comments left online and assess whether they might be a useful measure of healthcare quality.

## **Method**

I used a variety of different approaches to examine patients' comments and ratings about care online. I performed an examination of the comments left on the NHS Choices website, and analysed whether there was a relationship between the comments and traditional patient surveys or other measures of clinical quality. I used discrete choice experiments to look at the value patients place on online care reviews when making decisions about which hospital to go to. I used natural language processing techniques to explore the comments left in free text reviews. I analysed the tweets sent to NHS hospitals in England over a year to see if they contained useful information for understanding care quality.

## **Results**

The analysis of ratings on NHS Choices demonstrates that reviews left online are largely positive. There are associations between online ratings and both traditional survey methods of patient experience and outcome measures. There is evidence of a selection bias in those

who both read and contribute ratings online – with younger age groups and those with higher educational attainment more likely to use them. Discrete choice experiments suggest that people will use online ratings in their decisions about where to seek care, and the effect is similar to that of a recommendation by friends and family. I found that sentiment analysis techniques can be used to classify free text comments left online into meaningful information that relates to data in the national patient surveys. However, the analysis of comments on Twitter found that only 11% of tweets were related to care quality.

## **Conclusions**

Patients rating their care online may have a useful role as a measure of care quality. It has some drawbacks, not least the non-random group of people who leave their comments. However, it provides information that is complementary to current approaches to measuring quality and patient experiences, may be used by patients in their decision-making, and provides timely information for quality improvement. I hypothesise that it is possible to measure a 'cloud of patient experience' from all of the sources where patients describe their care online, including social media, and use this to make inferences about care quality. I find this idea has potential, but there are many technical and practical limitations that need to be overcome before it is useful.

## CONTENTS

|                                                                                                                                                |            |
|------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>ABSTRACT</b>                                                                                                                                | <b>1</b>   |
| <b>LIST OF TABLES:</b>                                                                                                                         | <b>5</b>   |
| <b>LIST OF FIGURES:</b>                                                                                                                        | <b>6</b>   |
| <b>ACKNOWLEDGEMENTS</b>                                                                                                                        | <b>7</b>   |
| <b>STATEMENT OF CONTRIBUTION</b>                                                                                                               | <b>8</b>   |
| <b>LIST OF ABBREVIATIONS</b>                                                                                                                   | <b>10</b>  |
| <b>CHAPTER 1: DRIVERS OF PATIENTS DESCRIBING THEIR CARE ONLINE:</b>                                                                            | <b>11</b>  |
| 1.1 Introduction                                                                                                                               | 11         |
| 1.2. Driver 1: Public reporting of healthcare provider performance                                                                             | 15         |
| 1.3 Driver 2: The rise of patient experience as a measure of healthcare quality                                                                | 40         |
| 1.4 Driver 3: Societal Trends and the Internet                                                                                                 | 56         |
| 1.5 Conclusion: The evolution of an idea - patients describing their care online                                                               | 62         |
| <b>CHAPTER 2: INTERNET BASED HEALTHCARE FEEDBACK</b>                                                                                           | <b>64</b>  |
| 2.1 Introduction to ratings websites                                                                                                           | 64         |
| 2.2 The literature around unsolicited patient comment online                                                                                   | 65         |
| 2.3 Usage of Internet patient reporting websites                                                                                               | 75         |
| 2.4 Learning from other industries                                                                                                             | 82         |
| 2.5 Gaps in the literature                                                                                                                     | 85         |
| 2.6 Online feedback in healthcare in the UK                                                                                                    | 88         |
| 2.7 An opportunity to study in detail: NHS Choices                                                                                             | 93         |
| 2.8 A 'Cloud of Patient Experience': Is it possible to collect data from other sources?                                                        | 100        |
| 2.9 Summary                                                                                                                                    | 104        |
| 2.10 Aims and Objectives                                                                                                                       | 106        |
| <b>CHAPTER 3: USAGE OF ONLINE COMMENTS AND REVIEWS ON NHS CHOICES AND COMPARISON WITH OTHER METRICS OF PATIENT EXPERIENCE AND CARE QUALITY</b> | <b>108</b> |
| 3.1 Background                                                                                                                                 | 108        |
| 3.2 Experiment 1: Online feedback for acute hospitals                                                                                          | 109        |
| Methods                                                                                                                                        | 109        |
| 3.3 Results                                                                                                                                    | 112        |
| 3.4 Experiment 2: Online feedback for primary care                                                                                             | 118        |
| Method                                                                                                                                         | 118        |
| 3.5 Results                                                                                                                                    | 123        |
| 3.6 Discussion                                                                                                                                 | 130        |
| <b>CHAPTER 4: THE VALUE OF ONLINE RATINGS AND REVIEWS IN HOSPITAL CHOICE: A DISCRETE CHOICE EXPERIMENT</b>                                     | <b>140</b> |
| 4.1 Background                                                                                                                                 | 140        |
| 4.2 Method                                                                                                                                     | 141        |
| 4.3 Results                                                                                                                                    | 149        |
| 4.4 Discussion                                                                                                                                 | 160        |
| <b>CHAPTER 5: USE OF SENTIMENT ANALYSIS FOR CAPTURING PATIENT EXPERIENCE FROM FREE-TEXT COMMENTS POSTED ONLINE</b>                             | <b>170</b> |
| 5.1 Background                                                                                                                                 | 170        |
| 5.2 Methods                                                                                                                                    | 172        |
| 5.3 Results                                                                                                                                    | 176        |
| 5.4 Conclusions                                                                                                                                | 180        |
| <b>CHAPTER 6: ANALYSIS OF PATIENTS' COMMENTS ABOUT THE ENGLISH NHS VIA TWITTER, AND COMPARISON WITH PATIENT SURVEYS.</b>                       | <b>184</b> |
| 6.1 Background                                                                                                                                 | 184        |
| 6.2 Method                                                                                                                                     | 186        |
| 6.3 Results                                                                                                                                    | 189        |
| 6.4 Discussion                                                                                                                                 | 205        |

|                                                                                     |            |
|-------------------------------------------------------------------------------------|------------|
| 6.4.5 Conclusion                                                                    | 210        |
| <b>CHAPTER 7: DISCUSSION</b>                                                        | <b>212</b> |
| 7.1 Key findings of the thesis                                                      | 212        |
| 7.2 Future policy and practice: Online ratings and reviews                          | 217        |
| 7.3 Drawing all this information together to create a 'Cloud of Patient Experience' | 226        |
| 7.4 Future research and questions to answer                                         | 230        |
| 7.5 New contributions from this thesis                                              | 239        |
| 7.6 Conclusions                                                                     | 241        |
| <b>REFERENCES:</b>                                                                  | <b>245</b> |
| <b>APPENDIX 1: OUTPUTS FROM PHD</b>                                                 | <b>267</b> |
| <b>APPENDIX 2: DETAILS OF REGRESSION ANALYSIS IN TABLE 14</b>                       | <b>270</b> |
| <b>APPENDIX 3: DETAILS OF REGRESSION ANALYSIS IN TABLE 15</b>                       | <b>272</b> |
| <b>APPENDIX 4: EXAMPLE SURVEY FOR DISCRETE CHOICE EXPERIMENT</b>                    | <b>274</b> |

## List of Tables:

|                                                                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 1: Background to provider choice in the English NHS.....                                                                                                                                      | 27  |
| Table 2: Dimensions of healthcare effectiveness .....                                                                                                                                               | 43  |
| Table 3: Different attempts to incorporate patient experience in to healthcare quality framework.....                                                                                               | 44  |
| Table 4: Proportion of reviews positive .....                                                                                                                                                       | 80  |
| Table 5: Questions asked by NHS Choices .....                                                                                                                                                       | 95  |
| Table 6: Total comments/rating from July 2007 to March 2013 on NHS Choices .....                                                                                                                    | 96  |
| Table 7: Patient rating domains on NHS Choices and scores for 2009/10 .....                                                                                                                         | 113 |
| Table 8: Hospital trust level associations between web-based patient ratings and clinical outcomes .....                                                                                            | 115 |
| Table 9: Hospital trust level associations between web-based patient ratings and healthcare-acquired infection (HCAI) rates.....                                                                    | 115 |
| Table 10: Hospital trust level associations between web-based patient ratings and paper-based surveys of patient experience .....                                                                   | 116 |
| Table 11: Comparison of associations between web-based/paper based ratings and clinical outcomes .....                                                                                              | 117 |
| Table 12: Mean patient ratings posted for questions on NHS Choices .....                                                                                                                            | 124 |
| Table 13: Descriptive statistics of practice characteristics .....                                                                                                                                  | 124 |
| Table 14: Associations between whether a practice is rated with population and practice characteristics .....                                                                                       | 125 |
| Table 15: Associations between the proportion of patients recommending a practice with population and practice characteristics .....                                                                | 126 |
| Table 16: Associations between web-based patient ratings and conventional survey measures of patient experience .....                                                                               | 127 |
| Table 17: Associations between web-based patient ratings of whether the user would recommend the GP practice to a friend and various clinical quality indicators .....                              | 128 |
| Table 18: Comparison of the characteristics of the NHS Inpatient Survey and ratings on the NHS Choices website.....                                                                                 | 132 |
| Table 19: Attributes included in the discrete choice experiment .....                                                                                                                               | 145 |
| Table 20: Characteristics of the sample.....                                                                                                                                                        | 150 |
| Table 21: Survey results on attitudes towards choice and online health behaviours, including breakdown by demographic and socio-economic subgroups (unadjusted) .....                               | 151 |
| Table 22: Adjusted analysis of survey results .....                                                                                                                                                 | 152 |
| Table 23: Estimates of utility for various attributes in hospital choice, including comparison in terms of waiting time equivalent (simple model not incorporating individual characteristics)..... | 156 |
| Table 24: Estimates of utility for various attributes in hospital choice .....                                                                                                                      | 158 |
| Table 25: Accuracy of different approaches to machine learning.....                                                                                                                                 | 177 |
| Table 26: The 10 words or two word phases with the most predictive accuracy.....                                                                                                                    | 178 |
| Table 27: Comparison of patient survey responses and machine learning predictions .....                                                                                                             | 179 |
| Table 28: Characteristics of hospital trusts on Twitter.....                                                                                                                                        | 189 |
| Table 29: Codebook of tweet classification including frequency of occurrence.....                                                                                                                   | 194 |
| Table 30: Examples of tweets in each of the categories .....                                                                                                                                        | 195 |
| Table 31: New information about advantages and disadvantages of online feedback .....                                                                                                               | 221 |
| Table 32: Potential sources of information for the 'Cloud of Patient Experience' .....                                                                                                              | 229 |

## List of Figures:

Illustration on title page: A cloud of words used in all negative comments about hospitals on the NHS Choices website 2008 -2012

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1: The drivers leading to online feedback in healthcare .....                                                                       | 12  |
| Figure 2: Pathways by which public reporting of performance data can lead to increased performance .....                                   | 17  |
| Figure 3: Public satisfaction with the NHS over time .....                                                                                 | 55  |
| Figure 4: Availability of the Internet in the UK .....                                                                                     | 56  |
| Figure 5: Internet use by age group, England, 2012 .....                                                                                   | 57  |
| Figure 6: Internet use by weekly income, England, 2012 .....                                                                               | 57  |
| Figure 7: A typical page from Yelp, a popular website for health reviews .....                                                             | 64  |
| Figure 8: Cumulative number of ratings on the RateMDs website in the US .....                                                              | 76  |
| Figure 9: iWantGreatCare (designed to spur selection) .....                                                                                | 92  |
| Figure 10: Patient Opinion (designed to spur improvement activity) .....                                                                   | 92  |
| Figure 11: Provider view incorporating latest comments, NHS Choices .....                                                                  | 98  |
| Figure 12: High level comparison view, NHS Choices .....                                                                                   | 98  |
| Figure 13: Histogram of frequency of web-based ratings per hospital trust .....                                                            | 113 |
| Figure 14: Scatter plot of number of ratings and proportion recommending hospital to a friend .....                                        | 114 |
| Figure 15: Histogram of rating frequency for all GP practices .....                                                                        | 123 |
| Figure 16: Map of frequency of primary care ratings on the NHS Choices website .....                                                       | 129 |
| Figure 17: Increasing number of ratings in the US on RateMDs - (NB figure is cumulative) .....                                             | 134 |
| Figure 18: Increasing number of hospital ratings in England on NHS Choices – (NB figure is cumulative) .....                               | 135 |
| Figure 19: An example of a discrete choice question .....                                                                                  | 146 |
| Figure 20: Information people would use to choose a hospital .....                                                                         | 153 |
| Figure 21: Information people would use to book a hotel .....                                                                              | 153 |
| Figure 22: Relationship between utility and waiting time demonstrating a linear function .....                                             | 155 |
| Figure 23: Relationship between utility and travel time demonstrating a linear function .....                                              | 155 |
| Figure 24: Effect of various hospital attributes on hospital choice, expressed as ‘weeks willing to wait’ .....                            | 157 |
| Figure 25: Effect of individual characteristics on choice behaviour .....                                                                  | 159 |
| Figure 26: Comparison of the proportion recommending a hospital using sentiment analysis and traditional paper based survey measures ..... | 179 |
| Figure 27: Number of tweets per trust .....                                                                                                | 190 |
| Figure 28: Frequency of tweets by day over the period studied .....                                                                        | 191 |
| Figure 29: Frequency of tweets about NHS hospitals by hour of day .....                                                                    | 193 |
| Figure 30: Example of a tweeted picture .....                                                                                              | 204 |
| Figure 31: Trust level association between Twitter sentiment and the NHS patient survey .....                                              | 205 |

## Acknowledgements

I am sincerely grateful to Dr Christopher Millett and Prof Ara Darzi for providing excellent supervision and advice in writing this thesis. I am grateful to Prof Azeem Majeed for his supporting the work. I thank both the Department of Primary Care and Public Health and the Centre for Health Policy at Imperial College for providing supportive environments to work in. I also thank John Robinson, Paul Nuki and the team at NHS Choices for providing me with access to their data.

I would like to thank Utz Pape, Anthony Laverty, Peter Burge and Daniel Ramirez Cano for providing statistical and technical support for various aspects of the work. I also thank Liam Donaldson and Bob Wachter for help clarifying some of my thinking around the ideas in this thesis.

I would like to acknowledge the support of the London Public Health Training Programme and the NIHR Collaboration for Leadership in Applied Health Research for North London for funding my salary during my studies, and the Department of Health for providing funding for the experiments in Chapter 4.

## **Statement of contribution**

The work presented in this thesis is my own but I received direction from my supervisors, Christopher Millett and Ara Darzi. In addition, Utz Pape advised me on the statistical analysis for Chapter 3. Peter Burge supported the analysis of discrete choice experiments in Chapter 4, both advising on design of the experiment, and developing the explanatory models using Alogit software. Daniel Ramirez Cano assisted the work using natural language processing in Chapter 5. We worked together closely to design the approach, but he was responsible for the implementation of the Natural Language processing software. He also provided code to collect tweets for Chapter 6. Anthony Lavery contributed to the qualitative analysis of tweets in Chapter 6, and also advised on several of the statistical techniques used in other chapters. Liam Donaldson, Azeem Majeed and Bob Wachter all provided additional guidance.

## **Copyright**

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

## Note on work published elsewhere

Parts of this thesis have been previously published in five peer-reviewed articles:

- **Greaves F**, Pape UJ, King D, Darzi A, Majeed A, Wachter RM, & Millett C. (2012). Associations between Web-Based Patient Ratings and Objective Measures of Hospital Quality. *Archives of Internal Medicine*, 172(5), 435–436.

Data from the article in *Archives of Internal Medicine* is reproduced with their permission. Copyright is held by the American Medical Association.

- **Greaves F**, Pape U J, King D, Darzi A, Majeed A, Wachter RM, & Millett C. (2012). Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ quality & safety*, 21(7), 600–5.
- **Greaves F**, Ramirez-Cano D, Millett C, Darzi A, & Donaldson L. (2013). Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. *BMJ quality & safety*, 22(3), 251–5.

Parts of the articles in *BMJ Quality and Safety* are reproduced with their permission. Copyright is held by BMJ Publishing Group.

- **Greaves F**, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, & Millett C. (2012). Patients' Ratings of Family Physician Practices on the Internet: Usage and Associations with Conventional Measures of Quality in the English National Health Service. *Journal of Medical Internet Research*, 14(5), e146.
- **Greaves F**, Millett C. Consistently increasing numbers of online ratings of healthcare in England. *J Med Internet Res*. 2012 Jun 29;14(3):e94. [Letter] ,

Articles in the *Journal of Medical Internet Research* have been published under a Creative Commons Attribution License.

Figure 2 is reproduced with permission of the American College of Physicians.

Table 3 is reproduced with permission of the Oxford University Press.

Figure 16 is reproduced under a Creative Commons Attribution License.

## List of abbreviations

|        |                                                                  |
|--------|------------------------------------------------------------------|
| AHRQ   | Agency for Healthcare Research & Quality (US)                    |
| API    | Application programming interface                                |
| BP     | Blood pressure                                                   |
| BME    | Black and minority ethnic                                        |
| DCE    | Discrete choice experiment                                       |
| GP     | General practitioner                                             |
| HBA1C  | Glycated haemoglobin                                             |
| HCAHPS | Hospital Consumer Assessment of Healthcare Providers and Systems |
| HCAI   | Health Care Associated Infection                                 |
| HES    | Hospital Episode Statistics                                      |
| HSCIC  | Health and Social Care Information Centre                        |
| HSMR   | Hospital Standardised Mortality Ratio                            |
| IOM    | Institute of Medicine                                            |
| MRSA   | Meticillin resistant <i>Staphylococcus aureus</i>                |
| MT     | Modified tweet                                                   |
| NBM    | Naive Bayes Multinomial                                          |
| NHS    | National Health Service                                          |
| NHSC   | NHS Choices                                                      |
| NLP    | Natural Language Processing                                      |
| OECD   | Organisation for Economic Co-operation and Development           |
| QOF    | Quality and Outcomes Framework                                   |
| RT     | Retweet                                                          |
| SES    | Socio Economic Status                                            |
| SHMI   | Summary Hospital-level Mortality Indicator                       |
| WHO    | World Health Organization                                        |

# **Chapter 1: Drivers of patients describing their care online: Transparency, choice, patient experience and the rise of the Internet**

*“I believe government has been much too slow to make use of the enormous democratising power of information. People take it for granted that they will access other people’s reviews and ratings before buying something on eBay or Amazon, and yet we do not yet have systematic access to other people’s experiences when choosing a GP practice.”*

*Gordon Brown, forward to Public Services: On your side (1)*

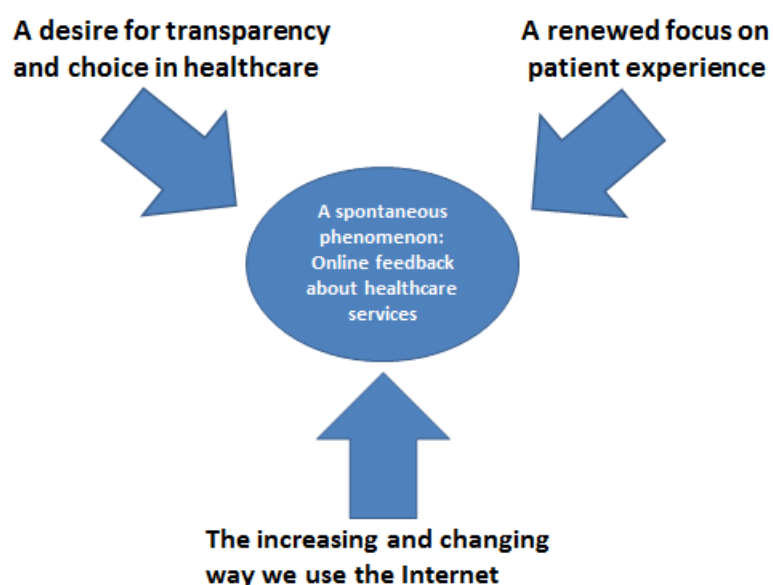
## **1.1 Introduction**

Consumers are using the Internet to rate services and products, for example, when they stay in a hotel or buy a product online. And people use this information to make decisions about what to buy and use. This increasingly applies to health care too. A number of websites allowing patients to rate their care have been developed by healthcare payers and the commercial sector, such as Patient Opinion in the UK and RateMDs in the US, in an effort to increase transparency and responsiveness of health systems and to help patients choose between providers (2,3). Although controversial, particularly among health care professionals (4), web-based hospital or physician rating sites represent a potentially important development in public reporting (5). Patients rarely use conventional publicly reported metrics when choosing their healthcare provider and often find these difficult to understand (6,7). They may prefer to use other patients’ ratings and experiences, particularly as more individuals instinctively turn to online customer reviews and peer opinions generated through online sites like Amazon, TripAdvisor and Facebook in making

many purchasing decisions. Rhetoric from the government about the worth of online feedback is strong and plentiful. However academic analyses to confirm, or refute, the arguments for this new phenomenon are weak and in many cases missing.

This thesis explores the topic of patients describing online their experiences of receiving care in an effort to find out how useful it might be. Online feedback about healthcare has appeared at the confluence of a number of current streams of policy, technology and culture, arising from new behaviours and new expectations about healthcare. In the first chapter, I examine some of these macro level themes and trends, which gave us an idea of the context in which this new phenomenon has emerged, and which may give some clues as to its potential value and risks. I have conceptualised these as three key drivers: 1) a desire for transparency and choice in healthcare, 2) a renewed focus on patient experience in healthcare and 3) the increasing and changing the way we are using the Internet. I take each in turn and explain how they are having an effect

**Figure 1: The drivers leading to online feedback in healthcare**



In the second chapter, I look at what is known about websites on which patients can rate either physicians or hospitals. I examine the limited existing literature from around the world to look for arguments in their favour, and against, and to understand how they are used. I also consider what the gaps in our current knowledge are, and set out my hypothesis that patients' descriptions of their care online might have a role in the landscape of clinical quality measures.

The third to sixth chapters are empirical research, each adopting a different methodological approach to looking at an aspect of the concept of online care description.

For the third chapter, I have obtained access to the data collected by the national, state-funded rating website in England, NHS Choices – possibly the largest such site in the world – and have performed the first national level study of activity on a website of this sort for both hospitals and primary care doctors. I describe the ratings in terms of how much of the health system they cover, and how positive they are. I compare the ratings with other conventional quality measures – such as survey measures of patient experience and mortality rates and I explore some of the links between population characteristics and usage rates.

In the fourth chapter, I use an economic experimental approach- a discrete choice experiment- to build a mathematical model of how people use online care ratings and reviews to make decision about which hospital to go to, in comparison to the many other

attributes that a hospital has. I also conduct a survey of knowledge of and attitudes towards online rating – the first in England.

In the fifth chapter, I adopt new analytical techniques of sentiment analysis from computer science and the marketing industries to try to make sense of free text information captured with online reviews. I demonstrate, using the NHS Choices dataset, that this is possible, and that the results are comparable to national patient surveys.

In the sixth chapter, I look at another source of online commentary about care online, the micro-blogging website Twitter. In an analysis of a years' worth of tweets about hospitals in the NHS, I qualitatively analyse what is said in them and explore their use as a measure of care quality.

As each empirical chapter is different in approach, I have included a discussion at the end of each, which I have then tried to bring together in a final chapter, where I take each of the findings and see how they combine together, what we have learnt, how this might affect policy, what we still need to know and where to go next with this work. This is an area on the policy agenda, but until recently neglected by academic scrutiny – it is ripe for examination and this analysis is carried out in the hope that it will provide useful and practical evidence for policymakers.

## **1.2. Driver 1: Public reporting of healthcare provider performance and the drive for greater transparency in health system**

### **1.2.1 The nature and purpose of public reporting**

Recent years have seen an increased movement towards the transparent display of information about the performance of health care providers – this has commonly been described in the policy literature as ‘public reporting’. The precise definitions used have varied slightly. The AHRQ proposed public reporting as “data, publicly available or available to a broad audience free of charge or at a nominal cost, about a health care structure, process or outcome at any provider level” (8). Cacace considers it “performance-related information about non-anonymous providers disclosed to the general public by applying a comparative approach” (9). In the policy discussion, it is often characterised by the language of ‘reports cards’ and ‘consumer reports’ (10). It has also been usefully defined by what it is not. These include: data which requires payment or position to access it, anonymous reporting (for example unlabelled organisational data), or data aggregated to regional or national level, or unsystematic feedback (7–9).

Along with these definitions, its purpose has been described variously. These include as a tool to promote transparency and provide information to allow informed choice, a device to stimulate quality improvement and a mechanism by which provide can be held accountable for the care that they deliver (7,10,11). A more philosophical viewpoint has noted the value of transparency for transparency's sake – suggesting that it is fundamentally a beneficial act, regardless of how it whether or how it produces an effect (12).

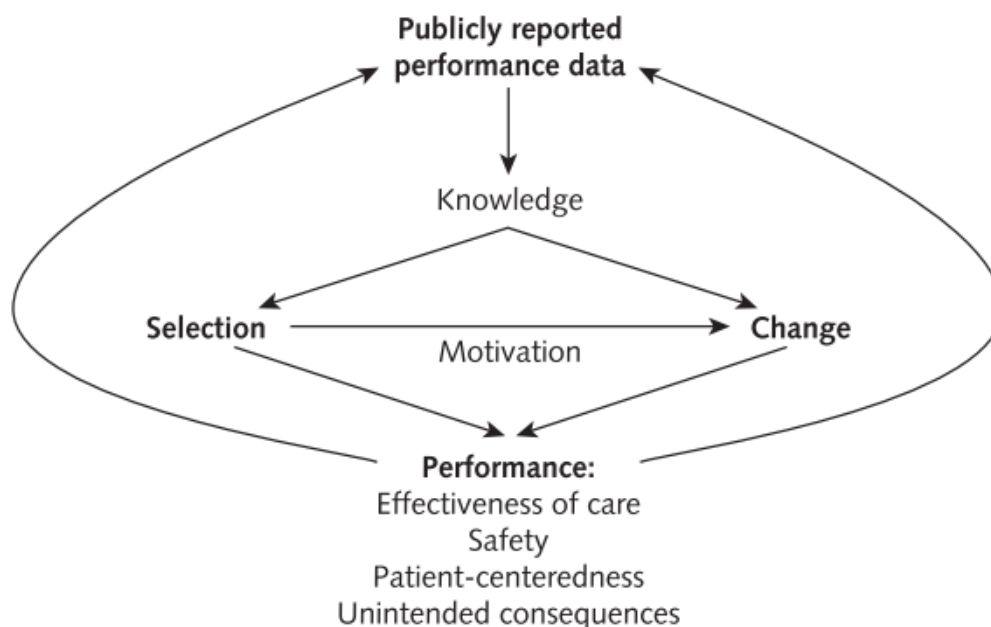
### 1.2.2 Mechanisms to improve quality

One of the most often stated expectations of public reporting is that it will improve the quality of care provided. It is known that there are both substantial variations in performance between clinical providers, and that average performance of healthcare providers is sub-optimal with much care not provided to the level recommended by clinical guidelines. Early seminal studies around hospital quality, notably the RAND Quality Study, demonstrated that patients received only 55% of the recommended steps for 30 common acute and chronic conditions in US hospitals (13). In 1999, the Institute of Medicine (IOM) report *To Err is Human* demonstrated that there are systematic problems with the way safety is managed in the US health system, and as many as 98,000 people may die a year as a result (14). Two years later their report *Crossing the Quality Chasm* argued that transparency of health system performance would lead to better quality (15). Although early work was done in the US, similar variations have been found in the UK. The NHS Atlas of Variation, published in 2011, showed substantial variation in activity rates and care quality (measured with process and outcome indicators) across England in a variety of common conditions including diabetes, depression, trauma and cardiac care (16).

The potential mechanisms by which transparency would lead to improved quality were best explained by Berwick and colleagues (17), where they set out two pathways by which the measurement and reporting of performance data within health systems could lead to the system improving. The first mechanism (which they termed improvement through ‘*selection*’) acts by altering the flows of patients (and their attached funding) as they make informed choice about the best place to seek care, using the information that is available publically. In the second pathway (described as improvement through ‘*changes in care*’) the

intrinsic pressure of competition on these organisations, when faced with transparent information comparing them to other providers, will drive those working in provider organisations to improve themselves. This theoretical model was summarised in diagrammatic form by Fung and colleagues in 2008 (see Figure 2). This demonstrates that the two pathways are not distinct, but linked. In particular, the threat of selection acts as the pressure to motivate the change to occur.

**Figure 2: Pathways by which public reporting of performance data can lead to increased performance**



Source: Fung et al. (7) – itself adapted from Berwick et al. (17)

The Berwick model has been expanded to include other quality improvement mechanisms – with the most substantial contribution from Hibbard who thought the original two-pronged pathway too limited. She suggested another pathway: concern for public image or reputation (18). This theory is based on her observation that making data public, as opposed to the same data being available within the organisation, leads to more active quality

improvement activity and improved performance. It is not the management being aware of the comparative data that has the effect; it is that the data is in the open. Her evidence to back up this assertion is strong, with both short and longer term controlled experimental studies carried out following the introduction of the Wisconsin QualityCounts initiative. Hospitals who received a public report were more likely to pursue quality improvement activity than both hospitals where the report was not public or where no report was made (19) (20). This leaves three established theoretical pathways of effect of public reporting to data: 1) selection, 2) internal change, and 3) reputational damage.

### ***Necessary assumptions***

All of these mechanisms rely on a set of assumptions. The 'change' and 'reputation' pathways requires those people working within their organisation to be aware of their performance, to be actively looking at how they compare to other similar services, and to be motivated and resourced to take part in improvement activity (17,21).

In contrast, for the selection pathway to improve quality, patients need to act as logical agents within a market: voting with their feet to move between providers, selecting good performers and discarding bad ones on the basis of published information. This requires acceptance of a neoclassical theory of economics in how a market in healthcare would behave, a marketplace of healthcare providers and a public that is able to choose between them (22). Within this theory, there is also a set of further conditions, summarized by Weintraub (23) as: 1) people have rational preferences among outcomes (they act logically), 2) individuals maximize utility (they know what they want and try to achieve it), and 3)

people act independently on the basis of full and relevant information. Further, this also assumes that performance is measurable and that all dimensions of effort are observable.

As with almost any market, these are idealised conditions that rarely, if ever, occur. In addition, healthcare has a number of particular nuances that might act to prevent this, particularly around patient access to information and rationality, potentially leading to market failure (24). Whether public reporting has an effect on quality, by any of these mechanisms, is still debated. Measuring the effect of public reporting has been attempted under a number of policy initiatives, across a number of different health systems, and we are still not certain whether it has a beneficial effect on outcomes.

### **1.2.3 The policy history of public reporting**

Public reporting has now become a mainstream health policy at the national level in a number of countries. In some cases, it has a voluntary activity (most notably in the Scandinavian countries); in others it is required by statute (in England, or in certain states in the US). For the purpose of this work, I will focus on the history of their development in the US – where the ideas were pioneered - and in the UK, which is relevant to the content and data I have studied in this thesis.

Although the seeds of the idea were planted more than a century ago by early pioneers, public reporting emerged in modern times in the 1990s. Florence Nightingale (1820 -1910), for example, advocated benchmarking between hospital in England to improve their performance in her monograph 'Notes on Hospitals' (1859) (25). Sixty years later, Ernest Codman (1869 – 1940) in the US was ostracised by the Massachusetts medical community

for suggesting that hospitals should measure and report their outcomes - or the 'product of a hospital' as he termed it (26).

Most modern considerations of this topic start with the ground-breaking work done by Mark Chassin and colleagues in New York State, publishing the performance of cardiac surgeons (27), as awareness began to grow about the huge variations in the quality of care that existed. The New York State Cardiac Surgery Reporting Program began the early 90s, and looked at the data for specific cardiac procedures. They found huge variations in mortality across 31 hospitals for heart bypass surgery, even when they accounted for the case mix of the patients (0.5% to 7.3%) - a fifteen fold difference, when they first looked systematically in 1990. And in a brave move, they made their entire finding public, including the performance and mortality rates of individual surgeons.

Even in a pre-Internet era, this public disclosure seemed to have an effect. The state-wide risk adjusted CABG mortality rate had dropped from 4.2% in 1989 to 2.5% in 2003, a fall of 41% (and faster than nationally) (28), and variation between providers was reduced (29). What is more, Mukamel and colleagues found that hospitals and doctors with better outcomes for CABG surgeries had higher rates of growth in market share – a 1% higher mortality rate led to a 1.8% reduction (for hospitals) and a 7% reduction for surgeons in market share growth rate (30). They even saw an impact on surgeons, with far more surgeons in the lower echelons of performance leaving practice. 19% of the bottom quartile left in a 2 year period, compared to 3% in the top quartile, after adjustment for surgeon characteristics (31). Transparency and public reporting seemed to work.

On the basis of this important foundation, numerous others organisation and areas followed suit with public disclosure of information. This varied by area, with some states requiring public disclosure of information, and by commercial organisation, with some groups of providers and health plans coming together and making their information public. Public reporting has been undertaken in several different settings, including hospitals (19,20), nursing homes (32,33) and care plans (34) and physician groups (35). Public reporting systems have developed in the UK. Scotland was the first to adopt the approach, publishing hospital mortality data as early as 1994 (36), but this has now spread to England and the other devolved nations (which is discussed in detail later). The free market nature of healthcare in the US made it a particularly important testing ground. Most recently, the Affordable Care Act (ACA) has increased the statutory requirement for public reporting. The ACA requires health plans offered through new Health Benefit Exchanges to report quality measures, both for consumers and to drive improvement activities from 2015 (37).

#### **1.2.4 Synthesis of data on the effectiveness of public reporting**

The variety of attempts to describe and measure the effects of these various new different public reporting systems have been captured in a series of systematic reviews that highlight the developing evidence. Marshall and colleagues in an early review in 2000 found 21 studies that looked for empirical evidence, based on only 7 different public reporting platforms, all of which were in the US (10). Their findings were mixed, and occasionally conflicting. Survey studies showed patients clearly wanted information on performance (38), but the information provided only had a limited effect on them making consumer decisions. Instead patients seemed hesitant to 'vote with their feet' in the face of published data from the Health Care Financing Administration (HCFA) mortality figures (39) (40). This

appeared to be due to a lack of choice in their real decisions, a lack of trust in the data, failure to understand it, and even a lack of interest. They noted that the early reports from the NY state cardiac studies (30) seemed to have exceptional results on this point.

On the other hand, Marshall and colleagues did note that there might be a measureable effect on what hospitals were doing. In particular, Longo's work looking at a maternity hospital reporting system in Missouri demonstrated that hospitals taking part in the public reporting system were more likely to engage in quality improvement activity after data was released, and that this was associated with improvements in patient experience and process measures of care (41). They also found evidence of quality improvement activity that matched public reporting in several other studies (42) (43). That said, they also found reports of a number of organisations that struggle to accept the data, did nothing or questioned the validity of the reporting system (44,45). In conclusion, Marshall suggested evidence on public reporting was scant and it was hard to make a strong comment about effectiveness, but that public reporting did appear to have some effect in hospitals – although the mechanism of improvement was not clear.

Their contemporaries Schuffler and Mordavsky, in a review in 2001, came to a less definite conclusion, claiming they found that consumer report cards do not make a difference in decision-making, improvement of quality, or competition (46). Nihilistically, they suggested that the whole endeavour should be reconsidered.

Fung and colleagues, reviewing the literature in 2008, found a further 27 articles not captured in the earlier review process. They looked in more detail to see the effect not just

on quality improvement activity, but also specifically at different clinical outcomes: effectiveness, safety and patient experience. They agreed with Marshall's earlier finding that "publicly releasing performance data stimulates quality improvement activity at the hospital level". However, they noted that going further and suggesting that there is an effect on any of the three outcomes would not be justified by the evidence in its current state (47). They comment on the lack of quality of many of the studies examined, a continued focus on particular public reporting systems (such as New York State) and an absence of data on some of the newer, but widely used commercial reporting websites, such as Healthgrades. For the first time in these reviews, they also only included one non US study (48) which examined the impact of star ratings in the English NHS. In this work, Mannion and colleagues used a multiple case study approach to compare the views of staff at high and low performing organisations. Although they found some staff considered them useful as a tool to drive improvement, the majority felt they did not represent a rounded view of organisation performance, and they caused a variety of negative unintended consequences including a distortion of clinical priorities, staff intimidation and reduced morale.

Taking their findings together, the authors of the review suggested three areas for further work: 1) more data needs to be collected from different systems, 2) more research was needed of the effects of design and implementation on the reports effect, and 3) there should be a careful look to see which of the causal pathways are responsible for the effect (47). Significantly, they found no substantial evidence of Berwick's 'selection pathway' in action. The AHRQ also reviewed the literature, hoping to identify the mechanism by which any quality improvement process took place (49). They concluded that although there was an observable effect, particularly around process measures of care, there was "no evidence or

only weak evidence that public reporting affects the selection of health care providers by patients or their representatives”.

In the latest published review, Berger and colleagues looked explicitly at the effect of public reporting on outcomes. Again they found a mixed picture. Of the 25 studies they identified that looked at the issue, 6 showed public reporting improved outcomes, 9 showed a mixed effect, 9 showed no measureable change and 1 showed a negative change (50).

Some authors remain critical of public reporting, with Rothberg and colleagues suggesting that the lack of consensus about how to present information and allow comparison means that there is a risk that public reporting systems may do more harm than good. In their own words – “rating services appear likely to confuse, rather than inform, consumers’ (51). A number of authors also describe the need for consideration of the potential adverse consequences. Marshall noted the risk of patient selection (whereby risky patients would be ignored because they might ruin surgeon’s statistics), a detrimental effect on morale, the potential for gaming and an inappropriate focus on those issues that were measured (with the potential neglect of those things that were not) (10). He also noted that public reporting is difficult to accept for many clinicians, representing a substantial challenge to traditional professional ideas of autonomy (10) (11). Looking at the New York State system, Omooigui and colleagues found New York residents treated at the Cleveland clinic had higher expected and observed mortality after public reporting was introduced in NY, suggesting that riskier cases were transferred out of the state to Cleveland (52). Moscucci et al. found higher risk patients were less likely to be treated with PCI in New York than Michigan, and

hypothesized that New York physicians were avoiding high risk patients for fear of lowering the publicly reported mortality figures (53).

More recent work, which has not yet been incorporated in the reviews, continues in the same vein. Joynt showed that in a comparison between states with public reporting and those without, there was no difference in mortality, but the PCI usage was lower in states where it was reported. Possible causes include a risk aversion in states with public reporting, or a more rational selection of patients (54). Work in the UK around the publication of cardiothoracic surgery performance data, appeared to show that publication was associated with falls in mortality for the providers monitored, without any evidence of surgeons showing risk averse behaviours towards the more high risk patients (55).

Taking all of these high-level, mostly observational studies together, we therefore see a pattern emerging. Data suggest that public reporting may have an impact on processes and quality improvement activity, but evidence is less strong on improving outcomes, and substantially lacking that there is an effect on patients choosing their hospitals. It might be that the expectations of those advocating public reporting to drive selection are – at least for the moment - unlikely to be met. Trying to understand how public reporting has an effect, academics have used a number of other different lenses – going from the most macro, system-level approaches looking at competition in the whole system, to studying the actions of the individual people making choices, to further understand these effects. Both have provided useful insights.

### **1.2.5 Choice and competition in the NHS**

A key development in the NHS that has occurred simultaneously with the development of public reporting is the increasing presence of choice of healthcare providers. This is driven by a belief that allowing competition will lead to improvement. Choice has not traditionally been a major feature of the National Health Service. For the first 40 years of its existence, care was normally provided at the closest provider, or the patient was referred further afield for specialist treatment. This has been changing over the last quarter of a century, as a reform agenda has been instituted by successive governments. At first slowly, with the introduction of the purchaser provider split in the late 1980s, and now in an increasing crescendo of policy initiatives, patients are being given the freedom to make choices about where they go to get treated.

A more nuanced description of the rise of the 'choice agenda' in England might place the increasing importance of competition and choice as one of three overlapping reforms since the NHS plan in 2000, as summarised by the Kings Fund (56). The first of these is a period of sustained increasing funding for the NHS running from 2000 until 2008. The second major policy agenda was performance management with substantial adherence to targets and potential punitive action for not reaching them - characterised as 'targets and terror' by Hood and Bevan (57). Unlike the first two reforms however, which have since finished (or at least reduced and changed emphasis), choice and competition do not appear to be short-term policy changes. The operationalization of the choice agenda has occurred in a series of step wide increments summarised in Table 1.

**Table 1: Background to provider choice in the English NHS**

|      |                                                                                                                                                                                                                                                                        |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1949 | The NHS is founded                                                                                                                                                                                                                                                     |
| 1989 | NHS internal market splits purchasers from providers for the first time                                                                                                                                                                                                |
| 2002 | First limited piloting of patient choice between providers                                                                                                                                                                                                             |
| 2006 | Patients guaranteed choice of at least 4 local hospitals for most conditions and national referral system 'Choose and Book' launched for secondary care appointments<br>NHS Choices website launched, providing systematic comparable data around provider performance |
| 2007 |                                                                                                                                                                                                                                                                        |
| 2008 | Patients given choice of all hospitals nationally for most treatments                                                                                                                                                                                                  |
| 2009 | NHS Constitution enshrines patients' right to choose provider                                                                                                                                                                                                          |
| 2011 | Choice of named consultant team within hospital is allowed                                                                                                                                                                                                             |
| 2013 | Health and Social Care Act open care provision up to any qualified provider                                                                                                                                                                                            |

Starting with local piloting and limited, local choice, patients are now able to choose between any provide in the English NHS system for their care. The right to choose between providers has been secured within the NHS constitution (58). Critically, the payment structures have also changed, with the payment-by-results policy acting as an activity based funding system where the 'money follows the patient'.

### ***The mechanisms and effectiveness of competition***

Beyond the notion of the effectiveness of public reporting, a broader question has emerged that instead focuses on the more macro level question of whether the presence of competition improves outcomes. That is to say, does the presence of a market lead to improved quality, however the effect is mediated?

Whether choice is a good thing is still much debated. Coulter laid out the arguments for and against (59). She noted that choice has been argued to increase plurality of providers, encourage innovation, may drive down cost, improve responsiveness, enhance patient influence and eventually lead to better patient experience and quality, perhaps even at

lower cost. The converse arguments were that choice increase fragmentation, reduces continuity of care, undermines population based services, increases transaction costs, may destabilised existing NHS providers and might increase inequalities.

This is an area of substantial research and debate in the English NHS and further afield. The international evidence around effectiveness of choice and competition is thin, a systematic review by Smith et al (60) failed to determine evidence of effectiveness. Gaynor et al (61) found mixed results, but evidence that competition may improve quality in systems with administratively set prices – drawing heavily on evidence from Medicare in the US. Nick Mays suggested that evidence for “supplier competition in healthcare concluded that it is complex and equivocal” (62). In this thesis, I will limit more detailed examination to the recent discussion around the UK’s choice reforms. Separate, but methodologically similar papers by Cooper (63), Bloom (64) and Gaynor et al (65) suggest that areas where there is more competition at the hospital level (with fixed prices) have better clinical performance - as measured by specific clinical indicators. For example, Cooper looked at variation in mortality rates for acute myocardial infarction, with a difference in difference approach, and found that mortality rates improved faster in hospitals that faced more competition. Other work using similar methodological approaches has suggested that choice may also bring down waiting times (66,67).

These findings, particular around mortality, were subject to substantial criticism from academic peers. The subsequent claim and counter claim in a series of letters in the Lancet demonstrate the ferocity of the argument, with both proponents and critics of the value of competition pulling apart the methodology and the ideology of the other side. The critics

raised 1) a failure to prevent confounding, 2) inappropriate selection of a limited indicator (30 day mortality for AMI) that reflects pure emergency rather than planned care 3) confounding of developments in PCI treatment was not adequately addressed and 4) that the economists authors had no grasp of the face validity of their findings (68). Responding en masse, the authors of the papers demonstrating a beneficial effect from competition argued the strength of their data, that sufficient adjustment had occurred and that their opponents were ideologically driven (69). It might be fair to say that the policy question is 'hotly debated'.

A final ingredient to throw into this policy mix is the impact of recent scandals. The debate about care quality and the growing demand for transparency must, in the NHS at least, be viewed with knowledge of recent events that have raised the issue of poor quality care into the public and political consciousness. Starting with events such as the Bristol heart surgery scandal (70), more recent scandals – in particular the mid-Staffordshire crisis– have been played out in the glare of the media spotlight in the form of a disappointed and perhaps surprised public, an incandescent press and a begrudgingly regretful health service. These events have put the variability of health service quality further to the forefront in the public's mind. Polling soon after the release of the Francis Report suggested that 50% of the British public thought some NHS hospitals probably have problems like those seen in Stafford hospital, with a further 28% think that most or all hospitals do (71)(72).

#### **1.2.6 Patient choice behaviours**

Moving from the macro level policy debate, to trying to understand the behaviours of the human actor working within it, we need to consider how people respond to information and

behave when faced with choices about health providers. Here I will focus in particular on the literature from the UK, which is more relevant to the system being analysed in this thesis.

### ***Are patients aware of choice, and do they even want it?***

The English Department of Health became particularly keen to explore the evidence in this area, and asked a series of specific policy questions under the Policy Research Programme. It briefly ran a series of surveys about attitudes towards and use of choice. The 2010 NHS Choice survey (the last year it was run before it was cut for financial reasons) received 69,000 responses from NHS patients who had recently been referred to hospital. It found that 54% of patients were aware that they had a choice of hospitals. This had gradually increased from 30% of the public in 2006, when the policy was introduced. Not everyone was offered a choice. 63% of patients who were aware of choice remembered being offered it. Only 32% of those not aware of choice recalled it (73). A separate survey (of 1000 patients) by the King's Fund found similar results about awareness. Just less than half (45%) knew they had a right to choose, before they sought care. And again, around half (49%) remember being offered a choice of hospital. 69% of those offered a choice picked their local provider (56).

Another fundamental question is whether patients want choice of their healthcare provider (which interestingly is not a question that is asked in the major NHS funded surveys). Dixon and others for the King's Fund found that 75% of patients thought choices was either important or very important to them - with older people, those with non-white ethnicity and those with no qualifications valuing it more (56). In a separate question in the long

running British Social Attitude Survey, 72% of patients thought they should have at least some choice about the hospital they went to (74). However, Fotaki and colleagues noted that when asked to think about the relative importance of hospital choice compared to other issues in the health system, its relevance fell. They found 'choice' to be only the 11th most important aspect of care (out of 16), beneath other items such as car parking (75).

### ***How do patients use public data to choose providers?***

A consequent question is how the public make use of data about hospital quality that is put in the public domain. In other areas of science, there is a rich body of information about how humans process the information put in front of them. This has been explored comparatively less in healthcare around public reporting (76). Detailed survey work by Dixon and colleagues looked into the information sources used by patients who had recently been referred to hospital, and the criteria they were most likely to use when making a decision (56). This suggested that patients valued the quality of care, the cleanliness of the hospitals and the standard of facilities. They also found that patients made little use of the information that was available to them - only 4% had looked at NHS Choices (the NHS's main source of comparative information – to be discussed in detail later), and only 6% had looked at leaflets comparing providers. Patients were much more heavily reliant on their own personal experience (41% identified this), their GP (36%), and their families and friends (10%).

In terms of studying choice behaviour, there are two principle ways to do it. First, you can look at peoples' actual choices by following a cohort of people through the system. The advantage is that this is real data, but it is expensive, and because researchers have no

control over what information is presented, they do not know which are the key factors driving the decisions. The second option is to give people hypothetical choice decisions about where to seek care, to see which factors are key to their decision (77).

In terms of what patients actually did, large national surveys suggest 69% went with their local provider if offered a choice. However, the results were slightly different in the hypothetical choice experiments when information on provider performance was put in front of them. When faced with a hypothetical choice, 45% went with non-local choices (56). Some people were obstinate in the support of their local provider, with 20% choosing it every time, regardless of performance. This potentially demonstrates the difference between the theoretical model and actual real life behaviour. Older and more educated people were more likely to travel to a non-local provider. They also found some findings that went away from expectation. For example, patients were keener to travel to other providers in non-urban areas. Further, an observational study by Lavery and colleagues has suggested that even when there is widely reported public awareness of poor performance (using high profile national scandals as an example) changes in utilization by the local population in the English NHS are both short term and limited in size (78).

### **1.2.7 The decision making process**

The classical economists might think that healthcare is a simple market, full of logical actors making rational decisions about their care, in a free and informed manner. Accepting that many people do not have access to perfect information, this group still feel that when information about organisational performance is present, patients will use it logically to

make choice to maximize their utility. A more pragmatic approach suggests that the complexity of the marketplace, and the inability of humans to deal with the sometimes divergent information present, leads to decisions that are not taken within such a logical model.

Marshall and McLoughlin caution against the notion that “providing information to patients to enable them to make choices between providers will be a major driver for improvement’ (6). Instead, they suggest that patients do not view healthcare as a commodity to be purchased in an egocentric decision, but instead use a much more complex, social process. Instead of a rational approach, they apply other thought processes to their choices- including past experiences (they give the example of not selecting a high quality hospital because their grandmother died there), the views of others and their own personal expectations and fears. They note the 'strong sense of loyalty and indebtedness' that influences many health care decisions. This more complex and irrational decision making process has been described as 'knowledge construction' and quite separate from the 'information telling model' of the rationalist (76).

Similarly, Schlesinger and colleagues describe how an abundance of complex information could drive irrational decisions. They note that when faced with an unintelligible amount of information, or information people do not fully understand, they may drop back to simple rules of thumb (heuristic) in their decision making process, using only a limited or 'bounded' portion of the information available (79). This is described as a 'theory of bounded rationality'. Marshall suggests that information could be made more useful by 1) presenting the information as a trusted source, 2) making it directly relevant to the decision they are

making (for example presenting the correct condition specific data for the operation they are about to have), 3) making the data visually attractive, 4) presenting different types of data - as some people are more drawn to different types than others.

### ***How should information be presented?***

The way the information is set out is important too. A team of behavioural economists, working with the King's Fund, looked at how people made decisions, and where the presentation of the information made a difference. They found that the way the information was presented on screen made a difference to patients' decisions (with the public paying most attention to what they see first), and that different behavioural prompts could be used to affect the amount of weight given to a particular metric. (80) However, they also found the ability of these nudges to work varied with a variety of personal characteristics such as age and education.

## **1.2.8 Transparency and public reporting in England: the current policy environment in the NHS**

*"As you know, transparency is at the heart of our agenda for Government. We recognise that transparency and open data can be a powerful tool to help reform public services, foster innovation and empower citizens."*

*David Cameron in a letter to the Cabinet 7 July 2011 (81)*

### ***Transparency across government***

Having reviewed the academic literature, it is also important to consider the policy environment. As in other areas of government, the views of academia and policy do not

entirely align. There has been a determined push by the government in England to make data about public services open and transparent. As the quote from the Prime Minister above shows, this desire for transparency has emerged from the highest levels. This has occurred across various sectors of government, from the release of crime maps that show you all the reported incidents around your postcode (82), to the spending of local councils (83) through to the annual release of school performance data (84).

These changes have proved popular with the English public. Substantial numbers of people make use of the new 'open data': parents consult league tables for schools before deciding where to move house, and MPs have felt the wrath of the people when their expenses finally reached the public light of day. From a policy analysis perspective, this movement can be seen from the government perspective in a series of documents and reports over the last 15 years, which, despite changes of administration, have demonstrated continued progress towards greater public availability of information.

Under the previous Labour Government, the Cabinet Office report '*Excellence and Fairness: Achieving World Class Public Services*' was clear that improved access to data to allow meaningful comparisons, together with access to choice in public services, were part of the government's plan to improve the quality of performance across many different parts of the state sector (85). The report stated that the government would put citizens in control by "giving all users - from patients, to parents and local residents - clear and comparable information about service quality." A year later, in another report from the Cabinet Office '*Working together: Public Services On Your Side*', Gordon Brown's administration proposed the notion of an 'Information Revolution' to profoundly alter the nature of the public's

relationship with state funded services. In 2011, Sir Frances Maude reinforced this notion claiming that the aim of government is to be “open by default” (86).

### ***Transparency in health***

This movement towards transparency has also been prevalent in healthcare in England - with a number of important health organization performance metrics reported in public on the government's NHS Choices website (to be described in more detail later). Data on hospital level performance has been reported for a number of years, initially using simple metrics such as mortality rates using data from the routine health information systems such as Hospital Episode Statistics (87). Other high profile issues, such as healthcare associated infections with meticillin resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile* have also been reported publicly by the Health Protection Agency.

The coalition government of 2010 set about making numerous changes to the structures and organizations of the NHS with the Health and Social Care Act 2012 (88). However, their language and ambition around the public availability of information was one of the few aspects of healthcare to feel consistent across the change of regime. The NHS information strategy *The Power of Information (2012)* describes, despite claims of novelty, a consistent desire to present data in the public domain in a usable format, and for citizen to use to make decisions and to improve services (89). It states an aspiration for: “A culture of transparency, where access to high-quality, evidence-based information about services and the quality of care held by Government and health and care services is openly and easily available to us all.”

The new organization with a responsibility for distributing information about health services, NHS England, has continued a transparency agenda in its operational plans. In the initial strategy document, '*Everyone counts: Planning for patients*' – they summarize the key message as: “more choice, more transparency, more patient participation and better customer service” (90). In their follow up future strategic plan they state a belief that “commitment to transparency and increasing patients’ voice are fundamental to improving patient care” (91).

More recently, there have been renewed calls for performance information at the level of clinical team, or even individual physicians (92). There was a specific commitment towards this idea by the Department of Health in 2010 (93), with the intention of publication of clinical team level performance data for both outcomes and patient experience (90). National surgeon level outcome data for vascular surgeons was published on NHSC in June 2013 as the start of this process (94).

What has been presented here is a series of government documents and policy objectives, some broad in focus and some more narrowly applied to healthcare, that state a common desire for increased transparency as a tool for service improvement. While the language is consistent in its aspirations to the point of duplication, this series of reports demonstrated a definite and considered set of aims: that transparency will allow the public to make better choices and drive improvement (via Berwick’s selection pathway). Further, as all mainstream parties are speaking a similar language on these topics, it seems unlikely that this course will change in the near future.

### **1.2.9 The relationship between transparency and choice**

In this description, I have considered the move towards both transparency and choice in policy together as a driver for the creation of online feedback systems. These two concepts, transparency and choice, are inter-related and often conflated in the policy discourse.

Certainly, choice depends on transparency of performance as a necessary assumption, if it is to operate rationally and deliver benefit to those patients in the system. It is, however, important to recognise that both are separate phenomena, and could be viewed as independent drivers of the development of online feedback. In particular, the transparency agenda could be viewed as a separate initiative, as according to both Hibbard and Berwick, it can have an effect independent of a choice agenda as hospitals are motivated internally to improve their performance – either because of an intrinsic desire to improve (17), or to protect their reputation (21).

### **1.2.10 Public reporting, choice, and selection: limited conclusions**

#### ***The myth of selection***

So far we have seen that there is a policy direction to put information on healthcare provider performance transparently in the public domain. And since the start of choice policy in the NHS, patients can use this information to select their location of care. With these building blocks, and Berwick's proposed mechanism of action, one would expect to see improvement in performance. However, despite this clear logic, it may not have worked as expected. As Berwick pointed out, neither the dynamics of the selection or improvement pathways work reliably because of a lack of capacity in the organisations and individuals acting within either route (17).

While it does appear that hospitals are motivated to act in response to comparative information, and perhaps even more so by that comparative information being made public, it also appears that the public is a curiously stubborn entity. Those economists hoping that the public will behave in a rational manner, using information to maximise their utility, seem to struggle to demonstrate that this occurs. Rather, the public appear to make choices influenced by a variety of other factors including their own experiences, trusted data sources (such as the views of their primary care doctors) and their capability to use and understand the information available. From the point of view of the 'classical economist' most patients making choices in the NHS are not the informed and empowered free agents of a perfect market. This all leads to a conclusion that of the various pathways that public reporting might improve quality, selection is the weakest.

### ***Web based feedback is an extension of public reporting***

A key future question for the NHS (and other market based health systems) is how to provide the correct mix of data in a presentable format, to drive optimal quality improvement. There is a desire to produce information about care quality that is relevant to the public and that they can understand and engage with. This desire for transparency and public engagement about healthcare quality has led to a novel type of public reporting: patients' providing feedback about their care online (which is discussed in detail in chapter 2). One potentially important question is whether this new, simple and human approach to public reporting could revive the selection pathway as a valid mechanism of quality improvement, or drive internal improvement activity by care providers.

## **1.3 Driver 2: The rise of patient experience as a measure of healthcare quality**

The last quarter of a century has seen an increased focus on the patient within healthcare, and an increased role for patient experience as a driver of health system improvement. This has been expressed through a variety of metaphors, phrases and techniques, including patient centeredness (95), co-design or co-creation of services (96), and a new language around customer service. (97). One of the seminal publications that cemented this new priority on experience was the World Health Report of 2000 (98), entitled *Health Systems: Improving Performance*, which set out three fundamental objectives of health systems. As well as expected goals of improving the health of the population they serve and providing financial protection against the costs of ill health, the document stipulates *responding to people's expectations* as a core aim of health systems around the world. Although this component is often neglected, and has the largest capacity for subjective assessment, this placed the topic as key goal for policy makers.

### **1.3.1 Defining Quality in healthcare**

At the same time, there has been an increasing awareness of a concept of 'quality' in medicine - that is to say a measurement of the performance of the healthcare system. This increasing focus on quality is a response to the threat of rising expenditures on health, an increasingly elderly population with complex multi-morbidities and a more developed awareness of the unwarranted and wasteful variations in care (99). Quality is a difficult term to define, at risk of meaning different things to different people, and so people have proposed a number of different frameworks with which to measure and understand it.

Early attempts to define quality included those of Donabedian, who suggested that the concept had generally been looked at from three perspectives. In a seminal work in the *Milbank Quarterly* in 1966, he described these components as: 1) structure, 2) process and 3) outcome (100). In this case *structure* would be those resources available to a service, and the pattern of organisation, which might include the number of doctors and nurses, money, or access to specific technologies and facilities (e.g. MRI scanners or stroke units) (101). *Process* would be those measures of the effectiveness of the internal pathways of care, such as the proportion of patients treated according to clinical guidelines or the proportion of patients waiting more than 4 hours in A&E. *Outcomes* are those eventual effects that the activity has on the population it serves, which might include mortality rates, functional status, quality of life or patient experience. While the Donabedian framework was widely adopted, he also noted that quality in health care was a problematic concept with multiple definitions and that “*more complete conceptual and empirical exploration of the definition of quality is needed* (100).”

Following this work, Maxwell also attempted to define quality, this time creating a set of functional criteria: effectiveness (does it work), efficiency (cost/benefits), relevance to need (appropriateness), equity, accessibility, locality and social acceptability (102). In this approach, he broke away from the sequential measurement of quality that is implied in Donabedian’s framework and moved towards a series of separate ways to view the outputs of the system.

Building on this is perhaps the best known definition of quality internationally, from the Institute of Medicine (IOM) in the US, which developed a multifaceted approach to considering the concept in their 2001 report, 'Crossing the Quality Chasm'. They termed quality: "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge". In a similar approach to Maxwell, they divided quality into a set of different components, which they viewed as: 1) effectiveness, 2) safety, 3) equity, 4) timeliness, 5) efficiency and 6) patient centeredness (15).

In England, there has been a sequence of different approaches to defining the concept of quality for the NHS. Initially drawing on the IOM's definition, quality was reconceptualised and reduced in the Darzi review as simply: 1) effectiveness of care, 2) patient safety and 3) patient experience (103). This formulation deliberately put a primacy on outputs of systems and outcomes for patients, in a reaction to the perception of an overreliance on the measurement of process that lost sight of the end result in a target driven environment. In this way, some of the earlier thinking by Donabedian and others been replaced with an emphasis on end results, rather than the measurement of the means which produce the ends.

Most recently, an even simpler description has appeared in the language of the Department of Health, proposed by the Secretary of State, which defines quality as either quality of treatment (incorporating traditional metrics of effectiveness) and quality of care – which is concerned with the experience of patients, and the dignity and compassion that they are treated with (104).

Given there are so many different definitions of quality actively in use, for the purposes of consistency in this thesis I have considered quality within the IOM's six-part definition.

### 1.3.2 The role of patient experience in Quality

This series of conceptualizations of quality described above demonstrate the variety of different definitions that have been attempted, and that are currently being used, in different health systems around the world. It is noticeable that within each of these systems, there are attempts to incorporate some form of patient experience.

Arah and colleagues in OECD work (105) reviewed a number of different health system measurement approaches across different countries, and found patient experience (in some capacity) incorporated in all of the national approaches (Table2).

**Table 2: Dimensions of healthcare effectiveness**

| Dimensions                                              | UK | Canada | Australia | USA | ECHI | Commonwealth fund | WHO | OECD |
|---------------------------------------------------------|----|--------|-----------|-----|------|-------------------|-----|------|
| Acceptability                                           |    | ✓      |           |     |      | ✓                 |     |      |
| Accessibility                                           | ✓  | ✓      | ✓         | ✓   |      | ✓                 |     | ✓    |
| Appropriateness                                         |    | ✓      | ✓         |     |      | ✓                 |     |      |
| Care environment and amenities                          | ✓  |        |           |     |      |                   |     |      |
| Competence or Capability                                |    | ✓      | ✓         |     |      |                   |     |      |
| Continuity                                              |    | ✓      | ✓         |     |      | ✓                 |     |      |
| Effectiveness or Improving health or Clinical focus     | ✓  | ✓      | ✓         | ✓   | ✓    | ✓                 | ✓   | ✓    |
| Expenditure or cost                                     |    |        |           |     |      |                   | ✓   | ✓    |
| Efficiency                                              |    | ✓      | ✓         | ✓   |      |                   | ✓   | ✓    |
| Equity                                                  | ✓  | ✓      | ✓         | ✓   |      |                   | ✓   | ✓    |
| Governance                                              | ✓  |        |           |     |      |                   |     |      |
| Patient-centeredness or patient focus or responsiveness | ✓  | ✓      | ✓         | ✓   |      |                   | ✓   | ✓    |
| Safety                                                  | ✓  | ✓      | ✓         | ✓   |      |                   |     |      |
| Sustainability                                          |    |        | ✓         |     |      |                   |     |      |
| Timeliness                                              | ✓  |        |           | ✓   |      |                   |     |      |

Source: Adapted from Arah et al. 2006 (105)

When conceptualised in the figure below, the progression of experience of the patient can be seen to be slowly, but increasingly, prevalent in our descriptions of system measurement - at least within the English health system. It is almost a stairway to the increasing importance of experience (Table 3).

**Table 3: Different attempts to incorporate patient experience in to healthcare quality framework**

| Framework  | IOM (2000)           | Darzi (2008)       | Hunt (2013)          |
|------------|----------------------|--------------------|----------------------|
| Components | Access               | Patient Safety     | Quality of treatment |
|            | Safety               |                    |                      |
|            | Timeliness           | Effective Care     | Quality of care      |
|            | Effectiveness        |                    |                      |
|            | Equity               | Patient Experience |                      |
|            | Patient Centeredness |                    |                      |

This change in definition in England is in line with parts of the academic community. In a recent editorial in the NEJM, Manary and colleagues consider the value of patient experience in and of itself as a key measure of healthcare quality (106). They suggest that there is evidence that patient experience should be views as "robust, distinctive indicators of health care quality". That said, they also recognise a number of complications in its use as a metric around the need for standardized methodological approaches to reduce the risk of bias introduced by issues of survey design, the need for risk adjustment, and the mode and timing of survey administration. In particular, they reinforce the need for rigorous adoption of patient experience surveys incorporating a standardised approach to assessment.

Meyer and colleagues have also criticised the current proliferation of quality metrics. They suggest this has reduced the focus on measuring what really matters, and instead produced an almost incomprehensible variety of metrics, many of which have a marked process focus, as these are more useful tools for driving improvement. Instead, they suggest that indicators should be balanced between the needs of patients to judge quality and the needs of providers to drive improvement, but must also be parsimonious to measure quality, outcomes and costs based on end user needs – suggesting a larger role for experience (107).

Berensen, Pronovost and Krumholz took a more strident view when they claimed that they: “believe that measuring patient experience is not only important because it can facilitate care that improves clinical outcomes, but also because it represents an important outcome in its own right.” (108)

Going even further, there are those who argue that patient experience should be less a component of quality, and more the overriding consideration through which to see all components of quality. In his article, Patient Centred Medicine, Harlan Krumholz hints at an almost entirely hedonic approach to quality in which the subjective experience trumps all other considerations (109). Although this is not a mainstream view, the fact that it even being considered signals a gentle change in the way we think about care. This view was reinforced at the policy level by NHS England’s statement in its forward plan *Everyone Counts*, their strategic plan for the 2013-16, where they state: “The final arbiter of the outcome of any NHS interaction is the patient’s experience” (90).

### **1.3.1 Patient experience as a quality metric**

As a consequence of this shift in emphasis, there has been a move to measuring the effectiveness of care by incorporating the subjective, patient point of view. This is in contrast to the previous focus on effectiveness, which tended to look at more binary objective outcomes, particularly mortality, cure and survival rates. This has led to the development of new metrics and patient experience measurement tools. As with all metrics, it is important to understand how they relate to one another. In Chapter 2 I look in more detail at the nature of the association between patient experience and technical quality, as for many this will define its usefulness as a quality measure. Here I consider the various forms available in the English NHS at the moment.

The NHS in England has led the way in terms of large scale, standardised patient experience measurement - particularly through the deployment of large national surveys. The two largest are the national inpatient and general practice surveys (although smaller service specific surveys are administered for a variety of specialty areas). Inpatient care in the NHS is measured by the NHS Inpatient Survey, a yearly national paper-based survey of patients in all NHS hospital trusts conducted by Picker Europe for the Care Quality Commission, England's health regulator. It was launched in 2001. It surveys a random sample of patients who have been treated in NHS hospitals for at least one night. Over 100,000 surveys are sent out yearly, and recently, the response rate has been over 50% (110).

A similar survey exists in primary care - the national General Practice Patient Survey. The NHS first performed a national survey of general practice experience in 1998 (111) but it was not until 2006 that the government launched an annual national survey. Since then, it

has been used to benchmark general practice performance, but also as a tool to inform payments to GPs through the Quality and Outcome Framework (QOF), the national primary care pay for performance system. This is a large, postal survey (sent to 2.7 million patients in 2011/12), including a sample of each of the more than 8000 practices in the country, and with a response rate of around 35% (112).

These national surveys are similar to those now used in some other countries. Notably the hospital Consumer Assessment of Health Provider and Systems Survey in the US (HCAHPS) – which has been running since 2008 (113). However, they have not yet been adopted fully in high-income health systems. For example, in Australia, Canada, New Zealand and several European countries there is no national level systematic approach to data collection for patient experience (114).

### **1.3.2 Limitations of the current survey approach**

Critics have asked if these are really the correct tools for a modern, responsive service to measure, understand and improve its performance. Advantages of the national patient surveys are that they provide comparable patient experience results from a standardised methodology, the random approach acts to reduce the selection bias, they are large in scale, and they achieve good response rates (by comparison with many surveys). However, they can be criticised for lack of frequency – as the surveys are deployed only once a year over a defined period - and lack of timeliness - as the survey results are only made available several months after the survey was first deployed. As an example, the National Inpatient Survey in 2012 was deployed in September 2012 to January 2013, for patients treated in

June to August 2012, and the results were only made public after a long period of delay in April 2013, a full 10 months after care was received (115).

### ***Using the information to drive improvement***

Evidence suggests that information collected in patient surveys is rarely used to spark debate on organisational improvement. National surveys are rarely used at the frontline to foster and encourage debate and improvement. Mountford and Shobjana described the disconnection between the measurement of experience indicators and their adoption and use in improvement activity (116). Robert and Cornwell suggest that the presence of national approaches may have created a disconnection between the reporting systems and the local organisations that use them - as taking part have become more of a 'box-ticking' compliance activity (114). They go on to highlight the need to think about practical approaches to measuring patient experience to drive improvement activities. They suggest the need for a standardised simple question, collected as close to real time as possible, together with a limited number of questions about relational aspects of care (114). A review by DeCourcy and colleagues looking at the output of the various studies done with the NHS inpatient survey results, noted that the patient survey in itself is not a quality improvement tool and that 'simply providing feedback' does not lead to quality improvement activity. (117).

### ***Issues of sampling, bias and representativeness***

These surveys also have issues in terms of the population they sample, the methodology and size, and how well they represent the broader public. This has been studied particularly in the GP survey. Paddison and colleagues demonstrated that although the case mix of patients in each practice has little effect on the results for the majority of practices, for a

small proportion serving vulnerable groups, unadjusted scores underestimate patient satisfaction (118). Lyratzopoulos also highlighted issues in reliability related to the sample size. He found that questions answered by fewer respondents had less reliability than those answered by more, and consequently suggest that questions answered less frequently might be less robust indicators of care (119). These studies do however give us clues about what drives patient satisfaction. In a further study, Paddison and colleagues also found that doctor communication followed by helpfulness of the reception staff were the main drivers of overall satisfaction with primary care (120). In contrast, the ability to obtain an appointment had a much weaker contribution with overall satisfaction.

#### **1.3.4 Other approaches to measuring patient experience in England**

Other methods are also used for monitoring and responding to patient experience in the NHS, particularly for those who have bad experiences of care. These have been studied less from an academic point of view, but are widely used by the public.

##### ***Patient Advice and Liaison Services (PALS)***

All trusts were made to set up Patient Advice and Liaison Services (PALS) services in 2002 by the Department of Health, in order to provide a new route for patient engagement and service user involvement. A national survey demonstrated that they exist in almost every hospital trusts in England, and when self-assessed, meet the standard set by government (121). Academic evaluations are limited, but South and colleagues reviewed the grey literature and concluded that the PALS model has had limited effectiveness 'bridging the gap' between users and the health service – and instead tend to defend service interests.

They argue that health services should instead put resources into fostering stronger partnerships with local communities and the voluntary sector (122).

### ***Complaints***

Complaints in the NHS are managed by a statutory procedure. They are both common and badly managed. In 2011/12, the NHS received more than 160,000 complains (3000 a week) (123). Rates of complaint about the NHS are rose by 8% from 2011- 12 (124). Commentators have noted that the NHS often fails to respond or learn from them adequately (125), and tends to handle them badly (126). The Parliamentary and Health Services Ombudsman (PHSO), which is the highest-level body for dealing with complaints handled 4,732 complaints that could not be resolved at a lower level in 2012 (127).

These complaints do however give us an insight into where the NHS is going wrong, from a patient experience perspective. For acute services (excluding primary care), 46% of complaints related to the medical profession. Nurse and midwives together accounted for 22%. Cutting the data a different way, 32% were about in patient service, 28% outpatient and 9% A and E. In terms of general topic area, 46% were about clinical care, 12% about staff attitude, 10% about communication and 8% about delays (127).

### **1.3.5 The rise of new metrics in patient experience:**

Other methods of measuring patient experience have also emerged over the last five years, particularly within the NHS context.

### ***Friends and Family Test***

In an efforts to improve responsiveness post-Francis Inquiry, the government has announced a 'Friends and Family Test' (128). This is a simple test, whereby a sample of patients are asked whether they would recommend their care provider to their friends and family (measured on a five point Likert scale), either while in hospital, or shortly after leaving hospital. The government has issued guidance on how it should be carried out (which may be by phone, text, paper surveyor in person), and initial results were published in August 2013 (129).

Despite the first results only recently being published, it has already encountered a number of criticisms. In particular, some have suggested that the question is offensive to ask (would a cancer patient ever recommend their treatment), the methodology is too unstandardized to allow any comparison and that different specialties are highly likely to get substantially different rates depending on the likelihood of successful outcome (for example between maternity care and pain control) (130). Qualitative testing by the Picker Institute (131) also found that many patients could not disentangle answering the question about the care provider they had seen, and their view of the NHS in general.

### ***PROMS***

The focus on experience has also seen the rise of patient reported outcome measures (PROMs) from the obscurity of academic debate, to mainstream policy and a national programme of delivery (132). In particular in the NHS, patient reported outcomes have been collected since 2009 for hip and knee replacements, hernias and varicose veins. Although not capturing experience of care, they capture the end result of treatment. For the

purpose of this thesis, however, I will focus on the patient reported experience measures – relating to their experience of care - rather than PROMS.

### ***Patient Experience Measures: Where we are***

Looking across this variety of indicators, we see that the NHS in England is taking a lead internationally in measuring patient experience. However, despite the variety of different systems in place, there are still limitations in the current approach, in terms of how well used it is by healthcare organisation to drive improvement, and how it used by patients to voice concerns.

#### **1.3.6 The policy clamour for feedback Patient feedback in the NHS**

Despite the various approaches described above. There is a desire in the NHS for stronger patient feedback systems about their health experiences. This has been spurred on by a series of care quality scandals, in particular at Stafford Hospital. The subsequent Francis enquiry into the events there found a number of weaknesses in the system that had allowed poor care to occur, but central amongst them was a lack of focus on the voice of the patients. Patients had been speaking out about the poor quality of care for a substantial period before the crisis was recognized by the formal organizational structures (133). Two quotes from this report make his point.

*“It is service users, including visitors and families, who are likely to be the first to witness poor outcomes or signs that standards are slipping”*

*“What is really important is information accessible and usable by all allowing effective comparison of performance by individuals, services and organizations”*

Francis Inquiry into Mid Staffordshire NHS Trust, February 2013

The public are also supportive of a stronger feedback mechanism. A poll by IpsosMori in 2013, shortly after the Francis report on Mid Staffs scandal was published, found that 40% of people rated ‘listening to patients about how services could be improved’ as the highest priority to prevent further failings (72).

The desire for improved feedback mechanisms is clear at the central policy level in the UK too. In the NHS Commissioning Board's planning document for 2013/14, they highlight areas for commissioners to deliver including: 1) rapid comparable feedback on the experience of patients and carers and 2) building a capacity and capability in both providers and commissioners to act on patient feedback (90). A national ‘Patient Feedback Challenge’ was started by the Department of Health in 2012 (134) to encourage adoption of patient feedback mechanisms.

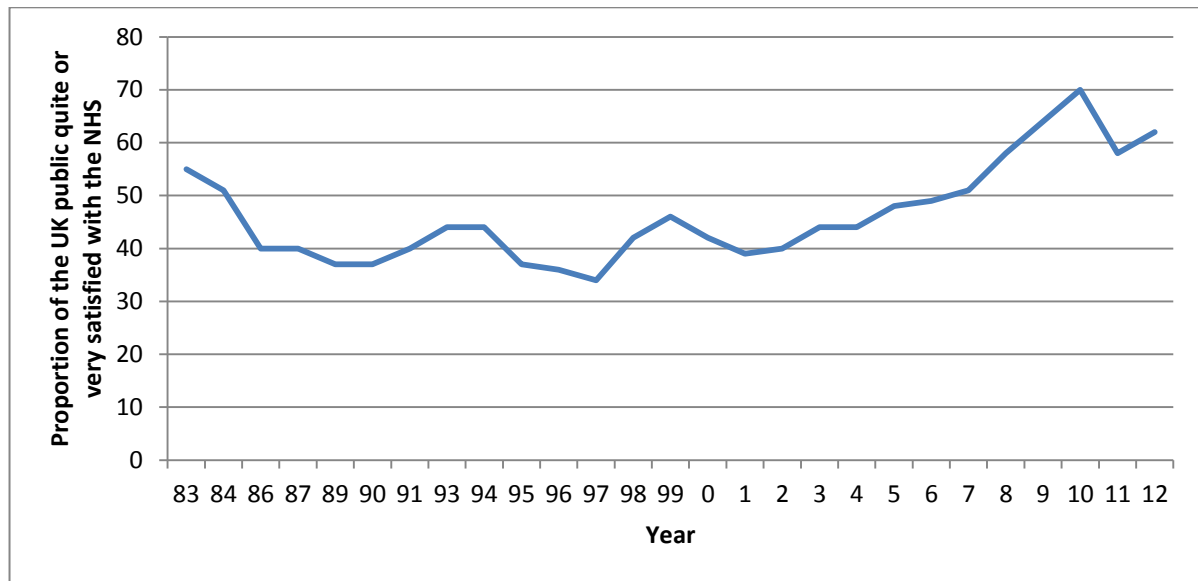
The policy think-tank NESTA has also suggested that our relationship with public services is changing, involving a more two-way discussion between citizens and the state than traditional, paternalist one-way interaction. As a consequence, the state needs to become better at responding (135). It highlights the capacity to learn from comments and complaints – and that the characteristics of new public service (which they describe as a *relational* state) are profoundly different to the way they used to be run (described as a *delivery* state).

### **1.3.7 Changing expectations of healthcare**

There is evidence of a changing set of expectation from the public about what health care should be like. The demographics of the English population are changing, and in particular ageing. The 10 million people currently over 65 expected to increase to 15 million by 2030 and 19 million by 2050 (136). As the bulge of the baby boomers hits the age where they might be expected to make use of more NHS services and have an increasing burden of chronic disease (137), data from public polling suggests we might face a miss-match between what the NHS can deliver, and that the public expects.

The British Social Attitudes Survey has been monitoring national attitudes toward the NHS since 1983, with a standardised survey methodology that lends itself to year on year comparison (138). For example, we see that on a societal level, satisfaction with the NHS had been increasing for a period of 10 years to a maximum in 2010. Since then there has been a striking fall in public satisfaction from 70% to 58% between 2010 and 11 (Figure 3). Several explanations have been put forward for this recent dip- including dissatisfaction with new government policy in the area, uncertainty about reform, a consequence of flat funding for the NHS and increasing expectations (139).

**Figure 3: Public satisfaction with the NHS over time**



Source: British Social attitudes Survey

Taken together, we see that there is a renewed emphasis on the importance of patient experience in health systems, and expectations about health experience are changing, possibly becoming more demanding. At the same time, there are limitations in our current approaches to measuring experience, and there is a clear desire for better feedback systems in health systems.

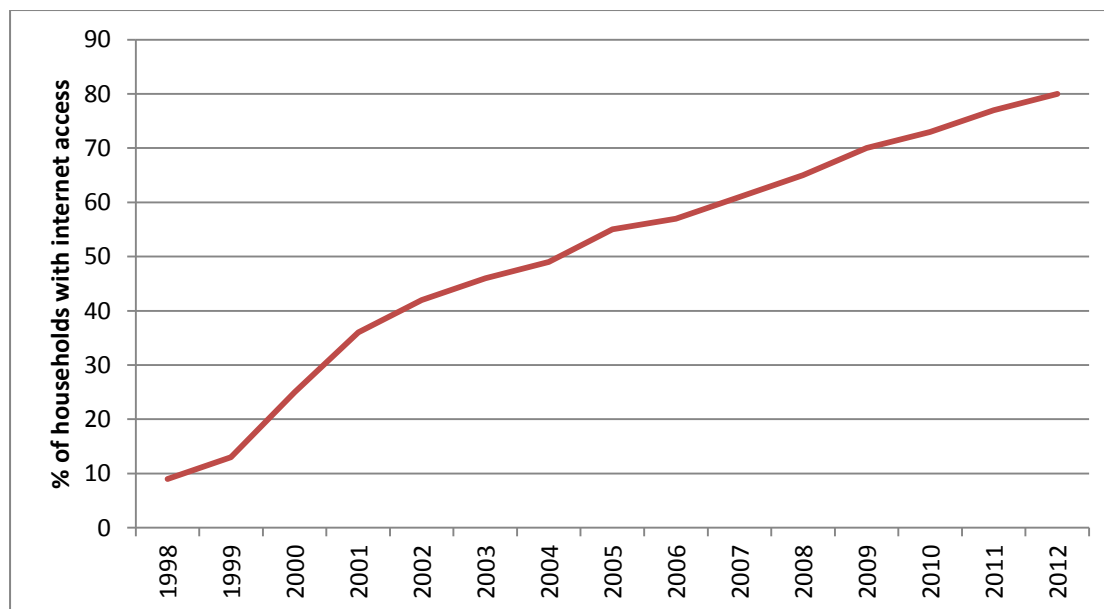
## 1.4 Driver 3: Societal Trends and the Internet

### 1.4.1 The rise of the Internet

The arrival of the Internet has profoundly changed the way we live many aspects of our lives. In England, as with many other developed countries, its presence is pervasive.

According to ONS statistics, 21 million households in Great Britain (80 per cent) had Internet access in 2012, an increase from 19 million in 2011, and from 7.1 million since 2006, when records started to be collected in a comparable fashion (140) (see Figure 4). In addition, 67% of adults use a computer every day. Worldwide the figures are more dramatic. In 1999, 38 million people had broadband Internet via any mechanism; in 2013 1.2 billion have it on their phones (141).

**Figure 4: Availability of the Internet in the UK**

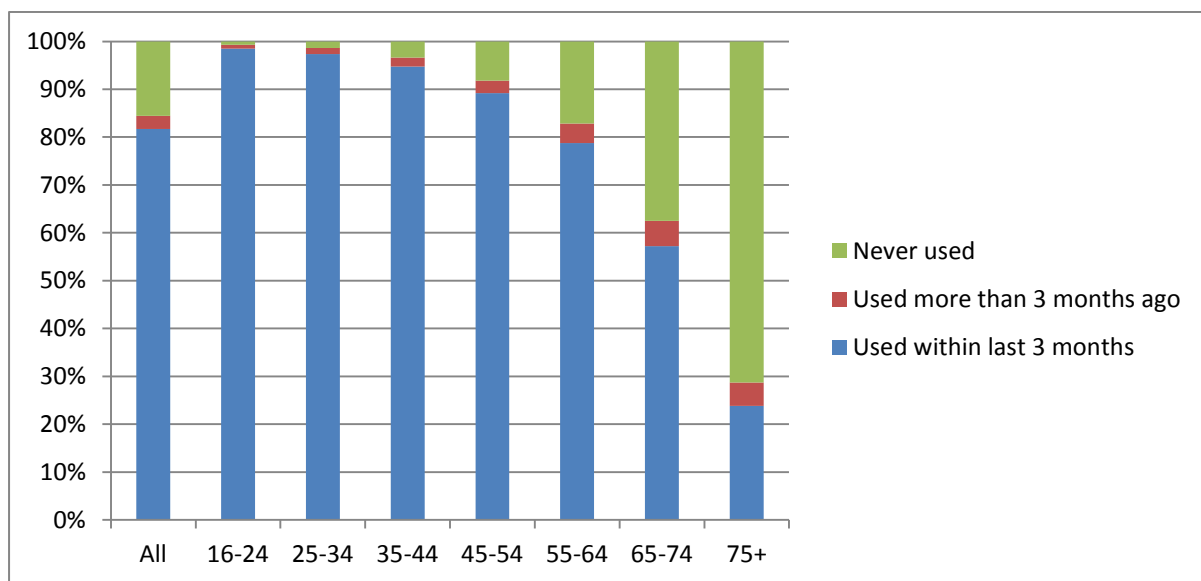


Source: Office for National Statistics

Despite the rapid increase in use, there are some groups who are less engaged. Important factors include age and income. While more than 99% of people aged between 16 and 24

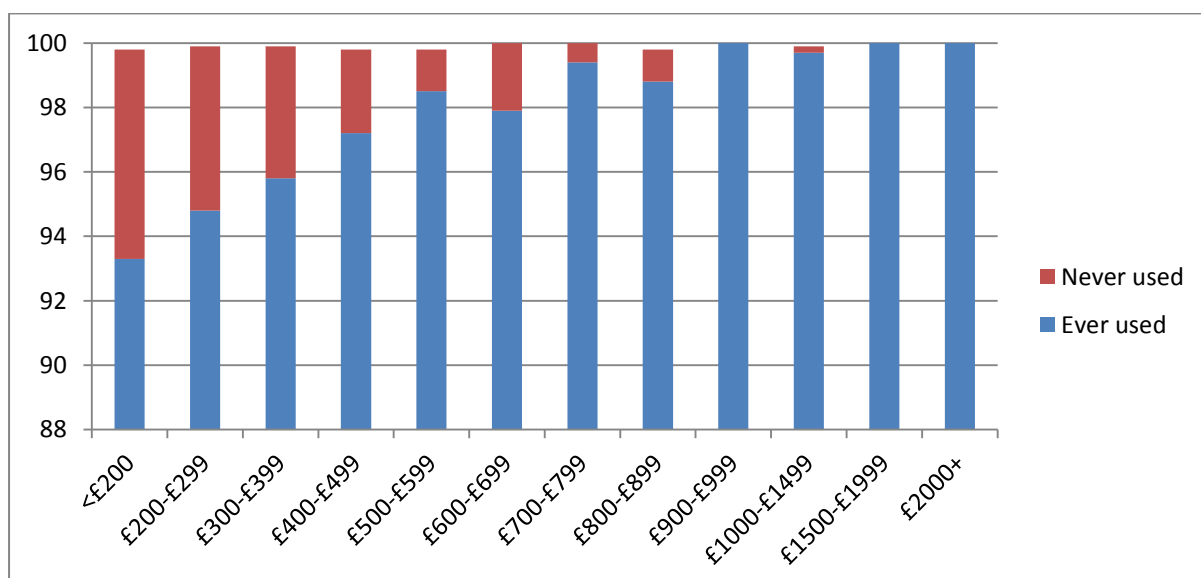
have used the internet, this figure is much lower in those over 75 years, where only 29% have used it (Figure 5). Similarly, Internet use is higher in those groups with high incomes, with a noticeable drop off at the lowest end of the income scale (Figure 6). These differences in access have been categorised as a 'digital divide' (142).

**Figure 5: Internet use by age group, England, 2012**



Source: Office for National Statistics

**Figure 6: Internet use by weekly income, England, 2012**



Source: Office for National Statistics

### **1.4.2 Changing behaviours around the Internet**

Data about Internet use tends to both astonish, and be hyperbolic. It is also often provided by the companies involved themselves, so must be noted with a sceptic's caution. Internet commerce is certainly big. Google's advertising revenue (\$43 billion in 2012) is bigger than the revenue of all print media (143). Amazon's revenue (\$61 billion in 2012) is bigger than the GDPs of the poorest half of the world's countries. Social networking websites such as Facebook grow closer to a notion of species level penetration, with 1.11 billion users and 665 million daily active users (144). The website Twitter, a micro blogging service allowing users to leave short messages (of less than 140 characters), has more than 500 million accounts, and 340 million Tweets (the short messages) are sent daily (145). Internet commerce can also be disruptive. Redit (a popular discussion website) has 63 million visitors a month, 4.4 billion page views and users spend 39 minutes on the site on each visit on average. It has 22 employees (141).

Along with this increasing usage and availability, the Internet has altered many of our commercial and social behaviours in the UK. Data from the British Social Attitudes Survey are revealing and representative. 98% of music singles tracks are now bought online (146). In 2009, 66 per cent of adults aged 16 to 74 in the UK had used the Internet to purchase goods in the last 12 months (147). It has changed the way many people, and particularly the young, communicate with friends, organisations and companies. Of those Internet users aged 16 and over, 44% had a profile on a social networking site by 2009, doubling from 2007 (when it was 22%). The Internet has also had an effect on information seeking behaviours and sources: 46 per cent of adult Internet users in the UK had obtained information from a public authority website in 2010.

### **1.4.3 Web 2.0: The changing way we use the Internet**

As Internet use increases, the patterns of use have also changed. A useful concept is the notion of Web 2.0 (148). This is a difficult concept to define, but rests on a notion of increased interactivity with the Internet by the people using it. Rather than just receiving information, the user also participates in the generation of new information. Tim O'Reilly - a internet pioneer - describes a series of values where 'collective intelligence is harnessed', and the 'users add value' (149). Web 2.0 has includes the idea of 'user generated content' where the people using a website also contribute information that is useful for others. This might be the review of a book on Amazon, the review of a seller on eBay, uploading a homemade video to YouTube or updating your status on Facebook. What is more, this user-generated content has been demonstrated to be a powerful tool for including consumer purchasing, particularly when it comes from friends in a social network. Iyengar and colleagues performed a series of empirical experiments on a large Korean social networking and found evidence that purchases were influenced by friends behaviour (150). In the marketing literature, it has been argued that the provision of this information, the ease it is obtainable and the breakdown in hierarchies of information is leading to the rise of the 'empowered consumer' (151).

### **1.4.4 Online behaviours around healthcare**

Despite these substantial changes in other fields, the extension of online interaction into healthcare remains relatively limited - so far. Data in the US is collected more regularly than in the UK, and there we are aware in 2013 that 81% of adults use the internet, and 59% say they have looked online for health information (either for themselves or a friend) in the past year (152). What is more, 35% have used the Internet to try to diagnose their own or

someone else's medical condition. This survey also looked into information about where patients got their information from when they had a serious health problem – and while unsurprising 70% went to a doctor or health professional, and 60% received informing from family or friends, 24% stated they received information from other with a similar condition. Although imperfect data, this was a national, randomly selected phone survey, and probably represents the best snapshot available.

Data in the UK is sparser, and often derived from ad hoc surveys, rather than continued polling around fixed questions. The NHS Choices Public Omnibus survey, conducted in 2009, found that 52% of adults had used the Internet to look for health information. However, using the information to choose hospital seemed less likely. 3% said they had already used the Internet to choose a hospital or health service, and 23% said they were very or fairly likely to do so in the future. 39% said they would be very or fairly unlikely to look on the internet for this sort of information (147).

#### **1.4.5 The rise of online empowered patient**

At the same time, we have seen the rise of empowered patients using the Internet. Deborah Lupton describes the emergence of a new type of patient, which she terms the 'digitally engaged patient' (153), who is active in the management of their diseases online, and in the sharing of health experiences. Examples of these patients can be seen in rapidly growing patient communities online. Some act as discussion fora, where patients can share their own experiences of illness, and experiences of care and particular treatments. In the US, a social network for patients, *Patientslikeme*, has more than 100,000 members who leave

descriptions of their condition, treatment and symptomatology online in an open format (154). In the UK, companies such as *Healthunlocked* have seen rapid growth and usage. They now have over 700,000 people visiting their online discussion communities focussed around health topics each month (155).

At the same time (or perhaps in response to the ground swell of a grassroots movement), the medical establishment has started to respond to these more demanding and aware patients, with an editorial by Fiona Godlee in the BMJ recently declaring the start of a 'Patient Revolution' (156). Muir Gray, with characteristic precision, described this as a third healthcare revolution, following on from a first revolution in the 19<sup>th</sup> century (the reduction of mortality in line with public health infrastructure) and a second healthcare revolution (the development of highly technical care and evidence based medicine) (157).

The revolution already has its figureheads. Self-defined 'ePatient' Dave deBronkhardt, has been advocating for the role and importance of patients within health systems and care pathways (158). His appeals to give patients access to their health information online have viewed on YouTube more than half a million times (159). Indeed the language he and others use is closer to that of a 'rights movement' than simple patient centred advocacy. For example his popular talk, an appeal for ownership of his own patient record is called: "Gimme my Goddamn Data!".

In extreme examples, patients have put their medical records online, in an attempt to 'crowdsource' a diagnosis and advice about treatment. Salvatore Iacones, a young Italian designer and open data advocate, put his entire life history, medical record and biodata in

the public domain after being diagnosed with brain cancer (160) – in the hope that the uncontrolled crowd might contain a wisdom beyond his conventional treatment process. This behaviour is not normal, but may give us a taste of where health behaviours may be heading in the future.

#### **1.4.6 Summary of the role of the Internet and health**

Here we have seen that the Internet is pervasive in most of society (with some notable exceptions) and has changed the way we live many aspects of our lives – and our consumer behaviour in particular. Although health is behind many other industries in adapting to a new online economy, there is a nascent digital patient movement, and online health seeking behaviours are moving from being a rarity, to being the norm.

### **1.5 Conclusion: The evolution of an idea - patients describing their care online**

So far in this chapter I have described a series of different social and policy forces: the rise of transparency in health systems, the renewed importance of patient experience and the changing use of the Internet. At the confluence of these different forces, and as a consequence of their interactions, is the spontaneous appearance of a novel phenomenon: patients describing their experiences of healthcare on the Internet.

In Chapter 1, I looked at these enabling and driving factors at play in the policy and practice of the health system in England. In Chapter 2, I will go on to explore the effect of the

confluence of these ideas, considering the limited, but growing academic work that seeks to understand online feedback in healthcare in more detail.

## Chapter 2: Internet based healthcare feedback

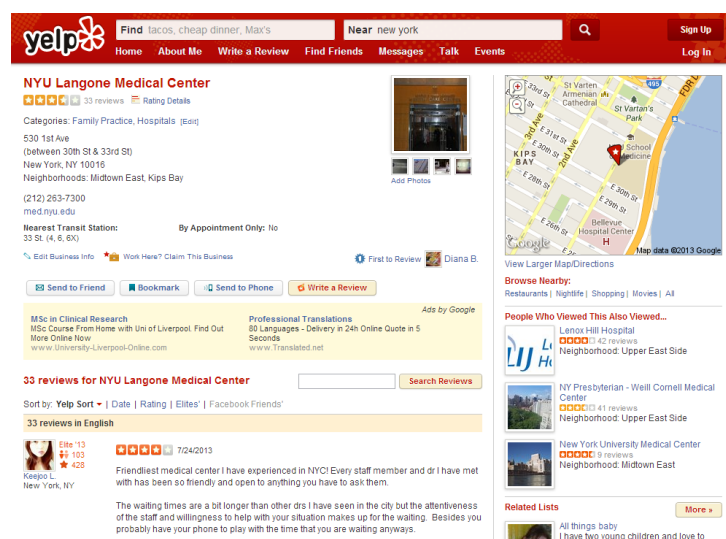
*"I wouldn't think of going on holiday without cross referencing two guide books and using Trip Advisor. We need to do something similar for the modern generation in health care."*

Ben Bradshaw, Minister of State for Health

### 2.1 Introduction to ratings websites

In the last chapter, I looked at a series of drivers that have led to the creation of a new phenomenon - patients rating and commenting on their healthcare online. In this chapter, I look at this in more detail – first considering healthcare rating websites. Online rating and review websites allow patients to describe the care they have received. They include a variety of formats, but crucially they ask a self-selecting group of patients to leave some form of commentary about their care. This is often in the form of rating aspects of their care on a scale, or describing the experiences they had in their own words in a free text section. Often they included both these approaches (see Figure 7 for an example).

**Figure 7: A typical page from Yelp, a popular website for health reviews**



Source: Yelp.com

Online feedback websites have been characterized repeatedly and simply as the ‘TripAdvisor approach to health’, including by the government minister in the quote above. However, it is possibly more complicated than that. I look at the early literature that exists in the field globally, to understand what we know about this practice. This includes looking at the extent of its use and its risks and benefits. I also consider what they try to measure, how people use them, and how clinicians and organizations respond to them. There is also a comparison to be made to the use of the idea in other industries, where online feedback is more mature as a consumer tool. I will also focus on a particular example in the UK, NHS Choices, which allows an opportunity to study this phenomenon in details at the national level for the first time.

## **2.2 The literature around unsolicited patient comment online**

### **2.2.1 The value of online review and comments**

At present, there is a lack of consensus about whether patients providing feedback online in healthcare is useful. Consideration of the usefulness of the practice in the literature, both popular and academic, tends to reflect a set of polarized views. Those making arguments around the importance of the transparency agenda have a tendency to propose them as a panacea against the historical inefficiency and hierarchy of the healthcare industry. Those arguing against, often - but not always - from the clinician viewpoint, tend to highlight the practical difficulties and the opportunities for misuse.

### **2.2.2 Arguments against**

Arguments against the value of online ratings and reviews have been summarized by McCartney and Lagu, and include the likely selection bias by those leaving reviews, a risk of gaming, risks of critical and unfair anonymous posting and the fact that good patient experience does not necessarily mean good care (4,161). I consider each in more detail below.

#### ***Good experience does not equal good quality***

It has been argued that patient reports on websites may damage the doctor patient relationship, as they set up an unreasonable requirement to meet patient's desires. McCartney argues that good patient experience does not necessarily equal good quality of care (4). Manary notes, a patient may mark you down for not giving you what they want (106). An often-quoted example is that of the GP asked by a worried patient to provide antibiotics for a cold. Being a viral infection, the antibiotics would be ineffective. Further, prescription of the antibiotics would have a secondary, harmful effect, as their unnecessary prescription would contribute to the development of antimicrobial resistance. The GP is therefore unwilling to prescribe, but fails to meet the patients' hopes and expectations. If this encounter was later rated, the GP would likely be rated poorly, where in fact have had acted in the best interest of the patient, and the wider population as a whole.

This argument rests on the notion that high quality technical care might not be the same as care that provides the patient with a good experience. As Manary and colleagues put it, "Patients are not credible critics- they have not been to medical school" (106) . They also noted that patients' views could be confounded by factors not related to the quality of care

provided, but the environment, for example hotel services. As a consequence, if a member of the public makes a decision about care on the basis of online ratings, they may not be picking the best technical quality of care.

Whether there is a link between patient experience and healthcare quality (measured as effectiveness of care) is an active area of research. The answer is somewhat uncertain in the health services literature. Separate studies at the hospital level in the US by both Jha and Orav have found correlation – not strong but statistically significant – between formal patient satisfaction surveys and technical, objective measures of care quality when using large routine data sources (162,163) – with better experience associated with better care. Also, Couling and colleagues found a relationship between patient experience of inpatient care and readmission rates from US hospitals (164). On the other hand, when examined at the person level, there has been a suggestion that good patient experience may be associated either higher cost and no better, if not worse, outcomes according to a large survey of 51,000 thousand people, linking experience surveys with electronic medical records (165). Similarly, Chang et al. studied the survey responses and clinical care in 276 patients over 65 years old and found no evidence of a link between patient experience and technical quality (166). The literature was reviewed by Doyle and colleagues (167) in 2013 – looking across 55 studies (including many very small studies), and they found evidence of a positive association between patient experience and measures of both clinical effectiveness and safety. However, it is noticeable that a number of recent negative studies were not included, which may have given a positive leaning to their results.

There is also an issue around standardizing patient experience, as different patient groups may have very different subjective experiences of care from the same objective event, driven by socio-cultural expectations. Strech and colleagues in Germany noted that although measures of patient experience are widely reported on rating websites, they are not risk adjusted and therefore should not be used by patients in making decisions about care (168). They also noted that if particular patient experience indicators are correlated with race or socio-economic status – which has been demonstrated elsewhere (169) - physicians might be tempted to shy away from treating certain groups of patients, creating an ethical risk. In a thoughtful analysis piece, Glyn Elwyn and colleagues ask "is what patients say they want the same as good quality care?" (170). They note that meeting the expectations of patients is necessary but not sufficient for high quality care. They suggest that the subjective experience of patients' needs to be respected, and standardized and validated tools need to be developed to assess patient experience.

Taken together, this evidence suggests a degree of uncertainty about whether unadjusted, unsolicited patient experience will be a useful measure of quality of care.

### ***Selection bias***

The nature of who uses online rating and review websites is another potential area of concern about the validity of comments left. All measures of patient experience – apart from those with a mandated and forced function - will by their nature have an element of a selection bias. Those with more extreme views may be more motivated to come forward and voice their opinion- even in the traditional large, random paper - based national surveys (4). Online comments are by their nature substantially unsolicited, and at the discretion of

the public. Therefore the risk of a selection bias altering the results is higher still than normal surveys.

It is possible to conceive that the group of people who go online to rate their care, may well not be representative of the wider public. And if a quality metric is unrepresentative, it might lead to inappropriate information being provided to the public, and inappropriate changes being made to services reflecting the goals of an unrepresentative group of people.

We know in other areas of public service provision, those of higher social economic status – sometimes described as the ‘sharp-elbowed middle classes’ – influence public services to their advantage (171). Similarly, we know that technology based interventions are adopted differently across different socio-economic status groups, possibly as a function of both income, access to technology and education status. Evidence suggests uptake of the NHS Direct service (which provide telephone based access to health), was lower than average in the most deprived groups (172). In addition, as internet access is higher in less deprived groups (173), they may be more likely to rate services and to make decisions based on ratings. This potentially creates access and equity issues for the new technology.

Internet technologies are used differently across different age groups. Younger, more tech savvy communities- for example the ‘*Millennial generation*’ born since 2000, have almost no experience of life before the Internet, and use it as a default mode for accessing any form of information, and see a right to create their own content and comment online as something that should be almost standard (174). We might therefore expect that a comment website would be more likely to be used by the young, wealthy, educated individuals. If usage, and

content generation, is substantially different across these groups, it presents a real problem of equity and the risk of the most benefit occurring for those who with the least need of care. Could this novel phenomenon just be one more device to shore up the reality of Tudor Hart's inverse care law (175)?

### ***Anonymity and the poison pen***

There have been continued voices of resistance against the idea of online rating and review from the medical profession relating to issues of anonymity. In particular, the British Medical Association have stated their concerns that there would be a poison pen mentality, with patients hiding behind anonymous profiles to make harmful, and defamatory comments about doctors and services (176). Dr Richard Vautrey, deputy chairman of the BMA's GP Committee, remarked (177):

*“A message board facility is open to abuse... There is the danger that this could become nothing more than a meaningless popularity contest which leaves patients with an unreliable source of information.”*

There is evidence of a similar professional hospitality globally. RateMDs, a popular physician rating website based in Canada, has been threatened with legal action, and both the Canadian Medical Association and the Canadian Medical Protective Association have demanded the names of critical anonymous posters (178). The limited academic literature includes an account of the experience of a plastic surgeon who had numerous highly critical reviews posted about her online. Her proposed solution to the problem, after she was able to identify the deliberately anonymised user from their other online activity, was to litigate

against the commenter -perhaps reflecting the particular pressures, and sociocultural construct, that falls upon a Californian plastic surgeon (179).

### ***Gaming***

Another argument against online rating and reviews is that there is a risk that the system could easily be 'gamed'. This means that online feedback measures could be artificially manipulated in response to an individual's particular agenda. For example, either inappropriately generating extra positive comments to boost the profile of their own service, or being excessively negative about another service, possibly without making use of the service at all. Further, a provider might be deliberately selective in terms of who they encourage to use a service. For example, if a consultation goes well, you might ask the patient to review you online, but if it goes badly, you do not suggest it. This potentially creates a particular kind of managed selection bias, unique to situation where comment is sought in an unsolicited fashion.

Strech et al. viewed these combined risks as a risk to physician welfare (180) - noting the potential for emotional and psychological burden to physicians, as well as potential financial consequences for physicians receiving poor ratings- whether justified or not. He suggested that there should be a minimum number of ratings before they are published publically, suggesting a figure of 5 - 10. Margaret McCartney, a prominent health activist and GP in the UK has called for the potential risk of physician rating to be explicitly measured (4).

### **2.2.2 Arguments for online rating and reviews**

Despite the opponents, there are also a number of voices that are in favour of online ratings and reviews. The arguments for online rating were put together in a coherent fashion by both Bacon (5) and Lagu (2). Arguments in favour include that doctors can often be poor judges of their patients' satisfaction and experience, feedback changes doctors' performance, this activity increase patient autonomy, and it is already happening, so why not make use of it.

#### ***Doctors as a judge of patient experience***

Doctors are not expert judges of how satisfied patients are with their care. Kinstry and colleagues used a survey methodology to compare doctors' predicted patient satisfaction with actual patient satisfaction after primary care consultations. Patient satisfaction was barely correlated ( $r^2$  values of between 0.07 - 0.13) with doctors' predictions of patient satisfaction or the doctors' own satisfaction (181). Similarly, Shannon and colleagues, working in a critical care environment, found physicians had a higher perception of the quality of care they provided than their patient, and the doctors tended to overestimate their patients views of how satisfied they were (182). As such we can speculate that the comments left online might be a useful tool to encourage doctors to modulate their ideas of how satisfied their patients were.

#### ***Feedback changes performance***

We also know that feedback, if done right, changes doctors' performance. Veloski and colleagues have reviewed the literature, and found that feedback can change physicians' clinical performance when provided systematically over a prolonged period from a credible source (183). For example junior doctor performance (as measured subjectively by the

consultant) has been improved in NHS hospitals using structured instant feedback mechanisms (184). Similarly, Reinders reviewed evidence of physician feedback from real patients in 15 studies across primary and secondary care, and demonstrated improved intended behaviour in all studies and improved actual clinical practice in four of the seven that measured it (185).

### ***Ratings and reviews provide autonomy***

From an ethical perspective, Strech and colleagues (180) note that physician rating websites provide patients with a further root to autonomy - and frames it within the context of increasing health literacy. If this provides a root to better and more understandable health information, there is an innate value, and a potential but unproved value in improving health systems.

### ***It's happening anyway***

The final key argument put forward is perhaps more driven by practicality than evidence. We know that people are already writing reviews online and using the information to make decisions about care, and they will do it whether the medical profession wants them to or not. Shaili Jai noted philosophically in an editorial in the New England Journal of Medicine that whatever his views on the fact, scores about him are available for public inspection, it will clearly continue to be a part his professional life, and he might as well use it to learn about his performance (186). As a consequence, rather than trying to prevent the activity occurring, a focus should be on making sure it capture information in a useful form, both for patients and clinicians.

These various arguments demonstrate that there is no consensus about whether rating healthcare online is a good thing or not. However, a number of the arguments, both for and against, are not deeply rooted in evidence (and where it is available, much of it is fragmented and based on small samples) – leaving scope for further investigation and improved quality of debate on the topic.

## **2.3 Usage of Internet patient reporting websites**

At present, there are rating and review websites in a variety of health systems, operated by different types of providers (both public and private), and with different approaches to capturing and reporting information. In the market driven US health economy, where choice of provider and physician are more established, ratings websites appeared early, and with a distinct focus on providing patients with information to make the choices about their care. The academic literature analysing these websites is therefore limited to studies largely in the US, with a small amount of work in the UK and Germany. In this section I look at this literature to understand what we know about them so far, and build on the early attempt to review the literature on this topic by Emmert and colleagues (187).

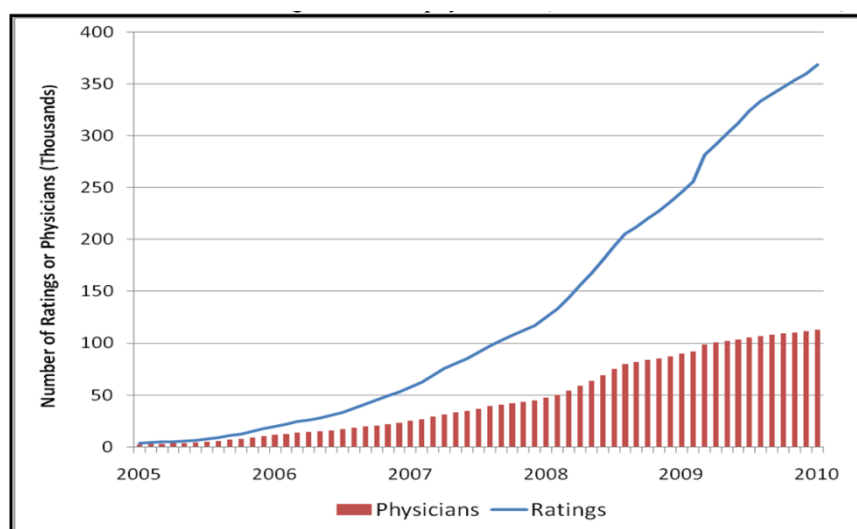
### **2.3.1 Use and coverage of these sites**

Information on how often rating and comment websites are used in healthcare is limited. Most publicly available data at present comes from studies of commercial websites in the US. One of the first attempts to look at this issue was done by Lagu and colleagues at Tufts University in 2010. Using Google to identify physician-rating websites, they looked for reviews of a random sample of 300 physicians in Boston. They found 33 separate sites, containing 190 reviews of 81 (or 27%) of the physicians.

The rate of usage of patients reporting on their experiences of healthcare on the Internet is assumed to be increasing, but there is relatively little public data available to back up this assertion. Actual data on usage is only available when a provider decides to release their

propriety information. A 2012 analysis of a national data set by Gao and colleagues set out to describe the trends in patients' online rating over time, and had privileged access to RateMDs, one of the bigger services (188). They found 386,000 ratings over a five-year period from 2005 to 2010. They were able to do some basic, descriptive epidemiology and found that the number of ratings was increasing, but not in an exponential fashion (Figure 8). They estimated that only 1 in 6 practicing physicians had on online review - a figure lower than seen before - but this is because they used only one website as a source. They found the certain specialties were more like to be rated – citing obstetrics and gynaecology as an example.

**Figure 8: Cumulative number of ratings on the RateMDs website in the US**



Source: Gao et al. 2012 (188)

Elimoottil and colleagues looked at 500 randomly selected urologists in 2012, and analysed how they were described on 10 popular physician rating websites (189). They found that 80% of physicians had a least one rating, and on average there were 2.4 ratings per doctor (range

0 – 64). They highlight the small number of ratings that could potentially lead to volatility of scores, and the possibility of being unduly influenced by a negative review.

### **2.3.2 What proportion of ratings are positive**

We have already seen the concern expressed by clinicians about how patients use the websites, and the risk that there will be a poison pen mentality, with a high number of negative comments. This is one of the simplest issues to analyse from a quantitative perspective. In their early analysis described above, Lagu and colleagues found that 88% of reviews were positive (190).

Kadry and colleagues identified the 10 most frequently used physician rating websites in 2012, and looked at each to see the type of information captured on them, including different questions and rating criteria (3). They found considerable discrepancy in how the information was captured, including measures of 35 different dimensions of care. They found that regardless of the measure used, the ratings tend to be substantially, if not overwhelmingly positive (averaging 77/100, 3.84/5 and 3.1/4, depending on the type of scale used) (3). They found that the patients single overall rating was well correlated ( $r = 0.71$ ) with the other dimensions of care described. On this basis, they suggested that a simple overall rating might have degree of relevance and validity.

Ellimoottil et al. found 86% of urological surgeons had positive ratings, with 36% being highly positive. They found no difference in the frequency of review based on the gender of surgeon, area of the size of the city they worked in (189). Similarly, Black and colleagues

looked at 16700 ratings from a commercial US website, and found that the average score was high (3.7-4.0 out of five) and that ratings were associated with higher staff ratings and scores around punctuality (191).

Gao et al, analysing their large national data set from RateMDs, also found results tend to be positive (3.93 on a five point scale) (188). Using a smaller sample of 3000 rated physicians in Virginia, they also found association between higher ratings and particular demographic details: notably higher ratings for board certified physicians, those from better medical school, and those who had never been sued. They concluded that they were observing a landscape of increasing popularity of rating, and no evidence of being “dominated by disgruntled patients”. In their discussion, they suggested that they had observed evidence of correlation between rating online and quality. However given the limited detail about quality available, restricted to the basic descriptive characteristics of the doctors (such as medical school attended, board certification and malpractice claims) rather than reliable measures of process and outcome, their conclusion is limited. Nether-the-less, it presents an interesting question for further research.

### **2.3.3 Nature of narrative comments**

Some authors have gone beyond simple quantitative analysis of the ratings, and have looked in more details at the narrative comments that are provided about care also with a variety of qualitative methodologies. Tara Lagu performed the first qualitative analysis of what people wrote in their comments. Her team took free-text comments about hospitals in England from NHS Choices, and examined 200 in detail, looking to see what the key issues

patients described were (192). They found the comments were particularly focused on experiences of clinicians and staff, and substantially positive (83%). 62% of the reviews discussed technical aspects of care including issues around medicines management, discharge irregularities and quality of care. More than a quarter of reviews described perceived medical errors. Many of the comments also discussed the softer, hotel side of care, including issues such as food and parking.

Lopez and colleague analysed the free text content of 700 reviews from two popular website, Rate MDs and Yelp in the US. They found 63% of reviews were positive, but noted that many reviews went beyond the doctor patient relationship, and incorporated issues of access, staff, convenience and treatment environment. Comments about particular physicians were more positive, those that looked at the broader focus of the practice were often more critical (193).

A summary table showing the proportion of positive comments is displayed in Table 4.

**Table 4: Proportion of reviews positive**

| Study      | Year | Size of sample                 | Sampled group               | Source                                        | Proportion positive                |
|------------|------|--------------------------------|-----------------------------|-----------------------------------------------|------------------------------------|
| Black      | 2009 | 16,700                         | Physicians in 4 US cities   | RateMDs (US)                                  | Not stated (mean 3.8 out of five)  |
| Lagu       | 2012 | 200                            | Hospitals in England        | NHS Choices (UK)                              | 83%                                |
| Lopez      | 2012 | 700                            | 445 primary care physicians | RateMDs / Yelp (US)                           | 63%                                |
| Kadry      | 2012 | 4,999                          | Physicians in 4 US cities   | 10 most commonly visited rating websites (US) | 58 – 74% (depending on scale used) |
| Lagu       | 2010 | 190                            | 300 Boston physicians       | 33 rating websites (US)                       | 88%                                |
| Elimoottil | 2012 | Not stated, calculated as 1990 | 500 urologists              | 10 most commonly visited rating websites (US) | 86%                                |
| Gao        | 2012 | 386,000                        | National physicians         | RateMDs (US)                                  | 88%                                |

### 2.3.4 Who is using online health rating and review websites

Information about who is using these websites, including their demographic and socioeconomic status, is limited. Awareness of hospital rating and commenting websites in England is also not high. A small 2012 survey by Mirada and colleagues in London showed that amongst members of the public, only 15% of respondents were aware of review and rating websites in healthcare, and only 3% had used one (194). That said caution should be taken with this, as it was a street side convenience survey of 200 people in an inner London

location. Despite these limitations, the survey provides the first real snap shot of who might be using these websites and the potential extent of selection bias.

### **2.3.5 Summary of the existing literature**

The summary of the work to date is a small but growing body of literature describing the activity of rating care online, and the variety of approaches taken. It is very US focused, and largely limited to a small number of commercial rating websites with the highest number of reviews – notably Yelp and RateMDs. The early work described has provided a number of insights into the value of these websites, but also highlights the lack of evidence in some areas and what we need to know more about.

## **2.4 Learning from other industries**

The phenomena of online review and commenting are now commonplace and normal consumer practice in other industries and businesses. In particular, people are used to the idea of reviewing a product they have bought on Amazon, or commenting on the seller they have just purchased something from on eBay. Although there is only a limited published scientific literature in this area, possibly because much of the information is proprietary and commercially sensitive, we can learn lessons from what has been made public. These may be applicable to healthcare.

### **2.4.1 Usage of online ratings and reviews in the travel industry**

In the travel industry, websites such as TripAdvisor are now common. More than 100 million reviews have been left on the TripAdvisor website, up more than 50% in the last year (195). In popular travel locations in the UK, the hotels have an average of 240 reviews. No overall figures are publicly available for commercial purchasing sites such as Amazon, but the rating and review facility is well used. In June 2013, the book *Harry Potter and the Deathly Hallows* had 7430 reviews on the US website (196).

Online reviews are also popular for making purchasing decisions. Some market research survey data is in the public domain. In a study of US shoppers (conducted by a market research company on behalf of a rating company), 79% of purchasers did online research before purchasing for at least half of their shopping. 15% did it for 90% of their shopping

(197). This study also found 59% of customers rated user generated consumer reviews as one of the two most important factors influencing their decisions.

Online reviews are also popular for hotel selection. An electronic survey of 2800 recent travellers about hotel booking by McCarty and colleagues found that more than 35% of leisure travellers used a review or rating website such as TripAdvisor when making a purchasing decision (198). They also found that seeking review information online was more likely from females than males. This survey also suggested that hotel visitors would be as likely to post a review online after a positive experience as they would after a negative experience.

### ***The nature of hotel reviews online***

Some findings in the commercial hospitality sector are similar to those observed in healthcare. Reviews tend to be positive. The average of the 100 million reviews on TripAdvisor is 3.7 stars out of five (195). Looking in more detail, Racheria and colleagues examined 3,200 hotel reviews and found a substantial majority (82%) of positive reviews. They also found difference between review types. Positive reviews tended to be shorter, and focused on value for money. Negative reviews instead focused on staff and management behaviours. They also note the limitations of a single overall rating scale, which may miss some of the nuances of commentary provided in free text (199).

The volume of reviews also has a relationship with the content. One study of the hotel market found a relationship between valence and volume with hotel reviews (200). That is, the first few reviews tended to be disproportionally negative, but as the number of reviews

increase, so did the level of positivity, and possibly becoming a more representative view of the hotel. Another study, looking at 1 star hotel ratings in Washington DC, found that the worst reviews tended to have a focus on front desk staff, bathroom issues, room cleanliness, and guestroom noise issues (201).

The studies in this field also found evidence of certain selection biases. For example, a study of the characteristics of Amazon.com reviewers found they were more likely to be older, male and better educated than the normal population (202).

### ***Do reviews drive traffic for hotels?***

Some researchers claim their economic analyses confirm what had long been expected in the industry: that social media reviews drive hotel reservations. A study looked at the impact of online reviews on pricing and found a link between rating and reviews online and their revenue performance. They found that if a hotel increases its score by 1 point on a 5 point scale – e.g. going from 3.5 to 4.5 out of 5, they can increase price by 11.2% and still keep the same level of activity and market share (203). This could be viewed as evidence that Berwick’s selection mechanism, if applied to hotel choice, does actually occur.

### **2.4.2 Highlighted problems**

Similar concerns about risks of user generated content have been expressed in the commercial sector as have in healthcare (204). As the concept is more mature in these other industries, evidence is also easier to come by. Examples of gaming in this field are perhaps more obvious. Much is anecdotal. For example, an established industry has been described in which people write deliberate and untruthful positive reviews about services online.

Investigation by the New York Times suggested that \$5 is the going rate for leaving a review

from a respectable account for a hotel on the Yelp website in the US (205). Similarly, press coverage reveals a hotel in Cornwall was recently criticized for offering discounts to guests- as long as they provided positive reviews in TripAdvisor (206). This in turn has created a whole subsidiary industry of companies specializing in online reputation management, and whose business model is protecting and improving ratings on TripAdvisor or similar platforms. Examples include Mainstreet Hub (207), which charges single hotels upwards of \$350 a month for the service. Data on the extent of gaming is limited, but the growth of these reorganizations suggests that the commercial market perceives it is real and a need to protect itself.

These problems have even sparked some creative solutions. Computer scientist have developed algorithms to detect unusual or suspicious patterns or reviews, either from the point of view of location, content or frequency (208). Although still in its infancy, it is believed that a number of large commercial companies use software of this, or a similar sort, to maintain the integrity of their reviews.

## **2.5 Gaps in the literature**

Having considered the existing literature in the field, I note that it is limited in scope and volume to simple descriptive analyses. It is a new area and many issues have not yet been studied, in particular how people and organizations respond to this data and how this data relates to other measures. If we are to be able to place these new online patient reported descriptions of care in the broader landscape of quality measures, we will need to know more details about all of these questions.

### **2.5.1 Understanding how patients use this information for selection**

There is a substantial question to be asked about how patients use the information on rating websites to make choices about where to seek care. This question is perhaps best viewed as an extension of a wider discussion around healthcare choice and the effect of public reporting that was discussed in Chapter 1. We know from other studies that the selection pathway proposed by Berwick is not supported empirically (if theoretically sound). Patients do not behave as if they are in a perfect and rational market place, moving as the classical economist might predict, to maximize their own personal utility in light of useful and reliable information about care quality such as mortality ratios. However, we have seen from other industries that online rating does seem to have an effect on selection of consumer products, and is widely acknowledged by consumers as being a substantial driver in their decisions about buying products. In particular, is there some component of online rating and reviewing which might have the capacity to better engage the public in their healthcare choices? If we consider the notion of bounded rationality discussed previously, and that that patients struggle to use or understand conventional publically reported data, we may find that the simple and human messages contained within comments and reviews are more effective than the dry and numerical data often found in report cares. Humans respond to stories and we listen to the views of others who we judge share common experiences and concerns. We can hypothesize that the reader is able to empathize with the comments, and this makes them more likely to drive a decision than the cold, hard logic of a mortality statistic. Could it be possible that user generated content, social media and web based patient feedback is able to revive the selection pathway?

### **2.5.2 Understanding how providers use this information for improvement**

We know little about how providers make use of the information left on NHS Choices, and whether it is a driver of quality improvement activity. As we know from our initial studying of the literature of public reporting (see Chapter 1), the effect of displaying data in public appears to be mediated by spurring organizations to improve, particularly to protect their reputation (19) (7). We are not sure, however, of the micro level causal pathway by which these improvements occur. What in particular causes organizations to respond? How do they go about looking at the information, and processing it through their organizational structure and bureaucracy, and how, if at all, is performance improvement activity started and measured? We are aware of the criticism of the patient survey approach – which many have accused of failing to spark performance activity, possibly because the comments are not sufficiently analysed or publicized (114). Will the open and immediate effect of online commentary alter this and prove a useful tool for quality improvement within organizations?

### **2.5.3 Understanding how valid online comment and ratings are as a quality indicator**

Picking the correct quality indicators for a health system is hard. Since the scale of the ‘quality chasm’ has been recognized, a huge number of indicators, metrics, targets and measures have been proposed, so much so that people have complained of a surplus, and called for parsimony (107). Could these online comments become a useful part of that landscape – capturing an important element of patient experience honestly and quickly?

The literature around indicator selection notes the need to validate indicators against one another (209). We therefore need to understand how online ratings and comments relate to other quality metrics (both technical measures of care and patient satisfaction), to understand how they fit in a landscape of quality measurement.

Although some prior research has shown an association between results on and clinical measures of quality, no previous study has examined the relationship between unsolicited ratings by patients and objective measures of quality. While some patients may be interested in reviewing other patients' ratings as a window into providers' and systems' interpersonal styles and office amenities, the degree to which such ratings are associated with objective clinical outcomes and conventional patient surveys may well determine their place in the landscape of transparency measures. In addition, decisions by policymakers and payers to support and promote such websites might well hinge on whether such associations exist.

## **2.6 Online feedback in healthcare in the UK**

### **2.6.1 Policy history for online feedback of health care**

We have previously discussed the policy literature around transparency of performance of NHS services, and then the stated desire for better feedback mechanisms within the NHS. Here we look briefly at the specific policy documents that support the idea of using the Internet to provide online feedback services in healthcare.

A history of the evolving policy enthusiasm for web-based feedback by patients can be traced in a line of government reports that have promoted the idea. In the UK, the Cabinet Office report *Excellence and Fairness: Achieving World Class Public Services* (2008) (1) recognized that the National Health Service has traditionally paid too little attention to the needs of patients. At the same time, there has been an expressed desire for responsiveness with government systems, allowing them to learn and improve. A subsequent Cabinet Office report, *Public Services: On Your Side* (2009), stated that the government aimed to create more responsive public services, and recommended using the Internet as a mechanism to collect feedback on NHS services.

During the change in administration between the Brown and Cameron governments, these ideas were given strong emphasis in the coalition agreement (2010) (3), the core document describing the plan for the subsequent administration, stating that: “we will enable patients to rate hospitals and doctors according to the quality of care they received”. The government's white paper on health reform published not long after (4) added they would encourage more widespread use of patient experience surveys and real-time feedback.

In May 2012, the English national *Information Strategy for the NHS* (89) made a series of statements that feedback, and online feedback in particular, would have a substantial role in the future NHS. In its preamble it states that it is hoping to create: “a new culture of transparency, where we will be encouraged to provide instant feedback about services, and where information about the quality and performance of health and care services is easily accessible.” Further, in one of their visions statements for patients they state (from a theoretical patient’s viewpoint), “I will be able to leave feedback about my health and care

experiences whenever I use services, in a way which is convenient and feels more like my experience elsewhere on the Internet.”

*Putting patients first*, the NHS England plan for 2013 - 2016, is more specific on this point, reinforcing the notion that “transparency and participation are key to transforming customer service in health and care” and that patients and the public need to be able to be able to offer feedback via the 'latest digital technologies' if they are to 'take control' of their local services (91). This document also states that the government will: “use modern techniques such as social and digital media to supplement other forms of insight, so that we hear views from all sections of society and understand what people are saying about the health services they want to receive.”

Given this prominence of patient experience measurement and feedback on a national agenda, and the proliferating presence of organizations in the market that provide an online patient feedback capability, it seems likely that these approaches are becoming a more permanent, if not universally accepted, part of the UK healthcare landscape.

### **2.6.2 Online rating in the UK**

This policy rhetoric has created a permissive landscape in which the idea of patients rating their care on the Internet has evolved in England. As I will discuss later in the chapter, this has taken the form of a large government-run rating and review service, on the website NHS Choices, but also the development of a number of smaller, privately owned organisations that provide a similar service. These include, but are not limited to PatientOpinion (210) and iWantGreatCare (211), GoodCareGuide and CareOpinion.

### 2.6.3 Stated aims of feedback systems

Despite this sense of urgency that has developed around the need for patient feedback systems, and the focus in policy rhetoric on allowing patients to make informed choices, there remains a diversity of views about what the purpose of feedback systems are – and which of Berwick’s pathways (discussed in Chapter 1) they are trying to affect. There are two potential, and contrasting, uses of patient feedback data: to drive the selection pathway in allowing patients to make better decisions, or driving the improvement pathway by spurring healthcare organizations and professional into action. This has also been described by Hodgkin as a notion of ‘Choice’ (to empower patients to make decisions between services) or of ‘Voice’ (to improve services) (212).

This tension can be seen in the different philosophies and measurement approaches of different websites in the UK. *IWantGreatCare* is, right from the title, about empowering patients to make choices for themselves. Individual clinicians are scored and comparable (Figure 9). *PatientOpinion*, on the other hand, has removed the ability to leave numerical ratings. Their philosophy is clear that the organization is designed to spur on change. The metric they are most interested is how many times a comment leads to an actual change in the system. Rather than present comparative ratings, their main graphical representation is of the feedback loop for each comment, so that the public can see whether change has been made as a result of the comment (Figure 10).

Figure 9: iWantGreatCare (designed to spur selection)

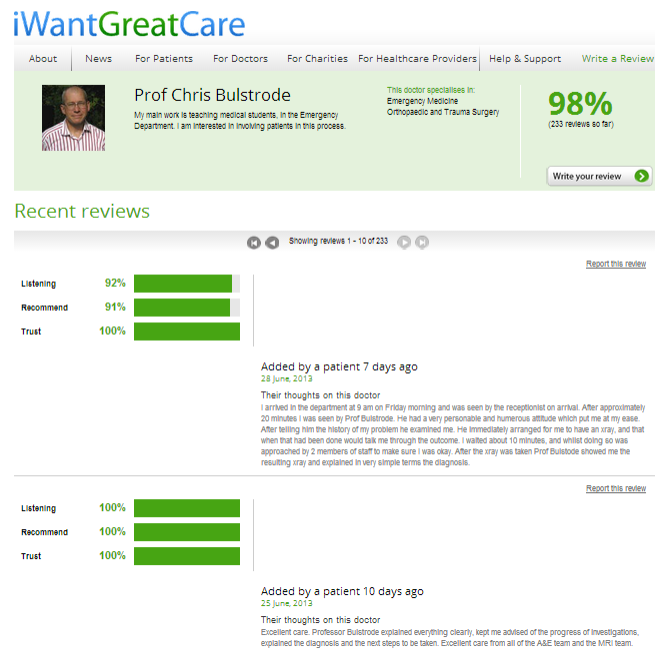
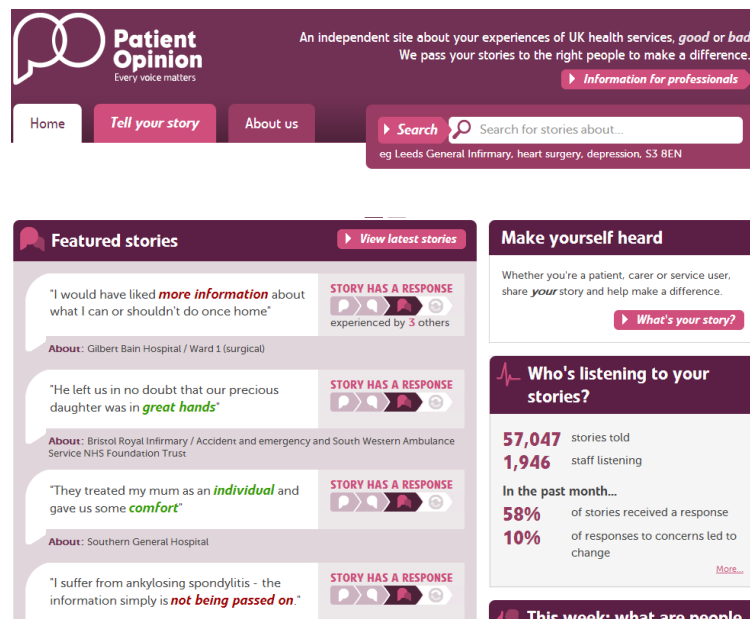


Figure 10: Patient Opinion (designed to spur improvement activity)



## **2.7 An opportunity to study in detail: NHS Choices**

I have highlighted some of the current questions that could be answered about online rating and reviewing of healthcare. In this thesis, I have been presented with an opportunity to answer some of these questions, using a large national system for rating and reviewing.

### **2.7.1 Purpose**

In England, consumer comments and ratings have been collected on NHS hospital services since June 2007. The NHS set up a state run platform, the NHS Choices (NHSC) website (213), to provide a “one stop shop” for health information. At launch, it had three principle aims: to increase health literacy by providing a library of accessible and evidence-based content on a wide range of common treatments, conditions and wellness topics; to encourage better informed health care choices by providing performance data for health services in a useful and understandable “scorecard” for patients; and to increase patient engagement with the NHS by allowing users to comment on and rate NHS services directly. Unlike many other public reporting systems, NHSC has used anecdotal patient comments and ratings as a prominent public indicator since its inception.

Since its launch, NHSC has become the biggest health information site in the UK in terms of traffic, attracting an average of 16.3m unique visitors and 24.3m visits a month between November 2012 and April 2013 (214). This includes 250,000 monthly views of pages that compare performance data of health care providers.

That said, we only have a limited and fragmented understanding of who uses NHSC, who leaves anecdotal comments and how representative the views expressed within them are. A user survey suggests 85% of visits to NHSC are for personal information, and 15% are by healthcare professionals (215). Further, while NHSC is well used, only small proportions (7%) of users come to the website with the specific purpose of either commenting on their care and health service. Most were either looking for medical information (39%) or specific information about health services such as location and contact details (23%) (216).

### **2.7.2 The development and design of the ‘comments’ section of the NHSC website**

The NHSC website currently allows users to both rate and leave comments on all services which deliver care funded by the NHS in England. Services were added in a stepwise approach, starting with hospitals in 2007 and then including primary care practices (known as general practitioners or GPs) in 2009. This has now been extended to dentists, opticians, pharmacies and care homes. The format of information sought from patients has changed over time, but the features now common to the rating forms for all organization types include:

- An overall rating for the organization in question based on a five point Likert scale
- A series of three to five specific questions about the organization, answered by use of a five point rating scale
- A free text area in which more detailed user views are invited to comment on their care

The questions asked relating to hospitals and primary care is shown in Table 5

**Table 5: Questions asked by NHS Choices**

| <b>Question Type</b> | <b>Hospitals</b>                                                                                              | <b>Primary care practices</b>                                                                                   |
|----------------------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Ratings on a scale   | How likely are you to recommend this hospital to friends and family if they needed similar care or treatment? | How likely are you to recommend this GP surgery to friends and family if they needed similar care or treatment? |
|                      | How satisfied are you with the cleanliness of the area you were treated in?                                   | Are you able to get through to the surgery by telephone?                                                        |
|                      | How satisfied are you that the hospital staff worked well together?                                           | Are you able to get an appointment when you want one?                                                           |
|                      | How satisfied are you that you were treated with dignity and respect by the staff?                            | Do the staff treat you with dignity and respect?                                                                |
|                      | How satisfied are you that you were involved in decisions about your care?                                    | Does the surgery involve you in decisions about your care and treatment?                                        |
|                      | How satisfied are you that the hospital provides same-sex accommodation?                                      | This GP practice provides accurate and up to date information on services and opening hours.                    |
| Free text commentary | Give your opinion in your own words. The more detail you can give, the more useful your review will be.       | Give your opinion in your own words. The more detail you can give, the more useful your review will be.         |

The development of the site's anecdotal reporting system has been iterative and driven by user and stakeholder input, as well as the broader web conventions established by sites such as Amazon and TripAdvisor in this area. The system has been in near constant development since the website was launched and could be described as a 'learning system', with aspects of design and functionality frequently altered in response to feedback and evaluation.

Over the last six years, 845,363 patient ratings and 177,854 comments have been collected (see Table 6) – however, as no demographic data is collected when comments are left, very little is known about the population who are leaving these comments and reviews. Growing

site traffic, an extension of the services that consumers are able to comment on and substantial design changes to the comment capture/display interface mean new comments are now being gathered at a rate of about 7000 a month (4,000 for GPs, 2000 for hospitals and 1,000 other). Progress was initially slow with early designs proving less than optimal for easy use and understanding by health consumers, but have increased with redesigns. The iterative nature of the NHSC is reinforced via the current plan to renew and rebrand NHSC as a new national ‘customer insight’ platform planned in 2014, which hopes to build on the existing successes, and is anticipated to continue to give a substantial role to anecdotal reporting (216).

**Table 6: Total comments/rating from July 2007 to March 2013 on NHS Choices**

| Organisation type | Number of Ratings | Number of Comments |
|-------------------|-------------------|--------------------|
| Hospital          | 220288            | 65361              |
| GP                | 490679            | 87051              |
| Dentist           | 129408            | 22357              |
| Optician          | 475               | 502                |
| Pharmacy          | 1961              | 2034               |
| Other             | 2552              | 549                |
| Total             | 845363            | 177854             |

### **2.7.3 How it is regulated**

The commenting part of the website has a strict set of rules (217), developed in order to try to mitigate some of the potential risks and uncertainties of unsolicited information of this

sort. For instance, comments are allowed to report on organizations and services, but not individually named clinicians. This was a specific requirement of the BMA during negotiation processes. There is a detailed set of rules around moderation and what can and cannot be posted. This includes specific rules on: relevance, civility, abusive language, stereotypes, multiple posts, commercial services, giving specific medical advice, the age of posters and English language. All of the comments that are posted online are reviewed by a person to ensure this is the case.

There is also a strict set of policies around anonymity and data confidentiality (218). All people leaving a comment area required to provide an email address, but do not need to register or to provide a real name. Comments are allowed to be posted anonymously, but NHS Choices should always have the contact details of the individual leaving the comment. This means that they can act as a conduit for discussion between the individual and the health service if that becomes necessary.

At the other end of the process, providers are able to respond to comments publicly, and there is encouragement to do so. The providers' ability to respond to comments online has been taken up variably by different organizations around the country. 54% of comments about hospital care have a response from the organization. (219). Some respond directly and personally to all comments, others ignore them, and some simply respond with a generic reply asking them to seek a different route to resolution of the issue (for example by going through the hospitals traditional patient liaison system). I am not aware of comparative data on this point from the US.

Figure 11: Provider view incorporating latest comments, NHS Choices


**St Thomas'**  
 Telephone: 020 7188 7188  
 Address: Westminster Bridge Road, Westminster Bridge Road, London, SE1 7EH  
 Website: <http://www.guysandstthomas.nhs.uk>


[Leave review](#)  
 Based on 55 ratings for this hospital

[Overview](#)
[Departments and services](#)
[Facilities](#)
[Maps, directions and contact details](#)
[Reviews and ratings](#)
[Leave review](#)

### Overview

Guy's and St Thomas' NHS Foundation Trust includes two of London's most famous teaching hospitals, providing patients with excellent treatment and care from two central London sites near Waterloo and London Bridge.

We provide a full range of hospital services for our local communities and have integrated community services in Lambeth and Southwark into the Trust. We also provide specialist services for patients from further afield, including cancer, cardiac, kidney, women's and orthopaedic services, and we are home to the Evelina Children's Hospital.

### Maps, Directions and Contact details

**General contact details**  
[communications@stth.nhs.uk](mailto:communications@stth.nhs.uk)

St Thomas'  
 Westminster Bridge Road  
 Westminster Bridge Road  
 London  
 SE1 7EH

Tel: 020 7188 7188

Website:  
<http://www.guysandstthomas.nhs.uk>

**Contact specific departments and services**



### Latest reviews of this hospital

★★★★★  
**I liked it.**  
 Just amazing. Was knocked off my bike, and dislocated my shoulder. The whole horrible experience was mitigated by the brilliant NHS. I en...  
 14 June 2013

★★★☆☆  
**poor communication and stressed staff**  
 I have recently been into hospital for rare condition. Pheochromocytoma. The ward was clean, cleaners were utterly dedicated & what I saw ...  
 9 June 2013

★★★★★  
**Happy to be treated in st thomas hospital**  
 I have cut my tendon & UDN in my left hand thumb 3 weeks back. I was referred to st thomas hospital by my GP & WMH. My thumb has b...  
 29 May 2013

★★★★★  
**An absolutely excellent hospital**  
 I recently had very severe malaria. I was written off by Homerton Hospital, but thankfully a team from St Thomas' came to take me to thei...  
 29 May 2013

**Some concerns about A&E care**  
 The trainee doctors and nurses were great but the stand in doctor was not. His English was poor, he had no empathy and told me I probably...  
 8 May 2013

[Leave your own review](#) [Read all 132 reviews](#)

Figure 12: High level comparison view, NHS Choices


**choices** Your health, your choices

[Health A-Z](#)
[Live Well](#)
[Care and support](#)
[Health news](#)

[Services near you](#)

### Results for Hospital in SE11 5PQ

Email 

[Narrow search](#) or [start new search](#)

Showing 1-10 of 207 results | Results per page:   | [See results on a map](#)

| Topics                                                                                                                                            | Key Facts | NHS Choices users rating                                | Responding to patient safety alerts                                                                                                           | Recommended by staff                                                                                             | Mortality rate                                                                                                                                                     | MRSA                        | Registration with Care Quality Commission                       |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-----------|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|-----------------------------------------------------------------|
| Sort by                                                                                                                                           | Nearest   |                                                         |                                                                                                                                               |                                                                                                                  |                                                                                                                                                                    |                             |                                                                 |
| <input type="button" value="Update results"/>                                                                                                     |           |                                                         |                                                                                                                                               |                                                                                                                  |                                                                                                                                                                    |                             |                                                                 |
| <b>St Thomas'</b>                                                                                                                                 |           |                                                         |                                                                                                                                               |                                                                                                                  |                                                                                                                                                                    |                             |                                                                 |
| Tel: 020 7188 7188<br>Westminster Bridge Road<br>Westminster Bridge Road<br>London<br>SE1 7EH<br>0.73 miles away   <a href="#">Get directions</a> |           | ★★★★★<br>53 ratings<br><a href="#">Rate it yourself</a> | <br>Good - All alerts signed off where deadline has passed | <br>In the best 20% of trusts | <br>Better than expected in hospital and up to 30 days after discharge (0.8294) | New information coming soon | Licensed<br><a href="#">More information about this licence</a> |
| <b>The Gordon Hospital</b>                                                                                                                        |           |                                                         |                                                                                                                                               |                                                                                                                  |                                                                                                                                                                    |                             |                                                                 |
| Tel: 020 8746 8733<br>Bloomburg Street<br>London<br>SW1V 2RH<br>1.02 miles away   <a href="#">Get directions</a>                                  |           | ★★☆☆☆<br>3 ratings<br><a href="#">Rate it yourself</a>  | <br>Good - All alerts signed off where deadline has passed | n/a<br>Data not available                                                                                        | n/a<br>Data not available                                                                                                                                          | n/a<br>Data not available   | Licensed<br><a href="#">More information about this licence</a> |
| <b>Guy's Hospital</b>                                                                                                                             |           |                                                         |                                                                                                                                               |                                                                                                                  |                                                                                                                                                                    |                             |                                                                 |
| Tel: 020 7188 7188<br>Great Maze Pond<br>London<br>Greater London<br>SE1 9RT<br>1.58 miles away   <a href="#">Get directions</a>                  |           | ★★★★★<br>23 ratings<br><a href="#">Rate it yourself</a> | <br>Good - All alerts signed off where deadline has passed | <br>In the best 20% of trusts | <br>Better than expected in hospital and up to 30 days after discharge (0.8294) | New information coming soon | Licensed<br><a href="#">More information about this licence</a> |

#### **2.7.4 An opportunity for research**

The existent of NHS Choices represents an opportunity for further research into the topic of online feedback. Because it is state run, the data are open and available. Because it is part of a large state run health system with well-established quality reporting systems in primary and secondary care, it presents a large natural experiment to compare online commentary with other measures on a scale not previously possible.

## **2.8 A ‘Cloud of Patient Experience’: Is it possible to collect data from other sources?**

Going beyond an initial question of physician and hospital review sites, we have also seen a rise in other approaches to talking about care online. Information about the quality of healthcare is routinely posted on the Internet by patients, whether that is on Twitter, blogs, or other social media – as well as the rating websites we have described. In a world that has placed a renewed focus on the experience of the patient, why not use this information to monitor the performance of healthcare providers?

### **2.8.1 Social media, healthcare quality and learning from other industries**

In the wake of scandals such as Mid Staffordshire in the English National Health Service (133)– when poor clinical performance was only identified in a post hoc analysis of a mixture of outcomes, process and experience data – we need new approaches to monitor how well healthcare organisations are working. The emergence of social media and novel technology to capture the voice of the patient presents an opportunity to do this.

I hypothesise that looking at patients’ opinions of their healthcare on the Internet in real time could act as an-early warning of poor clinical care. Traditionally patient experience in hospitals has been measured by means of annual, often paper-based, surveys but this is limited to providing data infrequently and is costly. Collection, automated analysis and aggregation of social media content could be done on a daily basis, at low cost and could provide a tool for continuous service monitoring.

Outside of the health sphere, the novel information from social media are being used for a wide variety of analytical and signal detection tasks, sometimes characterized as part of a 'Big Data' revolution (220). Recent examples include attempts at predicting election results (221), the next Hollywood blockbuster (222) or the closing price of the stock market (223). In healthcare, social media data is already used but, to date, has been largely limited to infectious disease surveillance such as the use of Google search activity to detect spikes of flu activity, (224) and more recently, Twitter analysis to monitor disease frequency in cholera and other disease outbreaks (225)(226). This use of digital information on the Internet to try to answer specific health problems has been characterised as 'infodemiology' or 'infoveillance' (227). Application to patient quality and safety issues has been more limited. However, as use of social media becomes an increasing societal norm, it may have considerable potential.

We are aware that more and more people are using the Internet as a platform to describe their care in both the US and UK (188), which we described in Chapter 1. Of the 85% of adults online in the US, 48% look at social networking sites daily, 34% have read about someone else's health experience on a website, and 15% have consulted online reviews of hospitals or other medical facilities (228). At the same time, work has demonstrated that sentiment analysis of content in patients' comments about their care on the Internet is possible (229,230). Sentiment analysis is the process of taking unstructured, often free-text information, and using a computer programme to assess whether the information contained within it is broadly positive or negative about a particular topic. Alemi and colleagues have proposed developing real-time patient satisfaction surveys using these analyses (229). Cambria and others have suggested that patient experience could be 'crowd validated' by

attempting to aggregate the various sources of unstructured information provided by patients online (231)(230).

The use of Twitter or Internet search activity to monitor the spike of a discrete and definite event such as a disease epidemic seems straightforward. The use of similar data to monitor the fundamentally subjective nature of patient experience remains largely unexploited.

However, I think that the growth in social media use, creating what I have characterised as a 'cloud of patient experience', combined with the effectiveness of the sentiment analysis techniques in other fields presents an opportunity to use this information to help understand the quality of health systems. The most immediate and feasible application of this approach would be in flagging potential examples of substantially poor performance—drawing a parallel with the use of surveillance systems for infectious diseases to detect an 'outbreak' or 'epidemic' threshold. It is possible that the consistent appearance of voices of concern about a hospital captured on the Internet, when compared to other peer hospitals, would be an indicator that standards need to be looked at, even if only to confirm that the dissatisfaction recorded is either a statistical blip or not representative of the real care provided.

### **2.8.2 How to capture and use the 'cloud of patient experience'**

To build up this social, or soft, intelligence, one would need to do two technically different and complex tasks: harvesting data, and then processing it into useful information. The harvesting part will involve collecting the free text from as many open sources of comment about healthcare providers as possible: from social networks, from Twitter, discussion fora

and from rating websites. Anywhere where people talk about their experience of care online would become a potential data source. The process would involve identification of appropriate websites, and then pulling relevant information off them on a regular, automated basis using specialised software- a process known in computer science as 'scraping'.

The second component is the data processing part. The patients' free text descriptions of their care would be converted into a social intelligence dataset by analysing the collected statements for sentiment and reliability, transforming them into an aggregated, quantitative measure of experience for each provider. This requires the use of algorithmic processes, known as natural language processing, to extract useful information from the data retrieved, looking for key themes- such as cleanliness - and key emotions such as anger, joy or sadness. The resulting data can be used to rate specific aspects of care at each hospital, such as the hospital environment and the quality of their interactions with staff, providing a score for each domain chosen.

I hypothesize that if these task could be achieved, this information, in comparison with traditional patient surveys, could then be used by health system regulators to identify poor performance, when particular warning thresholds were crossed, and by for managers and clinicians to suggest areas for improvement. In this thesis, we test whether some of these ideas are feasible.

## 2.9 Summary

Like it or not, people are describing their health care experiences on the Internet. From existing evidence – which is still at its early formative stage from a scientific point of view - we see that the act of patients discussing their care online is a real and growing phenomenon. It remains controversial, particularly in terms of how it is viewed by the medical profession, some of who see it as a challenge to their conventional professional autonomy. Despite the substantial professional resistance, there is public and political support. Although not yet well understood, there is a potential for both harm and benefit.

Unlike many policy ideas, this is not a potential policy option with a classic question of whether to implement or not. Implementation has already occurred, either through government or private sector websites. The question is instead how we should respond to it, how to make best use of it and where, if at all, it fits in the complex and cluttered landscape of quality measures.

Despite the high profile support for this approach, there is a lack of rigorous scientific evidence about the usefulness of patient generated feedback of this sort. There is plenty that we do not yet understand. We can draw parallels from other industries, which provided us with insights about where online feedback in health might go in terms of usage. However healthcare is almost always an exceptional case, where normal consumer behaviours rarely apply. We therefore need to look at the issue in more detail and better understand it within the unique context of healthcare. We need to better understand how it is used and by whom, and how it can be improved to maximize the potential benefit. We need better

information about simple, basic epidemiology such as usage rates in different groups and variation with geography. We need to understand how it compares to other quality metrics, both of experience and outcome, and that biases may exist in the data. We need to understand how patients use it to make decisions. We need to think about new techniques that could be used to make the data more usable.

The team behind NHSC has kindly provided me access to their information to analyse and learn from. I also have access to some of the new information being left online on websites such as Twitter, and new techniques from computer science. This will hopefully allow me to start to answer some of the questions about this new phenomenon.

## **2.10 Aims and Objectives**

### **2.10.1 Hypothesis:**

Patients' unsolicited descriptions of their healthcare experiences online represent a valuable resource for the health system, both for providers to understand how to improve services, and for patients to enable them to make decisions about care. They are a potentially useable as a measure of healthcare quality, and may drive people to select higher quality hospitals.

### **2.10.2 Research questions:**

In this PhD, I examine a number of novel questions in this emerging field. At the moment, there is no clear understanding of who is doing the rating, whether this patient generated information is a useful measure of healthcare quality, how this information could be used by providers to improve patient care, and how it is used by other patients in making choices about their healthcare. I will examine each of these questions in turn, through a variety of quantitative methodologies, and making use of new data sources that have not previously been accessible.

### **2.10.3 Aim:**

To examine the usage of patients unsolicited descriptions of their healthcare online and its potential contribution to improving healthcare quality.

#### **2.10.4 Objectives:**

1. To describe national usage patterns of Internet patient feedback in England, in both acute and primary care.
2. To quantify the nature of the relationship, if any, between Internet patient feedback and conventional measures of quality and experience at the healthcare organization level, in both acute and primary care settings.
3. To understand the extent to which the public use internet patient feedback to make decisions about choosing health care providers
4. To explore the potential for using further sources of user generated content and new techniques for their analysis, in particular:
  - a. To measure the effectiveness of sentiment analysis to generate useful information on care quality for patients and healthcare providers
  - b. To consider the amount of information relevant to care quality contained in Tweets about NHS Hospitals

## **Chapter 3: Usage of online comments and reviews on NHS Choices and comparison with other metrics of patient experience and care quality**

### **3.1 Background**

As gathering patient opinions and experiences becomes increasingly important, either because we care about them inherently, because they might drive improvement through transparency, or because we might want to use them in payment policy, it would be useful to compare patient experience captured via traditional surveys with ratings from these new websites. Little is known about how unsolicited ratings by patients left on websites relate to more conventional measures of clinical process and outcome measures. The amount of agreement between ratings and these conventional measures will be important in understanding the usefulness of these ratings in quality measurement.

In the first part of this experiment, I examine two years of unsolicited ratings posted on NHS Choices for all acute hospitals in the England, describing patterns observed and analysing associations with objective clinical outcomes and conventional measures of patient satisfaction obtained through formal surveys. I also compare the relative strength of associations between NHS Choices ratings and clinical outcomes and associations between patient survey measures of experience and clinical outcomes.

In part 2 of this experiment, I perform a similar analysis, but using online ratings and other metrics from general practice. This study seeks to examine usage patterns of patients' ratings of family physicians on the Internet. I describe how frequency and nature of rating

vary with practice and population characteristics, and present comparisons between web-based ratings and conventional measures of patient experience and clinical quality.

## **3.2 Experiment 1: Online feedback for acute hospitals**

### **Methods**

#### **3.2.1 Online rating data**

We obtained aggregate patient ratings for all NHS hospitals in England posted on NHS Choices for 2 years from 1<sup>st</sup> January 2009 to 31<sup>st</sup> December 2010 from the English Department of Health. The NHS Choices website allows patients to indicate whether they would recommend a hospital to a friend (yes/no) and rate hospitals on four specific domains of quality: the cleanliness of the hospital environment; whether they were treated with dignity and respect during their stay; whether they were involved in decisions about their care; and whether the hospital staff worked well together. The proportion of patients recommending a hospital trust was calculated by dividing the number of ratings that recommended the trust by the total number stating a view on recommendation. For ratings on a scale, the mean rating for each hospital trust was calculated.

As organisational performance in the NHS is generally measured at the level of a hospital trust, data were combined to this level. A trust is a single or small group of hospitals in a defined geographical area operated by the same management team. There are 166 acute care trusts in England. Data were combined by creating weighted arithmetic mean values for each score at each trust. Weighting was on the basis of the total number of comments

for each component hospital. If a hospital had no rating for a particular characteristic, it was excluded from the weighted average.

### **3.2.2 Comparison measures**

I selected clinical outcome measures based on three criteria: (1) the measures are commonly used in public reporting (2) there is known to be variation in hospital performance on the measures, and (3) the measures do not relate to a single condition, as web-based patient ratings were not available by disease group. I chose emergency readmission rate within 28 days, Hospital Standardised Mortality Rate (HSMR), standardised mortality rate from five high risk conditions (acute myocardial infarction, cerebrovascular disease, pneumonia, congestive heart failure and fractured neck of femur), standardised mortality rate from low mortality conditions, and standardised mortality rates among surgical inpatients with serious treatable complications<sup>13</sup>. These measures represent the IOM quality definition domains of effectiveness and safety. The latter two measures are Agency for Healthcare Research and Quality Patient Safety Indicators (232). These measures are adjusted for case-mix using a number of variables including sex, age, neighbourhood socioeconomic status, co-morbidities and diagnoses. Data used were for one year (financial year 2009/2010), and obtained from Dr Foster (233), a health intelligence company, and NHS Comparators (234), the NHS's comparative health system performance service. Both of these sources derive their information from the Secondary Uses Service, a national data system that captures standardised information on every NHS hospital inpatient admission in England (235).

We obtained data on Meticillin-Resistant *Staphylococcus aureus* (MRSA) bacteraemia (all cases) (236) and *Clostridium difficile* infection rates (in patients aged over 2 years) (237) per 10,000 bed-days for financial year 2009/2010 from the Health Protection Agency, which runs a compulsory national surveillance system. These infection rates are not adjusted for demographics or case-mix.

Conventional measures of patient experience were taken from the 2009/10 NHS Inpatient Survey (110) (238). This is the yearly national paper-based survey of patients in all NHS hospital trusts conducted by Picker Europe for the Care Quality Commission, England's health regulator. The survey is similar to the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) in the USA (113). It surveys a random sample of patients who have been treated in NHS hospitals for at least one night. 68,594 patients completed the survey between September 2009 and January 2010 (110). The response rate was 52% (239). I selected questions that were closely related to the domains reported on NHS Choices.

### **3.2.3 Statistical analysis**

We performed a descriptive analysis of the quantitative data from NHS Choices, calculating mean and median scores in each category, and measures of dispersion including the inter-quartile range.

We compared patient website ratings with other measures of performance using Spearman's rank correlation coefficient. This conservative non-parametric test was used

because the distribution of data in the patient ratings was not normally distributed according to the Shapiro-Wilk test ( $P < 0.05$  for four out of five NHS Choices variables described). Statistical analysis was conducted in Stata 10. When comparing patient ratings with infection rates, I used all 166 acute hospital trusts in England. When comparing ratings with clinical outcomes and patient experience, a reduced list of 146 hospital trusts was used that excluded trusts that only care for children or specific specialties (e.g. rheumatology or oncology).

To examine the relationship between patient experience and clinical outcomes, I compared both measures of patient experience (survey and online rating) with the three outcome measures using Spearman's rank correlation coefficient. I then compared the strength of the associations using Fisher's z-transformation. Statistical analysis was conducted in Stata 11.

### **3.3 Results**

Between 1<sup>st</sup> January 2009 and 31<sup>st</sup> December 2010, there were 10,274 ratings of hospitals posted on NHS Choices covering all 166 acute-care trusts in England. This represents a rating of 0.035% of hospital admissions (the NHS had 29,118,009 hospital admissions over the same period <sup>24</sup>). The mean and median number of ratings per hospital was 62 and 46, respectively (range 1-290, Figure 2). Of those who offered a view about recommending the hospital (9,349 ratings), 68.0% (6354) would recommend the hospital to a friend. The remaining 925 ratings did not express a view about recommending the hospital, but rated other aspects of the hospital. A summary of the rating results is displayed in Table 7. When

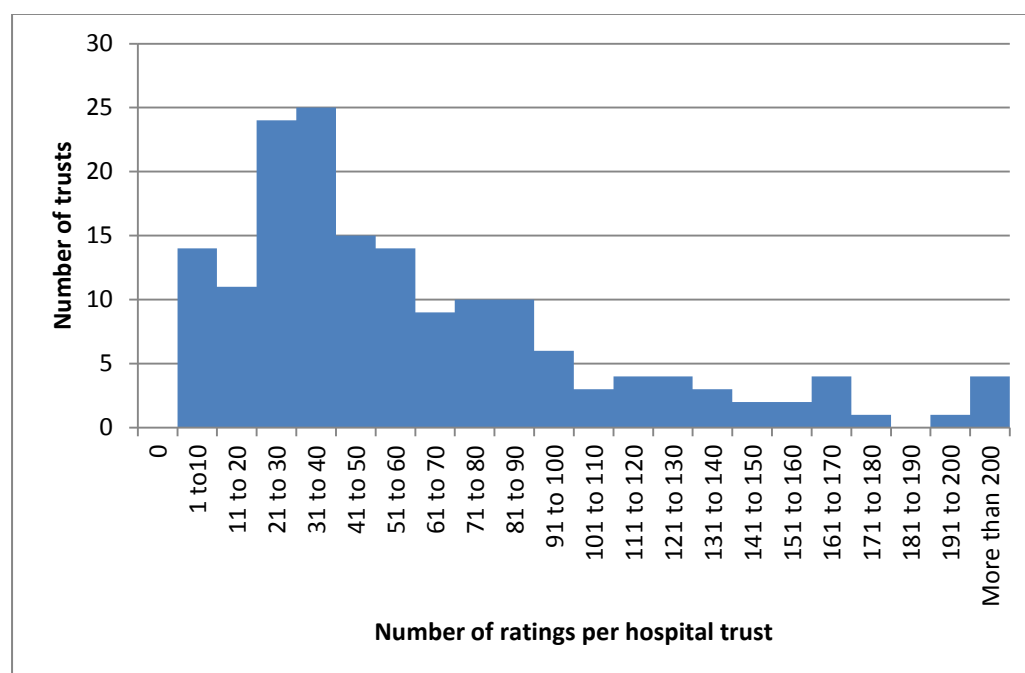
analysis was restricted to the 146 general hospitals trusts (excluding paediatric and specialist hospitals), there were 9,997 ratings and 67.4% recommended the hospital.

**Table 7: Patient rating domains on NHS Choices and scores for 2009/10**

| Score                                                           | 5                   | 4                | 3                | 2              | 1     | Mean rating* | Median * | Range*  | IQR*    |
|-----------------------------------------------------------------|---------------------|------------------|------------------|----------------|-------|--------------|----------|---------|---------|
| The environment where I was treated was...                      | Exceptionally clean | Very clean       | Clean            | Not very clean | Dirty | 3.6          | 3.6      | 2.6-5.0 | 3.4-3.8 |
| I was treated with dignity and respect by the hospital staff... | All of the time     | Most of the time | Some of the time | Hardly at all  | Never | 4.0          | 4.0      | 2.7-5.0 | 3.8-4.3 |
| I was involved with decisions about my care...                  | All of the time     | Most of the time | Some of the time | Hardly at all  | Never | 3.8          | 3.9      | 2.4-5.0 | 3.6-4.1 |
| The hospital staff worked well together...                      | All of the time     | Most of the time | Some of the time | Hardly at all  | Never | 4.1          | 4.1      | 2.9-5.0 | 3.8-4.3 |

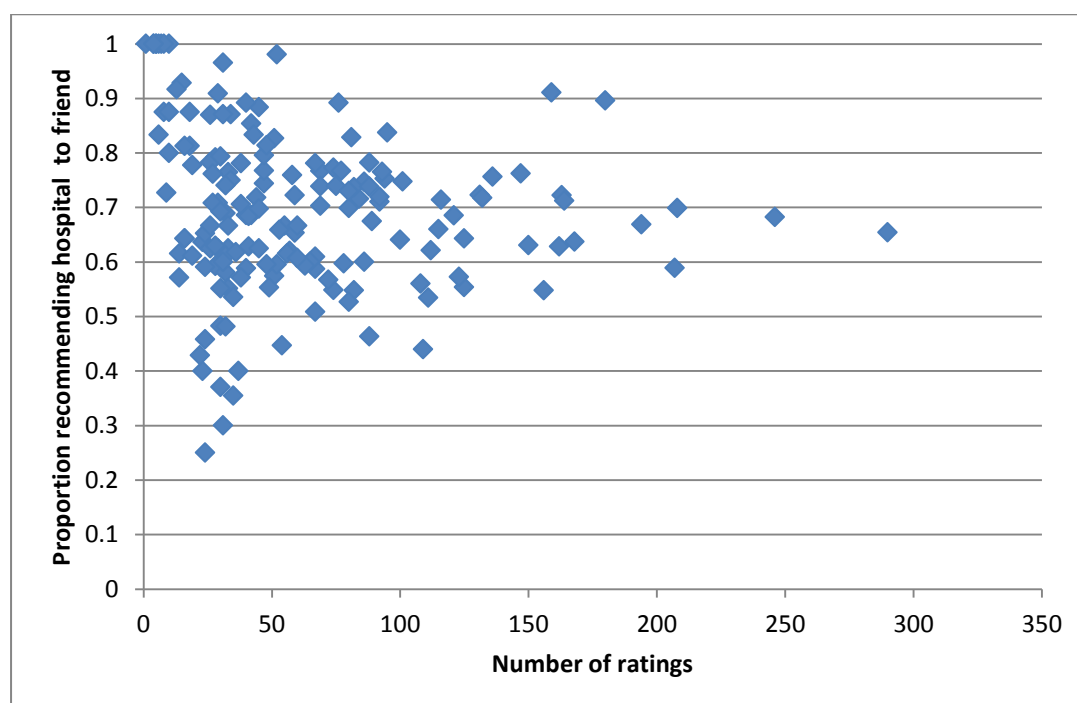
IQR: Interquartile range \*Mean, median and ranges calculated at hospital trust level

**Figure 13: Histogram of frequency of web-based ratings per hospital trust**



A scatterplot, Figure 3, of the proportion recommending their hospital trust against the number of responses for each trust, demonstrates a funnel shape. This suggests no clear association between the number of respondents and the proportion of patients recommending their trust ( $R^2=0.020$ ,  $p=0.07$ ).

**Figure 14: Scatter plot of number of ratings and proportion recommending hospital to a friend**



### 3.3.1 Comparing patient ratings with other measures of quality

Hospital trusts with a higher proportion of web-based patient ratings recommending the hospital to a friend had significantly lower hospital standardised mortality ratios (Spearman  $\rho = -0.20$ ,  $p=0.01$ ), lower mortality from high risk conditions ( $\rho = -0.23$ ,  $p=0.01$ ), and lower readmission rates ( $\rho = -0.31$ ,  $p<0.001$ ) (Table 8). The proportion of patients

recommending a hospital trust to a friend was not associated with mortality rates among surgical inpatients with serious treatable complications ( $\rho = 0.00$ ,  $p = 0.99$ ) or mortality from low risk conditions ( $\rho = 0.03$ ,  $p = 0.70$ ).

**Table 8: Hospital trust level associations between web-based patient ratings and clinical outcomes**

| Web-based patient rating            | Clinical outcome                                                                           | Spearman Rho | p value |
|-------------------------------------|--------------------------------------------------------------------------------------------|--------------|---------|
| Proportion of patients recommending | Hospital Standardized Mortality Ratio                                                      | -0.20        | 0.01    |
| Proportion of patients recommending | Standardized mortality rate for high risk conditions                                       | -0.23        | 0.01    |
| Proportion of patients recommending | Standardized mortality rate among surgical inpatients with serious treatable complications | 0.00         | 0.99    |
| Proportion of patients recommending | Standardized mortality rate from low mortality conditions                                  | 0.03         | 0.70    |
| Proportion of patients recommending | Emergency readmission rate within 28 days                                                  | -0.31        | <0.001  |

Hospital trusts with higher web-based patient ratings on cleanliness had significantly lower MRSA ( $\rho = -0.30$ ,  $p < 0.001$ ) and *C. difficile* ( $\rho = -0.16$ ,  $p = 0.04$ ) infection rates (Table 9).

**Table 9: Hospital trust level associations between web-based patient ratings and healthcare-acquired infection (HCAI) rates**

| Web-based patient rating            | HCAI rates                                                 | Spearman Rho | p value |
|-------------------------------------|------------------------------------------------------------|--------------|---------|
| Cleanliness of hospital environment | Rate of MRSA bacteraemia (per 1,000 bed days)              | -0.30        | <0.001  |
| Cleanliness of hospital environment | Rate of <i>C. difficile</i> infection (per 1,000 bed days) | -0.16        | 0.04    |

Patients website ratings of their experiences on hospital cleanliness, being treated with dignity and respect, staff working well together and being involved with decisions about

care were significantly associated with responses to similar questions in the national patient survey (Spearman Rho =0.31-0.49,  $p < 0.001$  for all) (Table 10). The proportion of people willing to recommend the hospital on NHS Choices was also associated with the overall rating of the hospital in the national inpatient survey (Spearman Rho 0.41,  $p < 0.001$ ).

**Table 10: Hospital trust level associations between web-based patient ratings and paper-based surveys of patient experience**

| <b>Web-based patient rating</b>     | <b>NHS inpatient survey question on patient experience</b>                                  | <b>Spearman Rho</b> | <b>p value</b> |
|-------------------------------------|---------------------------------------------------------------------------------------------|---------------------|----------------|
| Proportion of patients recommending | "Overall, how would you rate the quality of care you received"                              | 0.41                | <0.001         |
| Treated with dignity and respect    | "Overall, did you feel you were treated with dignity and respect while in hospital?"        | 0.34                | <0.001         |
| Staff worked together               | "How well would rate how well the doctors and nurse worked together?"                       | 0.38                | <0.001         |
| Cleanliness of hospital environment | "How clean was the hospital ward or room you were in?"                                      | 0.49                | <0.001         |
| Involved in decisions about care    | "Were you involved as much as you wanted to be in decisions about your care and treatment?" | 0.31                | <0.001         |

### **3.3.2 Comparing different measures of patient experience with clinical outcomes**

For those quality indicators for which an association has been demonstrated between online ratings and patient outcomes (i.e. standardised mortality ratios and readmission rates), I observed no difference in the strength of association between online ratings and outcomes as between patient survey experience measures and outcomes (Table 11).

**Table 11: Comparison of associations between web-based/paper based ratings and clinical outcomes**

| <b>Clinical outcome</b>                              | <b>Spearman Rho for NHS Choices rating and clinical outcome associations</b> | <b>Spearman Rho for NHS Inpatient Survey rating and clinical outcome associations</b> | <b>Z score for comparison of correlation coefficients</b> | <b>p value</b> |
|------------------------------------------------------|------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------|----------------|
| Hospital Standardised Mortality Ratio                | -0.20                                                                        | -0.16                                                                                 | -0.35                                                     | 0.73           |
| Standardised mortality rate for high risk conditions | -0.23                                                                        | -0.07                                                                                 | -1.39                                                     | 0.16           |
| Emergency readmission rate within 28 days            | -0.31                                                                        | -0.25                                                                                 | -0.55                                                     | 0.58           |

## **3.4 Experiment 2: Online feedback for primary care**

### **Method**

#### **3.4.1 Data sources and measures**

All patient ratings of family practices posted on the NHS Choices website between 14<sup>th</sup> October 2009 (the date the function started) and December 31st 2010 were obtained from the Department of Health, aggregated to the practice level. The NHS Choices website allows patients to indicate whether they would recommend a family practice to a friend (yes/no) and rate practices on a scale for four specific domains of quality: whether they were able to get through to the practice by telephone; whether they were involved in decisions about care; whether they were able to get an appointment when they wanted one; and whether they were treated with dignity and respect by staff. Data on the individual characteristics of those leaving ratings online was not available.

I obtained data on clinical quality of family practices from the NHS Information Centre (240), and from NHS Comparators (234), which are both central repositories of NHS process and outcome data. I selected clinical outcome measures based on three criteria: (1) the measures are commonly used in public reporting (2) there is known to be variation in practice performance on the measures, and (3) they represented the breadth of activities done in family practice, and span the IOM's six domains of quality (shown in brackets). The clinical measures chosen were: proportion of patients with diabetes receiving flu vaccination [effectiveness], proportion of hypertensive patients with controlled blood pressure (systolic/diastolic less than 150/90 mm Hg) [effectiveness], proportion of diabetic

patients with controlled HbA1C (less than seven %) [Effectiveness], cervical screening rate [effectiveness], admission rates for ambulatory care sensitive conditions [effectiveness/efficiency], and the proportion of achieved clinical Quality and Outcomes Framework (QOF) points from available points [effectiveness/efficiency/patient centeredness]. The QOF is the NHS's pay-for performance scheme in primary care, which awards points for a variety of clinical activities across acute and chronic disease management and disease prevention. The percentage of low cost statin prescribing (when a patient appropriately received a generic rather than proprietary branded treatment) was used as a measure of efficiency.

Survey measures of patient experience were obtained from the national General Practice (GP) Patient Survey at the practice level, and used to represent the IOM's domains of access/timeliness and patient-centeredness. This is a large mail based survey sent to 5.7 million patients in 2009/10, more than 2.1 million responses were received (a response rate of 37%) [9]. The GP Patient survey is not a representative sample of the English Public, but is a large, weighted survey sent to a sample of the patients at every practice. With more than 2 million responses, it is broadly representative of the UK population, but is weighted to sample higher proportions of young people and men, as these groups tend to have lower survey response rates (241).

The questions from the GP patient survey which we compared against online ratings were: the proportion recommending the GP practice to someone who has moved to the local area [patient centeredness], rating of doctors involving you in decisions about care [patient centeredness], rating of doctor treating you with care and concern [patient centeredness],

ease of getting through on the phone [timeliness] and the proportion able to book ahead for an appointment [timeliness]. No routinely reported measures of patient safety or equity could be identified.

Population and practice characteristics were obtained from the NHS Information Centre. Practice population variables were the proportion of patients over 65 years, the index of multiple deprivation (IMD) score (an area based measure of socio-economic status including components of income, employment, health, education, crime and housing), the proportion of the practice population who reported their ethnicity as 'white', the population density of the practice (measured as people per square kilometre) and the practice list size. Variables describing the practice were whether it was a singlehanded practice, a training practice and the type of contract it had with the healthcare payer. The type of contract was listed as either Personal Medical Services (PMS) or General Medical Services (GMS). The GMS contract is nationally agreed, while PMS is locally agreed. These variables were selected as they have all been demonstrated to be associated with general practice performance in a variety of observational studies looking at general practice performance measured by the Quality and Outcomes Framework pay for performance system (242–244).

### **3.4.2 Data linkage**

In order to create a complete set of family practices in England, I took the NHS's list of 8381 practices [12]. This excluded walk-in centres, out-of-hours services and prison health centres. I then excluded 165 practices with a list size less than 1000 and 127 military

practices. The total number of practices used in the study is 8089 (96.5% of the total number in England).

In order to link data about the practices from NHS Choices with other practice performance and demographic data, I matched practices by postcode using Excel. I created a computer program which extracted postcodes from the NHS Choices website using Python programming language. This match was checked manually by one person to ensure appropriate linking. Where more than one practice is located at the same postcode, they were manually checked to ensure the correct practice was listed using the names of the practice physicians.

We compared the NHS Choices rating data against patient survey and outcomes data from the 2009/10 financial year. I obtained only incomplete practice demographic descriptive data for 163 practices, but these practices were still included in the analysis as I still had information from the patient survey and NHS Choices. 24 GP practices (of 5362 with ratings) on NHS Choices could not be matched to the official list of practices and were excluded from this analysis. The 24 practices that could not be matched were all present in the NHS Choices data, but not in the formal NHS practice list, and as such we are unaware of their patient and practice characteristics. This suggests that they were probably branch surgeries of larger surgeries that are included in the analysis.

### **3.4.3 Statistical Analysis**

We examined associations between whether a family practice was rated and practice and population characteristics using logistic multivariate regression. A least square regression analysis was used to examine associations between the proportion of patients willing to recommend the practice and practice and population characteristics, as I was examining how a dichotomous categorical independent variable relates to a number of categorical and continuous dependent variables. In order to compare ratings with survey measures of patient experience, I compared ratings for questions on NHS Choices with results of the national General Practice Patient Survey using Spearman's rank coefficient of correlation. I selected questions from the national survey that most closely matched the ratings questions on the NHS Choice website. In order to compare ratings with measures of clinical quality, I compared the proportion of people who would recommend the practice on NHS Choices with the traditional quality indicators described above using Spearman's rank coefficient of correlation. For ratings on a scale, the mean rating for each practice was calculated. Analysis was done with Stata 11 software.

### **3.4.4 Mapping**

The level of NHS Choices usage to rate family practices online was mapped by ArcMap 9.3 software, using an Inverse Distance Weighted algorithm. The location of each practice was geocoded [13] and mapped with the corresponding data about ratings usage. Rate of using NHS Choices was measured as the number of ratings or comments divided by practice-registered population, expressed as ratings per 1,000 people. Where multiple practices share a postcode, the mean value of the rate was used

## 3.5 Results

### 3.5.1 Descriptive Statistics

Of the 8089 practices included in the study, 4950 (61%) had been rated on the NHS Choices website. There were 16952 ratings of these family practices. The mean number of ratings of each practice was 2.1, the median was 1. The range was from 0 to 149 ratings. A histogram of frequency of rating is shown in Figure 15. For those practices that had been rated, average ratings at the practice level for each of the questions by the rating website are shown in Table 12. The variation in practice characteristics is shown in Table 13.

**Figure 15: Histogram of rating frequency for all GP practices**

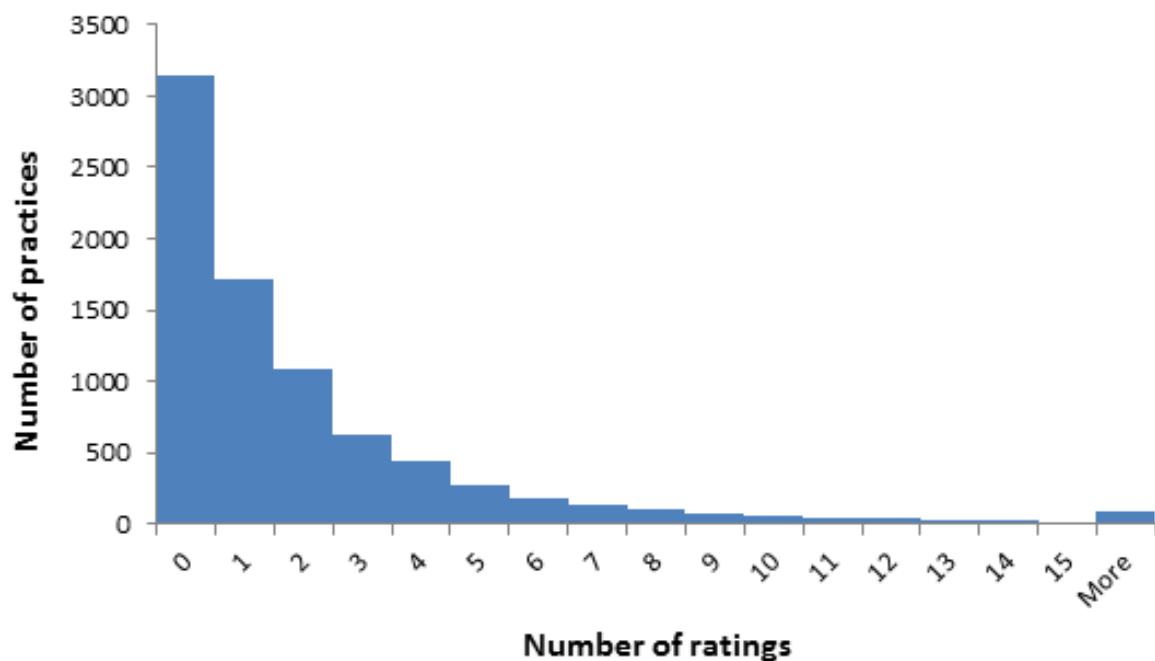


**Table 12: Mean patient ratings posted for questions on NHS Choices**

| Question asked                                                        | Mean rating                                 | Interquartile range |
|-----------------------------------------------------------------------|---------------------------------------------|---------------------|
| I would recommend this GP practice to a friend                        | 64.0 %                                      | 33.3 – 100%         |
| I am able to get through to the practice by telephone                 | 4.2 out of 5<br>(1 is lowest- 5 is highest) | 3.7 – 5.0           |
| This GP practice involves me in decisions about my care and treatment | 4.1 out of 5<br>(1 is lowest- 5 is highest) | 3.5 – 5.0           |
| I am able to get an appointment when I want one                       | 3.6 out of 5<br>(1 is lowest- 5 is highest) | 3.0 – 4.3           |
| I am treated with dignity and respect by the staff                    | 4.1 out of 5<br>(1 is lowest- 5 is highest) | 3.5 – 5.0           |

**Table 13: Descriptive statistics of practice characteristics**

| Variable                                            | Detail                            |               | IQR        |            | Range        |            |
|-----------------------------------------------------|-----------------------------------|---------------|------------|------------|--------------|------------|
|                                                     |                                   |               |            |            |              |            |
| <b>Practice descriptive variables (Continuous)</b>  |                                   | <b>Median</b> | <b>Min</b> | <b>Max</b> | <b>Min</b>   | <b>Max</b> |
| Practice List size                                  | Number of registered patients     | 5891          | 3411       | 9185       | 1000         | 39919      |
| Age over 65 years                                   | (% over 65 years old)             | 15.6          | 11.6       | 18.7       | 0            | 98.8       |
| Ethnicity                                           | (% White ethnicity)               | 96.3          | 85.7       | 98.5       | 4.6          | 100        |
| IMD score                                           | Multi component deprivation score | 22.3          | 12.6       | 37.2       | 0.61         | 82         |
| Population Density                                  | People/km2                        | 367           | 184        | 574        | 0.4          | 3860       |
|                                                     |                                   |               |            |            |              |            |
| <b>Practice descriptive variables (Categorical)</b> |                                   | <b>%</b>      |            |            |              |            |
| Contract type                                       | (% PMS)                           | 40.1          |            |            |              |            |
| Singlehanded practice                               | (% Single handed)                 | 14.5          |            |            |              |            |
| Training practice                                   | (% Training practice)             | 28.3          |            |            |              |            |
|                                                     |                                   |               |            |            |              |            |
| <b>Practice quality metrics (Continuous)</b>        |                                   |               | <b>IQR</b> |            | <b>Range</b> |            |
|                                                     |                                   | <b>Median</b> | <b>Min</b> | <b>Max</b> | <b>Min</b>   | <b>Max</b> |
| QOF Score                                           | Total score/points available      | 55.1          | 92.3       | 97.2       | 68           | 100        |
| Clinical QOF                                        | Clinical score/points available   | 97.9          | 95         | 99.5       | 67.2         | 100        |
| Diabetes HbA1c <7%                                  | % Patients achieving target       | 53.3          | 48.6       | 59.6       | 29.9         | 100        |
| BP < 150/90mmHg                                     | % Patients achieving target       | 79.3          | 75.3       | 83.4       | 61.5         | 100        |
| DM patients with 'flu immunization                  | % Patients vaccinated             | 91.3          | 88.1       | 94.5       | 100          | 72         |
| Recommend to friend                                 | % Who would recommend             | 84.5          | 76.4       | 90         | 32.8         | 100        |
| Involved in decision about care                     | % Very involved or involved       | 71.7          | 65.4       | 76.9       | 29.8         | 96.8       |
| Ease getting through on phone                       | % Easy or very easy               | 71.7          | 60.1       | 80.2       | 16.1         | 98.5       |
| Able to book ahead                                  | % Able or very able               | 72.6          | 61.2       | 82.8       | 11.3         | 100        |
| Low cost statins                                    | % Generic prescribed              | 77            | 71         | 82         | 45           | 100        |

### 3.5.2 Associations between population and practice characteristics and whether a practice is rated

The logistic regression model showed that a larger practice size, a lower proportion of older patients, lower deprivation, higher population density, and not being a single handed practice were all positively associated with the likelihood of a practice being rated (Table 14). Whether the practice was a training practice, the proportion of white patients and the nature of the practice contract did not appear to be associated with being rated. These results remained similar when the non-significant variables were consequentially excluded from the model (full details of the regression analysis are in Appendix 2).

**Table 14: Associations between whether a practice is rated with population and practice characteristics**

| Independent variable                                     | Practices that have been rated on NHS choices | Practices that have not been rated on NHS choices | Odds ratio | Std Error | Z statistic | P value |
|----------------------------------------------------------|-----------------------------------------------|---------------------------------------------------|------------|-----------|-------------|---------|
| Practice population size (Number of registered patients) | 7587                                          | 5554                                              | 1.000      | 0.000     | 15.41       | <0.001  |
| IMD score of patients (higher is more deprived)          | 25.1                                          | 28.2                                              | 0.988      | 0.002     | -7.78       | <0.001  |
| Population density (people/km <sup>2</sup> )             | 458                                           | 403                                               | 1.006      | 0.001     | 6.72        | <0.001  |
| Singlehander (% of practices which are singlehandlers)   | 10                                            | 20.7                                              | 0.054      | 0.054     | -4.61       | <0.001  |
| Percentage of population aged over 65 years              | 15.1                                          | 15.6                                              | 0.113      | 0.063     | -3.94       | <0.001  |
| Percentage of population who are white                   | 87.3                                          | 88.3                                              | 0.763      | 0.130     | -0.017      | 0.11    |
| Type of contract (% with PMS contract)                   | 42.7                                          | 39.6                                              | 0.964      | 0.049     | -0.73       | 0.48    |
| Training practice (% which are training practices)       | 32.3                                          | 22                                                | 1.022      | 0.064     | 0.35        | 0.73    |

### 3.5.3 Associations between population and practice characteristics and the proportion recommending

The least square regression model showed that smaller practice size, a higher proportion of white patients, lower population density, lower deprivation, being a training practice and not being a single handed practice were all positively associated with higher levels of recommendation (15). The type of practice contract and the age distribution of the patients were not associated with different recommendation levels. These results remained similar when the non-significant variables were consequentially excluded from the model (full details of regression analysis in Appendix 3).

**Table 15: Associations between the proportion of patients recommending a practice with population and practice characteristics**

| Independent variable                        | Coefficient | Standard Error | T statistic | p value |
|---------------------------------------------|-------------|----------------|-------------|---------|
| Proportion of population who are white      | 0.305       | 0.041          | 7.39        | <0.001  |
| Training practice                           | 0.088       | 0.013          | 6.61        | <0.001  |
| Practice population size                    | 0.000       | 0.000          | -4.61       | <0.001  |
| Population density                          | -0.001      | 0.000          | -4.16       | <0.001  |
| IMD score of patients                       | -0.001      | 0.000          | -3.93       | <0.001  |
| Singlehander                                | -0.080      | 0.021          | -3.81       | <0.001  |
| Proportion of population aged over 65 years | 0.230       | 0.127          | 1.81        | 0.071   |
| Type of contract                            | 0.007       | 0.012          | 0.64        | 0.52    |

### 3.5.4 Association between ratings and conventional quality metrics

Associations between patient ratings and patient experience measures from the family practice survey had Spearman's  $\rho$  values of between 0.36 and 0.48, and were all significant at the  $p < 0.001$  level (Table 16). A Spearman's  $\rho$  value of 1 represents a perfect rank correlation, 0 represents no correlation, and -1 represents a perfect negative rank correlation. Comparison between patient ratings and clinical quality indicators showed associations between better rating on NHS choices and better quality care for six of the seven variables ( $p < 0.001$ ), however Spearman's  $\rho$  values were all less than  $\pm 0.2$  (Table 17). There is a very weak negative correlation between ratings and low cost statin prescriptions.

**Table 16: Associations between web-based patient ratings and conventional survey measures of patient experience**

| Web-based patient rating                                              | NHS General Practice Patient Survey question on patient experience              | Spearman rho | p value |
|-----------------------------------------------------------------------|---------------------------------------------------------------------------------|--------------|---------|
| I would recommend this GP practice to a friend                        | Recommending GP practice to someone who has moved to the local area- % yes      | 0.48         | <0.001  |
| I am able to get through to the practice by telephone                 | Ease of getting through on the phone- % easy                                    | 0.43         | <0.001  |
| This GP practice involves me in decisions about my care and treatment | Rating of doctor involving you in decisions about your care- % good             | 0.38         | <0.001  |
| I am able to get an appointment when I want one                       | Able to book ahead for an appointment with a doctor in the past 6 months- % yes | 0.37         | <0.001  |
| I am treated with dignity and respect by the staff                    | Rating of doctor treating you with care and concern - Good                      | 0.39         | <0.001  |

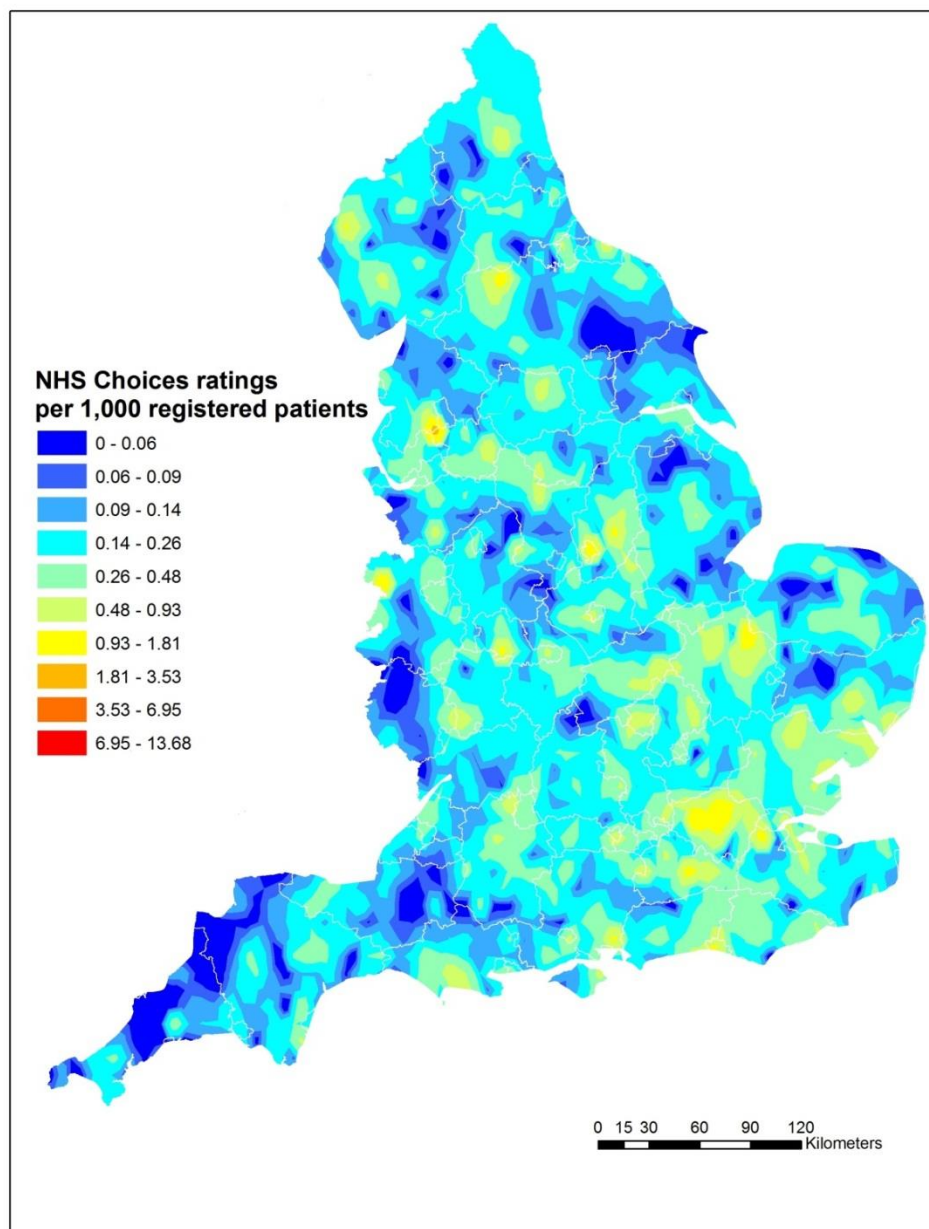
**Table 17: Associations between web-based patient ratings of whether the user would recommend the GP practice to a friend and various clinical quality indicators**

| Quality Indicator                                                                              | Spearman rho | p value |
|------------------------------------------------------------------------------------------------|--------------|---------|
| Proportion of patients with diabetes receiving flu vaccination                                 | 0.07         | <0.001  |
| Controlled blood pressure in hypertensive patients (systolic/diastolic less than 150/90 mm Hg) | 0.07         | <0.001  |
| Controlled HbA1C in patients with diabetes (less than 7%)                                      | 0.06         | <0.001  |
| % low cost statin prescribing                                                                  | -0.03        | 0.02    |
| Cervical screening rate                                                                        | 0.18         | <0.001  |
| Admission rates for ambulatory care sensitive conditions                                       | -0.15        | <0.001  |
| Total clinical QOF points available                                                            | 0.11         | <0.001  |

### 3.5.5 Mapping

Figure 16 illustrates the spatial variation in the use of online ratings by GP practice. The map indicates that in urban areas usage of NHS Choices tends to be higher, particularly around London. Rates are lower in rural areas, the southwest and northeast.

Figure 16: Map of frequency of primary care ratings on the NHS Choices website



## **3.6 Discussion**

Our study provides the first national picture of patient views of health services using data from rating websites. It demonstrates that, given the opportunity, many patients will provide ratings across a broad range of providers and geography, and that most comments about hospitals tend to be positive. Our results also demonstrate a relationship between patients' website ratings of hospitals and a number of commonly used and objective measures of organisational performance, such as mortality rates, infection rates and surveys of patient experience.

### **3.6.1 Proportion of ratings that were positive**

Our results show that despite fears by clinicians and administrators that rating sites would offer mostly criticisms, ratings of hospitals and primary care in England are largely positive. For hospitals, 68% of ratings were positive. For primary care practices this figure was 64%. These results mirror, or are slightly lower than, the earlier findings in other countries (privately-managed physician rating sites in the U.S.) discussed in Chapter 2, where between 60 and 88% are positive (3,161,188,189,191–193).

These results suggest concerns that online feedback mechanisms will only be used by disgruntled patients hoping to complain is not true. However, as the recommendation level online is lower than in the patient survey (where 82% recommend), the results indicate that there may be a selection bias towards less satisfied patients compared to when patients' views are randomly selected. It is also possible that the nature of the NHS Choices website,

funded by the government rather than comparable websites in the US which are privately owned, may create a selection bias towards less satisfied patients.

### **3.6.2 Associations with quality metrics**

#### ***Hospitals:***

Our results demonstrate that patients' website ratings of hospitals and more conventional measures of patient experience from large random surveys are significantly correlated, which is reassuring since it is likely that they are capturing overlapping but different populations of patients and some different dimensions of hospital quality and performance. Our results also demonstrate a relationship between patients' website ratings of hospitals and some objective measures of clinical quality, including mortality and infection rates.

In light of our early findings, Bardach and colleagues performed similar analyses in the US, comparing ratings online on the Yelp website with scores from the HCAHPS system (measuring patient experience) and measures of mortality and readmission rates. They found 25% of hospitals had scores on Yelp. Their correlation coefficient of high rating on Yelp and HCAHPS for patient experience was 0.49, our equivalent figure was 0.41. They also found high scores on Yelp were correlated with lower readmission rates for all measured conditions and lower mortality for MI and pneumonia (245) – and that the correlations were similar strengths. The consistency of these findings with our own, from a different set of data sources and in a different country is reassuring.

I also found no difference in strength of the association between clinical outcomes, such as hospital standardised mortality ratio, and self-selected patient comments left on a website

as with the random survey of patients obtained in the national survey. While traditional surveys have the advantage of random allocation (although they may still suffer from response bias), patients who offer unsolicited opinions about their care may be providing richer feedback. Unsolicited comments may be more likely to pick up low probability but high impact events affecting patient experience, particularly if people who have experience extremes of care, whether good or bad, are more likely to respond to online reporting schemes. These results do not prove that one method is better than the other, but they do help us to understand how they compare. A comparison of some of the characteristics of the two methods of capturing the patient voice are displayed in Table 18.

**Table 18: Comparison of the characteristics of the NHS Inpatient Survey and ratings on the NHS Choices website**

|                            | <b>NHS Inpatient Survey</b>                                                    | <b>NHS Choices ratings</b>                 |
|----------------------------|--------------------------------------------------------------------------------|--------------------------------------------|
| <b>Mechanism</b>           | Paper-based survey                                                             | Ratings left on a website                  |
| <b>Number of responses</b> | 69,000 per year                                                                | 5,000 per year                             |
| <b>Selection</b>           | Random; patients receive a survey requesting completion after leaving hospital | Self-selecting; patients are not solicited |
| <b>Proportion positive</b> | 79% rated their overall care as excellent or very good                         | 67% would recommend to a friend            |
| <b>Cost</b>                | Likely more expensive                                                          | Likely less expensive                      |

***Primary care:***

We found similar results across primary care as in secondary care, although the strength of some of the associations with other quality metrics was not as strong. Moderate associations between patients' ratings left on the web and more conventional surveys of

patient experiences were found (Spearman's  $\rho$  values of between 0.38 and 0.48). Due to the large number of practices in the analysis, these associations are all highly significant.

The association between ratings and clinical outcomes are less convincing. Although the results do show increasing levels of recommendation associated with better care across a number of indicators, for many of them the strength of the association is weak, and significant only because of the large number of data points. This may be because there is a genuine tension between patient experience and some aspects of technical quality; it is possible to do all of the right things technically in healthcare, and still get a bad outcome. It is also possible that patients' personal values may differ from the public health perspective captured in the quality metrics.

These data suggest that I can be more confident in the use of online rating data as a measure of patient experience, despite the many fundamental biases that are an inevitable consequence of this sort of data. However, the extent to which online ratings reflect the technical quality of clinical primary care is less clear.

### **3.6.3 Frequency of rating**

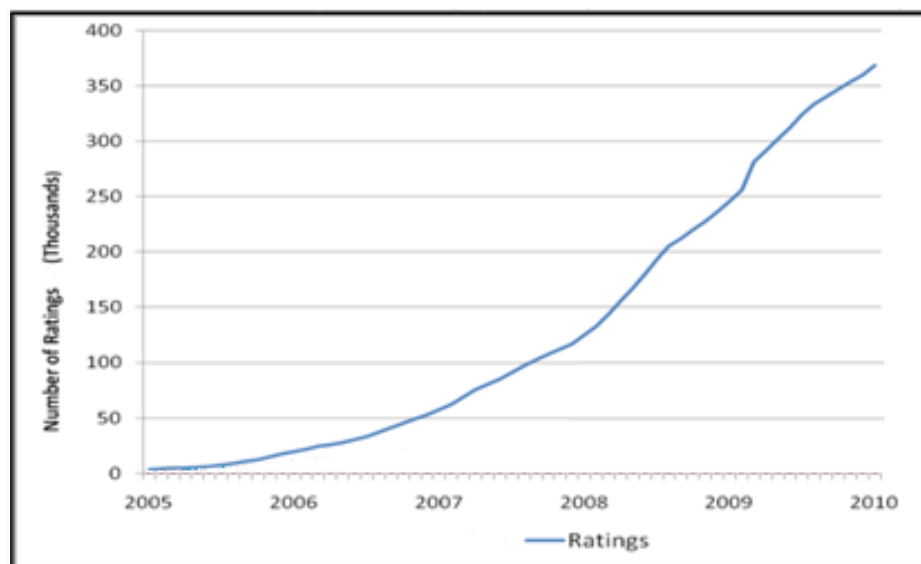
#### ***Hospitals:***

When the NHS launched its NHS Choices website, it was unclear how much the site would be used, both in terms of numbers of responses and geographical distribution. Our results show that every acute hospital trust in England has been rated. However, at present the numbers of ratings is only a fraction of the number of responses to conventional paper-

based surveys, and an even smaller fraction of healthcare contacts (only 0.04% of admission received a rating). There is also variability between different hospitals, which may reflect different levels of awareness of the rating site among patients or of levels of promotion by hospitals.

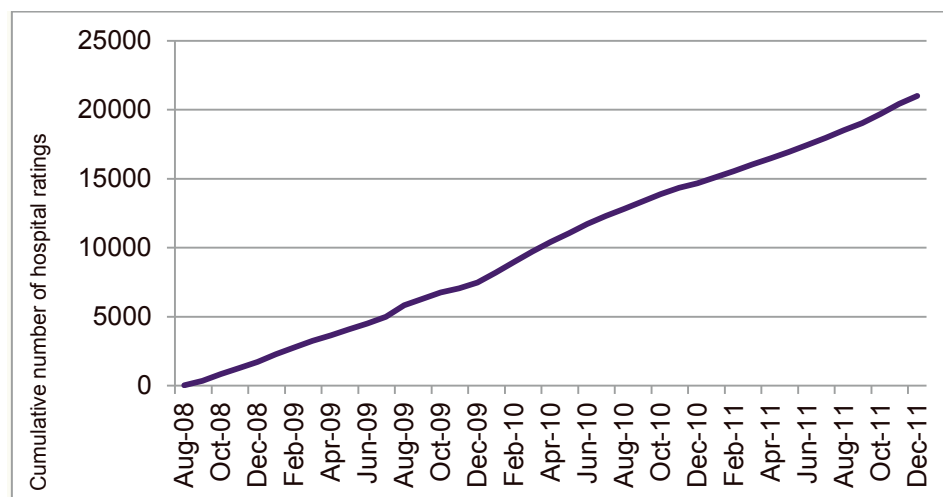
In their 2012 paper, Gao and colleagues demonstrated the growing number of internet-based ratings of physicians on a commercially-owned website in the USA (188). In our analysis I found there were 20,996 ratings of hospitals over the 40-month period from the start of commenting, fewer than in the US. I found a more gradual, linear increase in ratings in England compared with the accelerating growth in ratings seen on commercial sites in the US (Figures 17 and 18).

**Figure 17: Increasing number of ratings in the US on RateMDs - (NB figure is cumulative)**



Adapted from Gao et al 2012

**Figure 18: Increasing number of hospital ratings in England on NHS Choices – (NB figure is cumulative)**



We are not sure why the frequency of ratings is stable in England, but not increasing at the same rate as in the US. This may be because marketing budgets are lower for an English government run service compared to the more commercial advertising approach of US websites, leading to lower awareness of the websites in England. Alternatively, patients in England may be less culturally familiar with the concept of provider choice in healthcare, as the ability to choose between providers has only been introduced relatively recently in the English NHS while it may be a cultural norm in the US. This might result in English patients being less inclined to rate their care. This suggests that the increasing number of online ratings of healthcare is an international phenomenon, even if England is perhaps at an earlier stage on the curve than the US.

### ***Primary care:***

The results demonstrate that usage of patient ratings of family practices via the NHS Choices website has been variable with some practices having more than 100 ratings, but many

having none, and an average of only two ratings per practice over the 15 month period covered. A sizable minority (39%) of practices never received a single rating. The data analysed here represent the initial period of the rating function being available, so usage may well increase as it beds in. However, this usage of online ratings is higher than noted in other primary care settings. A study in Boston, MA found only 81 of a random sample of 300 physicians had been rated online (161), and a study found only 16% of primary care physicians had been rated on the *RateMDs* website in the US (188). When compared to the around 300 million primary care appointments in England each year [16], the number of responses looks rather small (this corresponds to 0.005% of GP consultations being rated online).

The results suggest that the level of usage of online ratings is different in different communities. Practices serving younger, more urban and less deprived communities were more likely to be rated. This is in line with previous work showing more usage of the internet as a health resource in those groups (173). The demonstration of an association between practice population characteristics and rating usage does not prove a link between individual characteristics and usage; there is a risk of the ecological fallacy, in which an incorrect inference can be drawn about individuals based on aggregated statistics about a group of people. However, these results do confirm that usage rates are variable around the country, and suggest that individual characteristics may have a role in influencing usage. Further studies, using individual level data, are required to understand the characteristics of those using ratings websites – which I attempt in the next chapter. The results also show that different practice characteristics are associated with different levels of satisfaction with service, measured as willingness to recommend. This is in line with a wide body of literature

on patient experiences, which notes that ethnicity and different socio-economic factors are associated with satisfaction with medical care (246)(247). These findings may suggest that ratings websites might want to consider ways to broaden their appeal beyond certain groups of users, potentially marketing themselves towards older or rural populations.

### **3.6.4 Limitations**

#### ***Hospital***

In this analysis, I compared patient ratings with a mixture of standardised (e.g. HSMR) and non-standardised performance variables (e.g. infection rates), and observed relationships between both. Although we were able to adjust for some patient level variation, we did not adjust ratings for hospital characteristics, which may have had confounding effects. This makes interpretation more complex.

I also accept that some of the quality measures are controversial, particularly HSMR (248,249). That said, measures of patient mortality are generally of interest to patients and providers, and are reported on many public reporting quality websites, including Medicare's Hospital Compare. In addition, patients may have some awareness of objective measures of hospital quality, and their ratings may be influenced by this knowledge.

I only had access to rating data and outcome data at the hospital trust level, not individual hospital units, so there may have been some loss of information when hospitals were aggregated together into hospital trusts for our analysis. Further, I was unable to examine the characteristics of patients who posted a rating on NHS Choices and the extent to which

ratings differed between age, gender, ethnic and socio-economic status groups. The number of ratings was small for some hospital trusts; however if trusts with less than ten ratings are removed from the analysis, all results apart from the association between *C. difficile* and rating of cleanliness remain significant.

The data from NHS Choices is from the complete years 2009 and 2010. I compared these to available data on quality during financial year 2009/10 – not quite an exact match, potentially limiting our confidence in the nature of the association.

Consistent, national recording of process measures of care, such as use of appropriately timed prophylactic antibiotics before surgery, is limited for hospitals in England, so I was unable to include any in this analysis. Finally, our analysis only reveals correlation and does not imply that positive patient ratings are due to higher quality.

### ***General practice***

As well as the limitations mentioned for the analysis at this hospital level, this study has a number of further limitations in the general practice analysis. I removed practices with a size below 1000 patients, as has been done in other analyses of practice performance (242), as smaller practices are often atypical such as those recently opened or being closed; or serving very specialised populations. In addition, 24 practices were excluded which could not be matched. However, as these exclusions represent less than 2% of practices in our sample this is unlikely to have a major bearing on our findings. In addition, the timing of the ratings and the other outcome and experience measures do not match entirely, but I have chosen the available data with the most overlap.

The use of practice population characteristics leaves some of our findings prone to the ecological fallacy, in which I make an inference about the nature of individuals, based on the characteristics of the aggregated population. In addition, the clinical quality measures used may not reflect true variations in quality between family practices. As with other online rating systems, there is potential for ratings on NHS Choices to be 'gamed' by organisations, or for fake or multiple entries to be left by individuals. I also note that some of the outcome metrics chosen here for quality of care are not entirely in the doctor's hands, but contingent on both the actions of the physician and the patient.

This study does however have some strengths. It uses a novel, largely complete national set of online ratings for primary and secondary care – something never studied before. The unique nature of the NHS's national patient surveys, performance related pay system in primary care and patient level hospital information allow detailed comparisons to be made.

## **Chapter 4: The value of online ratings and reviews in hospital choice: a discrete choice experiment**

### **4.1 Background**

There is little information on how patients incorporate online ratings and reviews by other patients into their decision making processes about choosing health care providers, and how they rate this information compared to other sources. It is possible that the more human voice of other patients' ratings and experiences provided by review websites might prove to be a more understandable and significant driver of patients' decision making processes than traditional public reporting metrics such as mortality (79). In other industries, reviews are common and effect consumer behaviour. The presence of good reviews online has been demonstrated to drive traffic towards hotels, and allow them to charge more and maintain market share (203).

Previous studies have demonstrated the effectiveness of discrete choice experiments (DCEs) in understanding decision making behaviour and public preference. The technique has been widely used in the transport literature, and more recently in the health sector too. It has been used to look at issues including preferred characteristics of physicians and consultations (71) and preferred treatment options, for example between cancer and radiotherapy in oncological cases (250). It has also been used to understand how patients choose hospitals, including the different effect of GP recommendation compared to conventional quality metrics (56) (251), and to look at how hospital choice behaviour varies with markers of socio-economic status (252).

This experiment examines the extent to which patient recommendations of hospitals online on a TripAdvisor style review platform affect people's choice of hospital compared to other measures of hospital performance (e.g. Hospital Standardised Mortality Ratios) or other hospital characteristics (e.g. waiting time). This will give us a better understanding of what factors people include in their decisions about choosing hospitals, and how they weigh them against each other. For example, would they rather go to a hospital with a good reputation or a hospital that is close by? This is important for researchers and policy makers to know, because it will allow for more rational planning of health service supply and demand. It will also be useful to explore whether websites such as NHS Choices are having a discernable effect in providing review data for patients to make decisions. As discussed in the introduction, I am aware that patient's selection of hospitals – one of the possible pathways by which public reporting of data could drive quality improvement – has been found to be limited in effect. Is it possible that online ratings and reviews will have a substantial effect on the choices people make and put more emphasis on the selection pathway?

## **4. 2 Method**

### **4.2.1 Theory**

We used a discrete choice experiment to analyse the value placed on different hospital characteristics by patients when making choices. Discrete choice experiments aim to understand human decision-making processes between different options. The approach assumes that people make rational choices and maximise their utility when making a choice. The process involves presentation of hypothetical options with different attributes, and

asking people to make a choice between them. The term discrete refers to the options being mutually exclusive. The experiment forces people to make a decision, trading off between the relative merits of the different attributes on offer. The discrete choice approach allows the quantification of the effect of the various drivers in the decision making process.

#### **4.2.2 Sample size and recruitment**

We recruited a sample of 2000 members of the English general public. Potential participants were identified from an existing online polling panel database run by a company called Research Now, of a similar sort to that used by YouGov or IpsosMori. The people chosen were selected deliberately to match the make-up of the England in various demographic characteristics such as age and sex using a stratified sample approach. Exclusion criteria included anyone under the age of 18 years. Participants had already signed up to the panel survey process, and regularly take part in Internet surveys.

Participants were emailed directly and asked to take part in the experiment. They were given information about the general nature of the study. They were not told the exact aim – to identify the weight placed on online ratings against other measures – as this may bias their responses. They were asked for consent to take part. Participants who agreed to take part were given a small financial reward as part of a structured reward programme for the research panel. This is redeemable once a certain number of surveys are completed. The value of the reward for this survey was in around £0.75 (for a survey taking 20 minutes).

### **4.2.3 Choice experiment and design**

Participants were asked to complete a series of discrete choice experiments online, given a hypothetical medical referral to hospital for a non-life-threatening condition. For each choice set, the participant was offered three different hospitals, and also the option of not being referred to any of them. The exact nature of the questions was adapted after initial user testing to maximise how easily they can be understood.

Although it is possible to calculate sample size for DCE when the likely influence of choice attributes is known, there is no agreed mechanism for calculating sample size in DCEs of this sort, where baseline studies have not been done. I note similar experiments in healthcare with comparable number of variables had used 500 patient responses (6) (71). However, I also note that some experiments have involved several thousand responses. On the basis of available funding, secured from the Department of Health, I was able to include 2000 respondents. After field-testing for the length of time taken, each participant was asked to complete 6 choice sets- giving 12,000 completed response sets.

In our approach, I built our choice experiment design, attribute categories and demographic questions on the basis of the earlier work on hospital choice measurement by Dixon et al. for the Department of Health in 2010 (56) and by Burge et al from RAND Europe (252,253). I used similar experimental design, including survey questions, phrasing and attribute selection, because these had all been extensively tested previously, and this experiment used a very similar target population. Earlier experiments focused on a broader variety of features and attributes to include in the model. I was fortunate that I could use this earlier work to focus on those factors known to influence decisions. I also added our own novel

variable, not included in previous studies: the presence of reviews online (which could either be positive, negative, or not reported). I took travel and waiting times that were realistic and related to policy initiatives – for example the maximum waiting time was 18 weeks, in line with the national maximum target. The attributes chosen are shown in 19. An example of a choice set is shown in Figure 19. The complete details of the survey and choice process are in Appendix 4.

I used a fractional factorial design to maximise the efficiency of the experiment. As described by Ryan and Gerard (254), this involves using a subset of the possible experimental options in the decision sets, chosen carefully to allow the maximum discrimination of the most important features affecting decisions. An experimental design matrix was generated using the commercially available software Ngene (ChoiceMetrics, 2012). For this study we used a D-efficient design with zero priors and near orthogonal blocking. The final design contained 60 choice scenarios and was divided in 10 blocks.

**Table 19: Attributes included in the discrete choice experiment**

| Attribute                                           | Levels                                 |
|-----------------------------------------------------|----------------------------------------|
| Travel time to hospital                             | 30 minutes                             |
|                                                     | 60 minutes                             |
|                                                     | 90 minutes                             |
|                                                     | 120 minutes                            |
| Waiting time to first appointment                   | 1 week                                 |
|                                                     | 4 weeks                                |
|                                                     | 8 weeks                                |
|                                                     | 16 weeks                               |
| MRSA infection rates                                | Low MRSA rate                          |
|                                                     | Average MRSA rate                      |
|                                                     | High MRSA rate                         |
|                                                     | MRSA rate not known                    |
| Standardized mortality rate                         | Low mortality rate                     |
|                                                     | Average mortality rate                 |
|                                                     | High mortality rate                    |
|                                                     | Mortality rate not known               |
| Whether recommended by GP                           | GP would recommend                     |
|                                                     | GP has no view                         |
|                                                     | GP would not recommend                 |
| Whether recommended by friends or family            | Friends and family have no view        |
|                                                     | Friends and family would recommend     |
|                                                     | Friends and family would not recommend |
| Whether recommended by patients on a review website | Low level of recommendation            |
|                                                     | Average level of recommendation        |
|                                                     | High level of recommendation           |
|                                                     | No online reviews                      |

**Figure 19: An example of a discrete choice question**

| Hospital Attribute                                           |  | Hospital A                                  | Hospital B                      | Hospital C                            | I would chose not to be referred to any of these hospitals |
|--------------------------------------------------------------|--|---------------------------------------------|---------------------------------|---------------------------------------|------------------------------------------------------------|
| Travel time to hospital                                      |  | 30 mins                                     | 60 mins                         | 90 mins                               |                                                            |
| Waiting time to first appointment                            |  | Within 1 week                               | 4 weeks                         | 8 weeks                               |                                                            |
| MRSA infection rates                                         |  | High infection rate                         | Average infection rate          | Low infection rate                    |                                                            |
| Standardized mortality rate                                  |  | High mortality rate                         | Average mortality rate          | Low mortality rate                    |                                                            |
| Recommended by GP                                            |  | GP actively does not recommend              | GP has no view                  | GP recommends                         |                                                            |
| Recommended by friends or family                             |  | Friends or family actively do not recommend | Friends or family have no view  | Friends or family recommend           |                                                            |
| Recommended by patients on a hospital review website         |  | More recommendations than average           | Average level of recommendation | Below average level of recommendation |                                                            |
| Please select one of the following options (please tick one) |  |                                             |                                 |                                       |                                                            |

#### 4.2.4 Analysis

We used a logit model to estimate the importance of different hospital characteristics in making choices between them, which has been used in comparable DCEs (7). This works on the principle that people make choices to maximise their utility (255). A model is built which attaches a coefficient to each attribute level, identifying the extent that it alters decision-making. The model considers a series of utilities associated with particular attribute variables, and that total utility is the sum of these smaller utility components. There is also

an error term that is included to measure unobservable factors. A model estimation process is used to determine the scale of these coefficients.

Two models were made. The first was a basic model with no account taken of differences between individuals (according to the various demographic, socio-economic and attitudinal characteristics captured in the survey). A second model did include these factors, looking for statistical significance between sub groups. In the model development process for the second model, a systematic process was used that considered each of the 24 demographic and behavioural variables, assessing to see whether choice behaviours was different across subgroups. Model creation was done in an iterative process. If the accuracy of the model was improved by incorporating subgroup data, then this was incorporated into the model. The models were estimated using Alogit software and are simple multinomial logit models (MNL). Both models were bootstrapped (using 50 sub-samples) so have corrected standard errors and t-ratios that take into account the fact that the observations are not independent and I have repeated observations from each respondent.

### ***Comparing utility***

Willingness to pay is often used in discrete choices such as this, to provide a comparative measure of the effect of different attributes. This is not a relevant issue in the NHS, where treatment is free at the point of use. Instead, to interpret the relative strength of the different attributes, the coefficient of waiting time was used as a numeraire (or standard unit) to provide a transferable metric of marginal rates, as demonstrated previously by Robertson and Burge (252). Column ordering effect was also assessed to see if there was

any different selection of particular choices based on position (i.e. was there a bias towards picking the first column, regardless of content).

### ***Demographic survey and knowledge and attitudes towards online rating***

In addition to the choice experiments, each participant was asked a number of survey questions about their demographic characteristics, online health behaviours and attitude to choice. These questions included age, gender and a number of measures of socio economic status (educational attainment, household income, and work status calculated using the abbreviated 5-point NS-SEC classification). Participants were asked about their health status (self-defined) and recent health experience and healthcare usage. I also asked about their knowledge about and use of online hospital rating websites, what information they would use to choose a hospital, and what they would use to choose a restaurant. Questions about online health behaviour were adapted from the Pew surveys of internet health activity (228). Details of the specific questions asked can be seen in Appendix 1.

One sample Z-tests for population proportion (two sided) were used to compare the characteristics of the sample to national population data, as the sample size was large (2011) and the population was more than 20,000 times the population. To analyse the survey results I present both the crude results, broken down by subgroups. To test significances in the subgroups I used simple logistic regression. I also carried out an adjusted analysis, where I constructed a larger logistic regression model including age, gender, GP usage, previous experience of care, whether referred recently, locality, education, income and ethnicity.

#### **4.2.5 Ethics**

Ethical approval to conduct this study was obtained from the Imperial College London Ethical Review Committee, because it involved interacting with and gaining opinions of members of the general public. Data were collected and managed to comply with the Data Protection Act 1998. Responses were stored at Research Now on their secure database, and also on secure servers at Imperial College London. Names or identifying data on participants (such as date of birth) were not recorded. Data were not disposed of immediately, to provide the opportunity for transparency of process if others want to review the data.

### **4.3 Results**

We obtained 2011 responses in total. The response rate to the initial email request to complete the survey was 12%.

#### **4.3.1 Survey results**

The nature of the respondents, including breakdowns by a variety of demographic and socioeconomic characteristics are shown in Table 20. In addition, survey participants rated their health on average as 73.8 (mean), (median 80, Interquartile range 68-90) on a scale of 0-100. Table 21 shows the results of the survey questions, unadjusted and broken down by demographic characteristics. Table 22 show the results of the adjusted logistic regression model that incorporated personal characteristics.

**Table 20: Characteristics of the sample**

| Characteristic                                     | Sample |      | English population* | P value |
|----------------------------------------------------|--------|------|---------------------|---------|
|                                                    | n      | %    |                     |         |
| Age                                                |        |      |                     |         |
| 18-24yrs                                           | 309    | 15.4 | 11.8                | <0.001  |
| 25-34yrs                                           | 325    | 16.2 | 17.2                | <0.001  |
| 35-44yrs                                           | 382    | 19   | 17.3                | <0.001  |
| 45-54yrs                                           | 322    | 16   | 17.7                | <0.001  |
| 55-64yrs                                           | 305    | 15.2 | 14.4                | <0.001  |
| 65yrs+                                             | 368    | 18.3 | 20.5                | <0.001  |
| Gender                                             |        |      |                     |         |
| Female                                             | 1,009  | 50.2 | 50.6                | <0.001  |
| Male                                               | 996    | 49.5 | 49.4                | <0.001  |
| Geography                                          |        |      |                     |         |
| North                                              | 523    | 0.3  | 0.28                | <0.001  |
| Midlands                                           | 682    | 0.3  | 0.3                 | 0.4179  |
| South                                              | 806    | 0.4  | 0.42                | <0.001  |
| Ethnicity                                          |        |      |                     |         |
| Any white background                               | 1837   | 91.3 | 87.5                | <0.001  |
| Any mixed background                               | 26     | 1.3  | 1.9                 | <0.001  |
| Any Asian Background                               | 86     | 4.3  | 7.4                 | <0.001  |
| Any black background                               | 32     | 1.6  | 2.9                 | <0.001  |
| Not stated/other                                   | 30     | 1.5  | 0.3                 | <0.001  |
| Employment status                                  |        |      |                     |         |
| In paid work                                       | 1084   | 53.9 | 60.2                | <0.001  |
| Unemployed                                         | 108    | 5.4  | 4.8                 | <0.001  |
| Retired from paid work                             | 443    | 22   | 21                  | <0.001  |
| Unable to work because of disability/ill health    | 59     | 2.9  | 4.1                 | <0.001  |
| Looking after family, home or dependants           | 126    | 6.3  | 4.8                 | <0.001  |
| In full-time education                             | 151    | 7.5  | 4.6                 | <0.001  |
| Education                                          |        |      |                     |         |
| No formal qualifications                           | 111    | 5.5  | 9.5                 | <0.001  |
| GCSE/O level/A level or equivalent                 | 802    | 39.9 | 50                  | <0.001  |
| Professional qualification below degree level      | 308    | 15.3 | 34.2                | <0.001  |
| Degree level qualification or higher or equivalent | 768    | 38.2 |                     |         |
| Location                                           |        |      |                     |         |
| City/large town                                    | 470    | 23.4 | 81.4                | <0.001  |
| Suburbs of city/large town                         | 588    | 29.2 |                     |         |
| Small town                                         | 564    | 28.1 |                     |         |
| Village/ rural area                                | 386    | 19.2 | 18.6                | <0.001  |
| Income                                             |        |      |                     |         |
| Less than £9,999                                   | 159    | 7.9  | N/a                 | N/a     |
| £10,000 – £19,999                                  | 394    | 19.6 | N/a                 |         |
| £20,000 – £29,999                                  | 399    | 19.8 | N/a                 |         |
| £30,000 – £39,999                                  | 325    | 16.2 | N/a                 |         |
| £40,000 – £49,999                                  | 230    | 11.4 | N/a                 |         |
| £50,000 – £74,999                                  | 183    | 9.1  | N/a                 |         |
| £75,000 or more                                    | 106    | 5.3  | N/a                 |         |
| Prefer not to say                                  | 215    | 10.7 | N/a                 |         |

\*Source: Office for National Statistics figures for ethnicity (256), age (257), education employment (258), and location (259)

**Table 21: Survey results on attitudes towards choice and online health behaviours, including breakdown by demographic and socio-economic subgroups (unadjusted)**

| Patient Characteristics                                                  |                       | In favour of choice (%) | Aware of choice of hospital (%) | Looked online for health info (%) | Aware of health rating website (%) | Looked at health rating website (%) | Comment on health rating website (%) |
|--------------------------------------------------------------------------|-----------------------|-------------------------|---------------------------------|-----------------------------------|------------------------------------|-------------------------------------|--------------------------------------|
|                                                                          | Overall               | 86.6                    | 62.6                            | 67.0                              | 29.5                               | 12.5                                | 5.4                                  |
| Age                                                                      | 18-34                 | 88.2 ref                | 57.1 ref                        | 70.9 ref                          | 36.4 ref                           | 20.8 ref                            | 10.6 ref                             |
|                                                                          | 35-54                 | 84.8 .                  | 57.8 .                          | 64.8 *                            | 25.5 **                            | 10.2 ***                            | 3.6 ***                              |
|                                                                          | 65+                   | 86.8 .                  | 72.8 ***                        | 65.7 .                            | 27.1 **                            | 7.2 ***                             | 2.5 ***                              |
| Location                                                                 | Urban or suburban     | 86.6 ref                | 62.9 ref                        | 66.4 ref                          | 30.0 ref                           | 14.5 ref                            | 7.0 ref                              |
|                                                                          | Small town or rural   | 86.5 .                  | 62.5 .                          | 67.6 .                            | 28.9 .                             | 10.2 ***                            | 3.6 *                                |
| GP visits in last year                                                   | No visits             | 79.6 ref                | 54.8 ref                        | 47.3 ref                          | 27.6 ref                           | 7.5 ref                             | 3.7 ref                              |
|                                                                          | 1 - 5 visits          | 87.2 *                  | 63.0 **                         | 68.7 ***                          | 29.5 .                             | 12.7 ***                            | 5.6 .                                |
|                                                                          | More than 5 visits    | 91.4 *                  | 70.1 ***                        | 81.9 ***                          | 31.9 .                             | 17.3 ***                            | 6.5 .                                |
| Education                                                                | Secondary or less     | 88.4 ref                | 59.0 ref                        | 62.4 ref                          | 25.4 ref                           | 9.9 ref                             | 4.2 ref                              |
|                                                                          | Higher than secondary | 85.0 *                  | 65.9 **                         | 70.4 ***                          | 33.1 ***                           | 14.5 ***                            | 6.5 *                                |
| Ethnic                                                                   | Non white             | 90.7 ref                | 60.9 ref                        | 76.4 ref                          | 38.5 ref                           | 28.9 ref                            | 14.7 ref                             |
|                                                                          | White                 | 86.1 *                  | 62.7 .                          | 66.4 *                            | 28.6 *                             | 11.1 ***                            | 4.7 ***                              |
| Gender                                                                   | Female                | 89.9 ref                | 63.7 ref                        | 74.5 ref                          | 29.1 ref                           | 14.1 ref                            | 4.0 ref                              |
|                                                                          | Male                  | 83.2 ***                | 61.5 .                          | 59.4 ***                          | 29.8 .                             | 10.7 *                              | 6.7 ***                              |
| Health status                                                            | Below Median          | 86.7 ref                | 64.0 ref                        | 70.0 ref                          | 27.8 ref                           | 13.8 ref                            | 6.2 ref                              |
|                                                                          | Above Median          | 86.4 .                  | 61.3 .                          | 64.0 *                            | 31.0 .                             | 11.3 .                              | 4.6 .                                |
| Income                                                                   | Less than £9,999      | 85.0 ref                | 56.0 ref                        | 61.0 ref                          | 25.7 ref                           | 12.3 ref                            | 9.1 ref                              |
|                                                                          | £10,000 – £19,999     | 87.1 .                  | 62.9 .                          | 65.9 .                            | 28.1 .                             | 11.9 .                              | 5.7 .                                |
|                                                                          | £20,000 – £29,999     | 88.2 .                  | 62.7 .                          | 68.3 .                            | 28.9 .                             | 11.3 .                              | 4.0 *                                |
|                                                                          | £30,000 – £39,999     | 85.1 .                  | 60.0 .                          | 69.3 .                            | 28.8 .                             | 13.9 .                              | 5.6 .                                |
|                                                                          | £40,000 – £49,999     | 83.8 .                  | 63.9 *                          | 70.2 *                            | 27.8 .                             | 11.0 .                              | 6.6 .                                |
|                                                                          | £50,000 – £74,999     | 85.1 .                  | 60.1 .                          | 65.0 .                            | 34.3 .                             | 13.1 .                              | 3.8 .                                |
|                                                                          | Over £75,000          | 87.6                    | 69.8                            | 74.3                              | 32.7 .                             | 15.2 .                              | 7.6 .                                |
| Past experience of healthcare<br><br>Quality of Past experience (if any) | Yes                   | 86.6 ref                | 63.6 ref                        | 68.3 ref                          | 30.0 ref                           | 12.6 ref                            | 5.7 ref                              |
|                                                                          | No                    | 85.4 .                  | 48.5 ***                        | 47.7 ***                          | 23.1 .                             | 9.9 .                               | 0.8 *                                |
|                                                                          | Generally good        | 84.1 ref                | 65.7 ref                        | 67.0 ref                          | 30.4 ref                           | 10.6 ref                            | 5.5 ref                              |
|                                                                          | Mixed                 | 90.0 ***                | 60.1 *                          | 70.5 .                            | 30.6 .                             | 16.3 **                             | 5.8 .                                |
|                                                                          | Generally poor        | 95.9 ***                | 59.7 .                          | 71.5 .                            | 23.1 .                             | 15.7 .                              | 7.4 .                                |
| Recent referred                                                          | Referred to hospital  | 87.1 ref                | 66.0 ref                        | 77.2 ref                          | 31.6 ref                           | 16.2 ref                            | 7.0 ref                              |
|                                                                          | Not referred          | 86.0 0                  | 60.2 **                         | 59.0 ***                          | 27.8 .                             | 9.6 *                               | 3.9 ***                              |
| Transport                                                                | Walk or bike          | 81.6 ref                | 64.5 ref                        | 67.3 ref                          | 37.5 ref                           | 20.0 ref                            | 15.4 ref                             |
|                                                                          | Car or Taxi           | 87.6 .                  | 63.3 .                          | 66.8 .                            | 27.8 *                             | 9.8 **                              | 4.1 ***                              |
|                                                                          | Public Transport      | 84.5 .                  | 59.6 .                          | 68.5 .                            | 33.4 .                             | 19.7 .                              | 6.7 ***                              |

\*P<0.05 \*\*P<0.01 \*\*\* P<0.001, in unadjusted logistic regression

Table 22: Adjusted analysis of survey results

| Question asked              | Gender |      |         | Age   |       |         |      |         | GP visits |        |         |      |         | Past experience of local hospital |      |         | Type of past experience (if any) |       |         |      |         |     | Recently referred to Hospital |         |
|-----------------------------|--------|------|---------|-------|-------|---------|------|---------|-----------|--------|---------|------|---------|-----------------------------------|------|---------|----------------------------------|-------|---------|------|---------|-----|-------------------------------|---------|
|                             | Female | Male |         | 18-34 | 35-54 |         | 65+  |         | None      | 1 to 5 |         | 5+   |         | None                              | Some |         | Good                             | Mixed |         | Poor |         | Yes | No                            |         |
|                             |        | OR   | P value |       | OR    | P value | OR   | P value |           | OR     | P value | OR   | P value |                                   | OR   | P value |                                  | OR    | P value | OR   | P value |     | OR                            | P value |
| In favour of choice         | ref    | 0.63 | 0.003   | ref   | 0.71  | 0.08    | 0.82 | 0.32    | ref       | 1.86   | 0.001   | 2.36 | 0.006   | ref                               | 0.92 | 0.78    | ref                              | 1.75  | 0.002   | 4.5  | 0.004   | ref | 0.944                         | 0.72    |
| Aware of choice             | ref    | 0.94 | 0.25    | ref   | 1     | 0.94    | 1.88 | <0.001  | ref       | 1.07   | 0.63    | 1.69 | 0.017   | ref                               | 1.67 | 0.01    | ref                              | 0.76  | 0.021   | 0.71 | 0.11    | ref | 0.9                           | 0.186   |
| Looked online for health    | ref    | 0.47 | <0.001  | ref   | 0.68  | 0.01    | 0.77 | 0.087   | ref       | 1.56   | 0.004   | 2.5  | <0.001  | ref                               | 2.07 | 0.001   | ref                              | 6     | 0.068   | 1.05 | 0.88    | ref | 0.46                          | <0.001  |
| Aware of rating sites       | ref    | 1.17 | 0.15    | ref   | 0.5   | <0.001  | 0.55 | <0.001  | ref       | 1.03   | 0.87    | 1.26 | 0.3     | ref                               | 1.66 | 0.05    | ref                              | 0.9   | 0.41    | 0.66 | 0.1     | ref | 0.85                          | 0.17    |
| Looked at rating websites   | ref    | 0.89 | 0.46    | ref   | 0.36  | <0.001  | 0.29 | <0.001  | ref       | 1.76   | 0.06    | 2.29 | 0.019   | ref                               | 2.12 | 0.09    | ref                              | 1.32  | 0.091   | 1.32 | 0.35    | ref | 0.57                          | <0.001  |
| Commented on rating website | ref    | 1.95 | 0.003   | ref   | 0.33  | <0.001  | 0.19 | <0.001  | ref       | 1.54   | 0.27    | 1.88 | 0.19    | ref                               | n/a  | n/a     | ref                              | 0.77  | 0.32    | 0.86 | 0.74    | ref | 0.59                          | 0.03    |

| Question asked              | Locality |                   |         | Education |        |         | Ethnicity |       |         | Income           |                   |         |                   |         |                   |         |                   |         |                   |         |                 |         |
|-----------------------------|----------|-------------------|---------|-----------|--------|---------|-----------|-------|---------|------------------|-------------------|---------|-------------------|---------|-------------------|---------|-------------------|---------|-------------------|---------|-----------------|---------|
|                             | Urban    | Small town /rural |         | Low er    | Higher |         | Non White | White |         | Less than £9,999 | £10,000 – £19,999 |         | £20,000 – £29,999 |         | £30,000 – £39,999 |         | £40,000 – £49,999 |         | £50,000 – £74,999 |         | £75,000 or more |         |
|                             |          | OR                | P value |           | OR     | P value |           | OR    | P value |                  | OR                | P value | OR                | P value | OR                | P value | OR                | P value | OR                | P value | OR              | P value |
| In favour of choice         | ref      | 0.99              | 0.954   | ref       | 0.75   | 0.05    | ref       | 0.43  | 0.042   | ref              | 1.06              | 0.83    | 1.14              | 0.64    | 0.95              | 0.87    | 0.9               | 0.75    | 1.01              | 0.97    | 1.23            | 0.61    |
| Aware of choice             | ref      | 0.85              | 0.12    | ref       | 1.52   | <0.001  | ref       | 0.88  | 0.56    | ref              | 1.17              | 0.45    | 1.1               | 0.65    | 1.05              | 0.8     | 1.36              | 0.19    | 1.04              | 0.87    | 1.64            | 0.09    |
| Looked online for health    | ref      | 1.09              | 0.41    | ref       | 1.52   | <0.001  | ref       | 0.76  | 0.29    | ref              | 1.17              | 0.47    | 1.48              | 0.08    | 1.48              | 0.08    | 1.66              | 0.041   | 1.24              | 0.4     | 2.38            | 0.007   |
| Aware of rating sites       | ref      | 1.02              | 0.84    | ref       | 1.4    | 0.004   | ref       | 0.77  | 0.23    | ref              | 1.35              | 0.22    | 1.33              | 0.28    | 1.42              | 0.14    | 1.3               | 0.31    | 1.66              | 0.05    | 1.61            | 0.12    |
| Looked at rating websites   | ref      | 0.86              | 0.39    | ref       | 1.42   | 0.03    | ref       | 0.47  | 0.003   | ref              | 1.17              | 0.62    | 1.05              | 0.87    | 1.44              | 0.26    | 1.11              | 0.76    | 1.16              | 0.68    | 1.4             | 0.43    |
| Commented on rating website | ref      | 0.72              | 0.17    | ref       | 1.59   | 0.04    | ref       | 0.49  | 0.025   | ref              | 0.77              | 0.51    | 0.44              | 0.05    | 0.69              | 0.34    | 0.88              | 0.76    | 0.42              | 0.08    | 0.78            | 0.65    |

Figures 20 and 21 show the various information sources people would use to book a hotel and a hospital respectively (when presented with a list of options).

**Figure 20: Information people would use to choose a hospital**

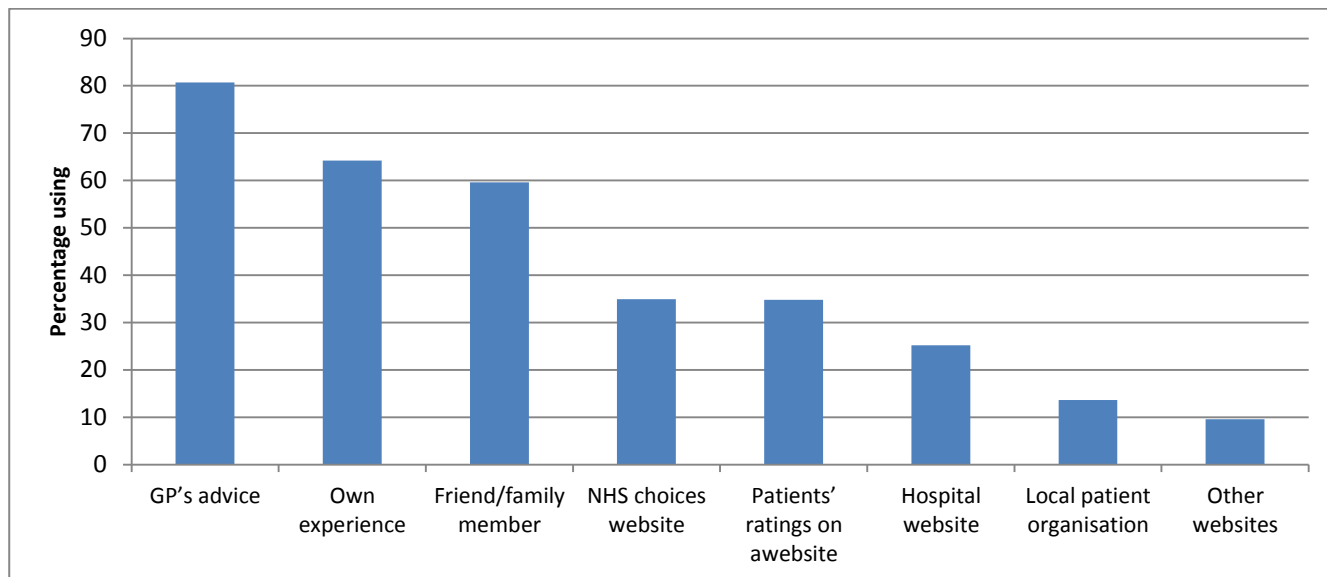
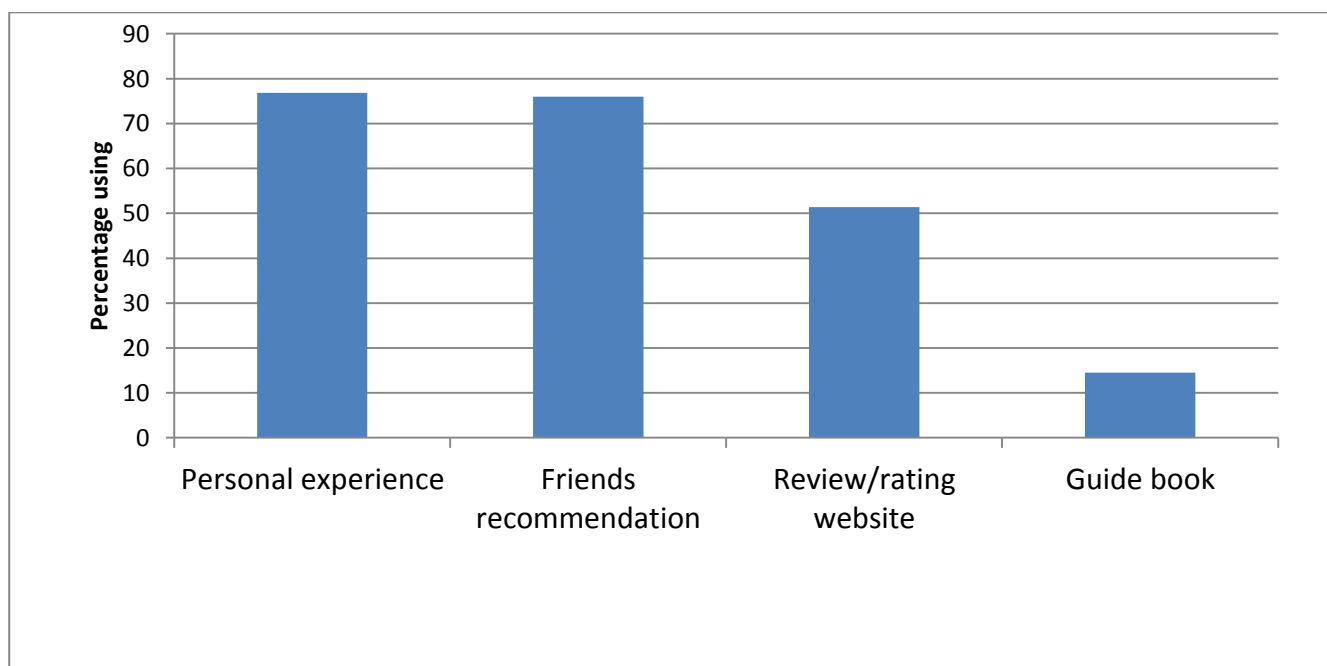


**Figure 21: Information people would use to book a hotel**



Further analysis of these results showed that those who used a TripAdvisor style rating to pick a hotel were more likely to have consulted a patient experience rating website (OR 1.3 in logistic regression,  $p=0.046$ ). This was no longer significant when adjusted for age, sex, gender, income and employment – with the majority of the effect mediated by age.

#### **4.3.2 Discrete Choice Experiments**

The effect of attributes on choices is expressed via coefficients of utility. A positive coefficient means the person is more likely to choose a hospital with this characteristic. A negative coefficient means they are less likely to. The size of the coefficient determines the effect, with large size (either positive or negative) meaning larger effect. Both waiting time for and travel time to the hospital appointment exhibited linear relationships with utility, as demonstrated in Figures 22 and 23. This means that it is possible to use either waiting time or travel time for a proxy measure of utility to the value of other attributes. For the purpose of this analysis, I have expressed utility in the both the form of 'weeks willing to wait for treatment' and 'hours willing to travel'.

**Figure 22: Relationship between utility and waiting time demonstrating a linear function**

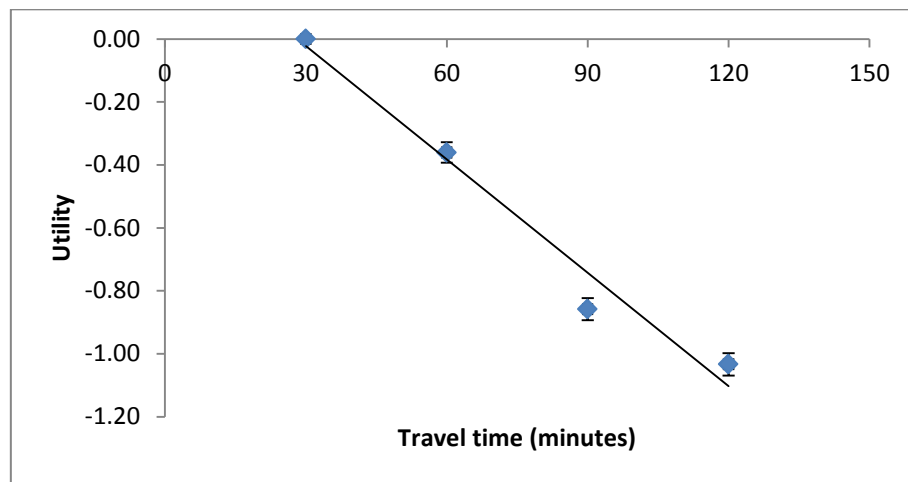
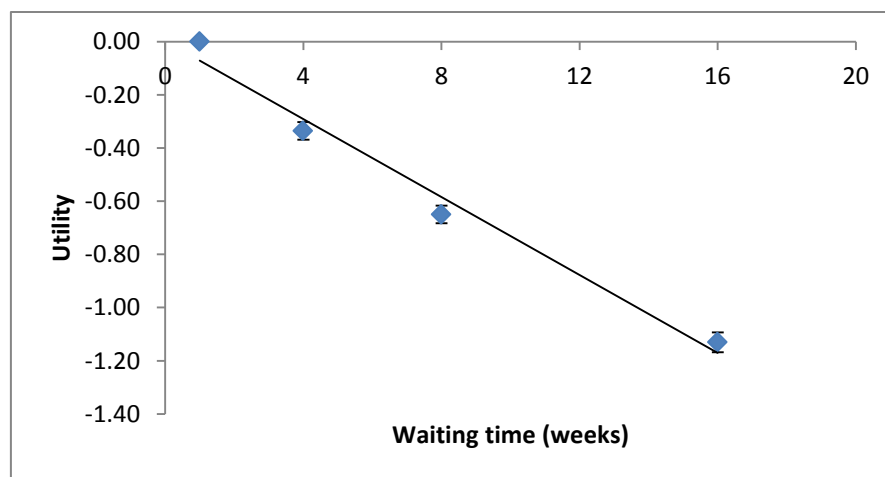


**Figure 23: Relationship between utility and travel time demonstrating a linear function**



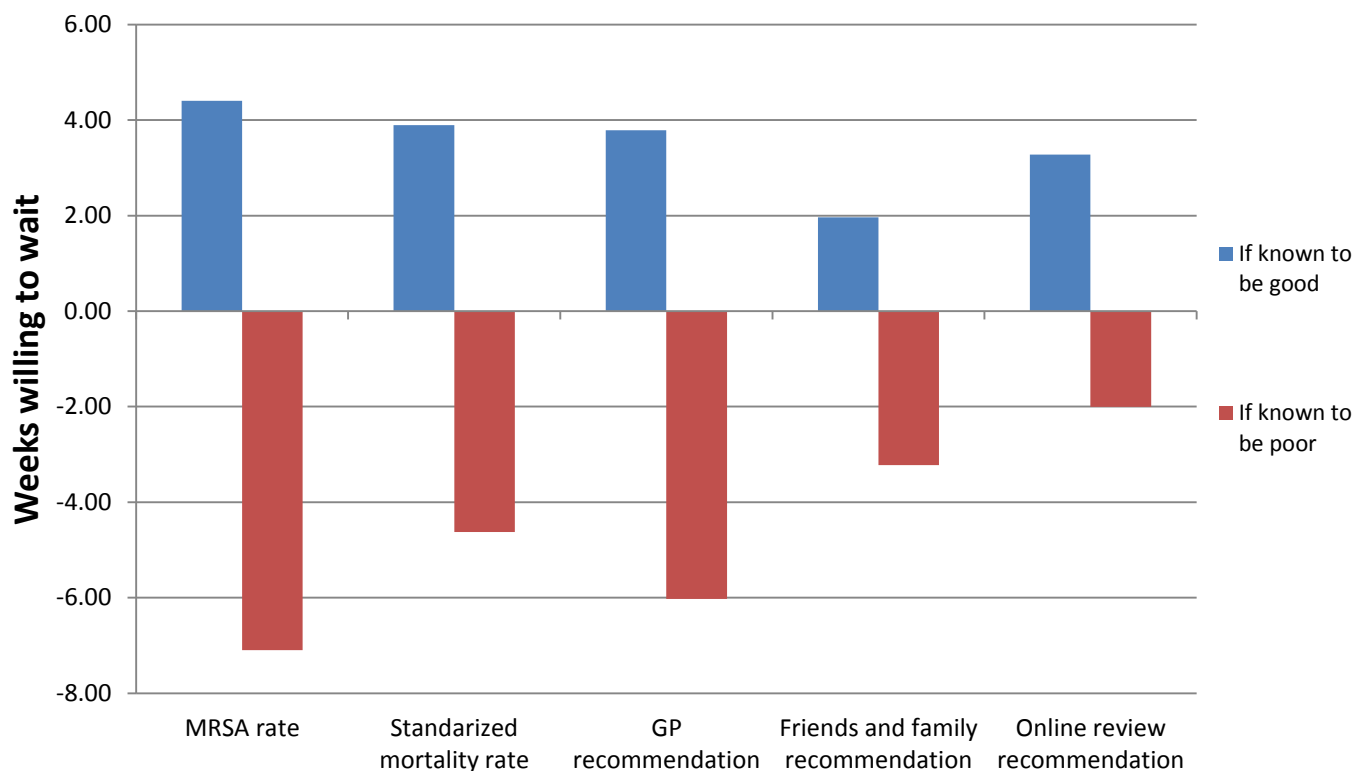
Coefficients for the various attributes and levels in the discrete choice experiment as estimated in the simple model are demonstrated in Table 23. This displays both the coefficient of utility, and the difference in time willing to wait (expressed as weeks) and time willing to travel (in hours). For example, in Table 24, an average member of the public would be willing to wait an extra 4.4 weeks to go to a hospital with a low MRSA rate, compared to not knowing. Similarly they would be willing to wait 7.1 weeks for a hospital where the MRSA rate is not known, compared to a hospital where the infection rate is known to be high (a negative coefficient of 7.1)

**Table 23: Estimates of utility for various attributes in hospital choice, including comparison in terms of waiting time equivalent (simple model not incorporating individual characteristics)**

| <b>Coefficient</b>                                  | <b>Estimate</b> | <b>t-ratio</b> | <b>Weeks willing<br/>to wait</b> | <b>Hours willing<br/>to travel</b> |
|-----------------------------------------------------|-----------------|----------------|----------------------------------|------------------------------------|
| Travel time (hours)                                 | -0.01           | -28.16         | -0.09                            | N/A                                |
| Waiting time (weeks)                                | -0.13           | -25.05         | N/A                              | -10.99                             |
| MRSA infection rates                                |                 |                |                                  |                                    |
| - "High infection rate" relative to "not known"     | -0.95           | -21.14         | -7.10                            | -1.30                              |
| - "Average infection rate" relative to "not known"  | 0.17            | 4.39           | 1.25                             | 0.23                               |
| - "Low infection rate" relative to "not known"      | 0.59            | 18.38          | 4.40                             | 0.81                               |
| Standardized mortality rate                         |                 |                |                                  |                                    |
| - "High mortality rate" relative to "not known"     | -0.62           | -13.88         | -4.63                            | -0.85                              |
| - "Average mortality rate" relative to "not known"  | 0.21            | 5.17           | 1.60                             | 0.29                               |
| - "Low mortality rate" relative to "not known"      | 0.52            | 13.88          | 3.90                             | 0.71                               |
| Whether recommended by GP                           |                 |                |                                  |                                    |
| - "Not recommended" relative to "Has no view"       | -0.80           | -18.39         | -6.02                            | -1.10                              |
| - "Recommended" relative to "Has no view"           | 0.51            | 16.92          | 3.79                             | 0.69                               |
| Whether recommended by friends or family            |                 |                |                                  |                                    |
| - "Not recommended" relative to "Have no view"      | -0.43           | -12.13         | -3.22                            | -0.59                              |
| - "Recommended" relative to "Have no view"          | 0.26            | 9.62           | 1.97                             | 0.36                               |
| Whether recommended by patients on a review website |                 |                |                                  |                                    |
| - "Below average" relative to "No review"           | -0.27           | -7.58          | -2.00                            | -0.37                              |
| - "Average" relative to "No review"                 | 0.18            | 6.72           | 1.34                             | 0.25                               |
| - "Above average" relative to "No review"           | 0.44            | 14.34          | 3.28                             | 0.60                               |
| Constants                                           |                 |                |                                  |                                    |
| - Hospital B relative to Hospital A                 | 0.09            | 3.12           | 0.66                             | 0.12                               |
| - Hospital C relative to Hospital A                 | -0.04           | -1.31          | -0.33                            | -0.06                              |
| - Choose none of the offered hospitals              | -1.34           | -23.26         | -10.06                           | -1.84                              |

These results can also be summarised graphically, as in Figure 24.

**Figure 24: Effect of various hospital attributes on hospital choice, expressed as ‘weeks willing to wait’**



The second model, incorporating individual demographic and attitudinal factors, is shown in Table 24. This reports the same results as Table 23, but also additional values where characteristics of the individual were found to have an additional effect on the choice patterns. Of note, only a very limited number of individual characteristics appeared to improve the model specification, which resulted in only six additional terms being added to the utility estimates for hospital characteristics (represented in Table 24 as ‘additional utility’).

**Table 24: Estimates of utility for various attributes in hospital choice, including individual characteristics**

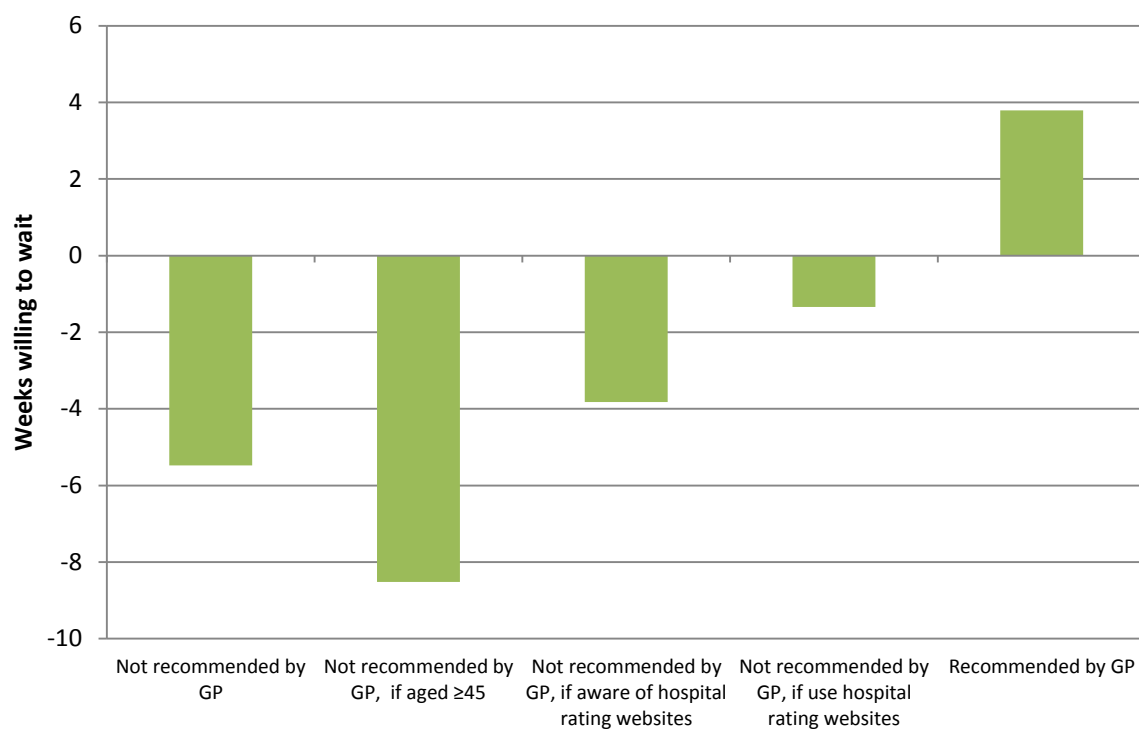
| Coefficient                                                                     | Estimate | t-ratio | Weeks<br>willing to<br>wait | Hours<br>willing to<br>travel |
|---------------------------------------------------------------------------------|----------|---------|-----------------------------|-------------------------------|
| Travel time (hours)                                                             | -0.01    | -28.2   | -0.1                        | N/a                           |
| Waiting time (weeks)                                                            | -0.13    | -25.6   | N/A                         | -11.0                         |
| MRSA infection rates                                                            |          |         |                             | 0.0                           |
| - "High infection rate" relative to "not known"                                 | -1.00    | -20.7   | -7.4                        | -81.4                         |
| - Additional utility on "Not known" if view hospital choice as very important   | -0.18    | -2.7    | -1.3                        | -14.4                         |
| - "Average infection rate" relative to "not known"                              | 0.13     | 2.9     | 0.9                         | 10.5                          |
| - "Low infection rate" relative to "not known"                                  | 0.55     | 14.3    | 4.1                         | 44.8                          |
| Standardized mortality rate                                                     |          |         |                             | 0.0                           |
| - "High mortality rate" relative to "not known"                                 | -0.50    | -10.0   | -3.7                        | -40.9                         |
| - Additional utility on "High mortality rate" if use patient rating websites    | -0.36    | -4.4    | -2.6                        | -29.1                         |
| - "Average mortality rate" relative to "not known"                              | 0.22     | 5.3     | 1.6                         | 17.9                          |
| - "Low mortality rate" relative to "not known"                                  | 0.52     | 13.8    | 3.9                         | 42.9                          |
| Whether recommended by GP                                                       |          |         |                             | 0.0                           |
| - "Not recommended" relative to "Has no view"                                   | -0.74    | -10.6   | -5.5                        | -60.4                         |
| - Additional utility on "Not recommended" if aged ≥45                           | -0.41    | -5.7    | -3.0                        | -33.5                         |
| - Additional utility on "Not recommended" if aware of hospital rating websites  | 0.22     | 2.6     | 1.7                         | 18.3                          |
| - Additional utility on "Not recommended" if use hospital rating websites       | 0.33     | 2.9     | 2.5                         | 27.3                          |
| - "Recommended" relative to "Has no view"                                       | 0.51     | 16.7    | 3.8                         | 41.7                          |
| Whether recommended by friends or family                                        |          |         |                             | 0.0                           |
| - "Not recommended" relative to "Have no view"                                  | -0.34    | -7.1    | -2.5                        | -27.6                         |
| - Additional utility on "Not recommended" if female                             | -0.20    | -3.2    | -1.5                        | -16.4                         |
| - "Recommended" relative to "Have no view"                                      | 0.27     | 10.0    | 2.0                         | 22.0                          |
| Whether recommended by patients on a review website                             |          |         |                             | 0.0                           |
| - "Below average" relative to "No review"                                       | -0.27    | -7.7    | -2.0                        | -22.2                         |
| - "Average" relative to "No review"                                             | 0.18     | 6.7     | 1.3                         | 14.6                          |
| - "Above average" relative to "No review"                                       | 0.44     | 14.4    | 3.2                         | 35.7                          |
| Constants                                                                       |          |         |                             | 0.0                           |
| - Hospital B relative to Hospital A                                             | 0.09     | 3.1     | 0.7                         | 7.2                           |
| - Hospital C relative to Hospital A                                             | -0.05    | -1.3    | -0.3                        | -3.7                          |
| - Choose none of the offered hospitals                                          | -1.29    | -9.4    | -9.6                        | -105.5                        |
| - Additional utility on "Choose none" if view hospital choice as very important | 0.37     | 3.6     | 2.7                         | 30.2                          |
| - Additional utility on "Choose none" if female                                 | 0.32     | 3.9     | 2.4                         | 26.0                          |
| - Additional utility on "Choose none" if live in suburb of city or large town   | 0.26     | 3.1     | 1.9                         | 21.3                          |
| - Additional utility on "Choose none" if in generally good health               | -0.41    | -3.7    | -3.1                        | -33.7                         |
| - Additional utility on "Choose none" if aged <35                               | -0.38    | -4.6    | -2.8                        | -30.7                         |
| - Additional utility on "Choose none" if in full time employment                | -0.25    | -2.7    | -1.9                        | -20.7                         |

An alternative explanation of a part of these results is show in Figure 25. Here you see that

for some groups, demographic characteristics of individuals influence how choice is made.

For example, the figure shows how the older group put more weight on GPs view than the young, and whether the GP would not recommend the hospital in particular. On the other hand, those who are aware of rating websites tend to be less influence by what the GP says, and those who have left a comment online even less so.

**Figure 25: Effect of individual characteristics on choice behaviour**



## **4.4 Discussion**

### **4.4.1 Survey results on attitudes towards choice and online health behaviours**

Our results show around two thirds of patients are aware that they have a choice of health care provider, and this is higher in older patients, those who visit their doctors more frequently, those with worse experience of health providers and those with higher educational attainment. Our results found 87% of respondents thought choice was very important, important or somewhat important. Those more likely to be in favour included women, those who go to their doctor more often, those with previous mixed or bad experience of care, and those with lower educational attainment.

We found that almost 70% of the sample looked for health information online. This was higher in those with higher educational attainment, females, people who have been referred to secondary care recently and those who visited their GP more often. Awareness of online rating of healthcare providers was not high – with only 29.5% aware. Higher rates were found in younger people and people with higher educational status. Of those who were aware of these websites, only 42% had looked at one (12.5% of the total surveyed). Those who had looked at them were again more likely to be younger, go to their GPs more often, and to be of non-white ethnicity. 5% reported having left a comment on a healthcare rating website. Those more likely to do so were younger people, men, those with higher educational attainment and non-white people.

Some of these findings are what might be expected. Taken together, online behaviours related to health are more common in younger age groups, those who make more use of the health system, and those with higher educational levels. At the same time, some factors, which might be expected, do not seem to be driving behaviour. For instance, I found little relationship between income and rating site usage - suggesting different effects by different types of socioeconomic status (however, this might also be due to the fact that household income was unadjusted for family size). Similarly, while I found a difference in usage between urban and rural location in the unadjusted analysis, this was not present when fully adjusted. The size of the effect of age in adjusted model might lead us to think that many of these unadjusted results were mediated by age.

The finding on ethnicity stands out also. There were concerns expressed the black and other minority ethnic groups might access this form of feedback less, leading to inequality. However, our findings suggest those with non-white ethnicity are more likely to be aware of and use online feedback websites than others. That said, this broad 'non-white' ethnicity category might mask differences that are present in some ethnic groups. I was unable to test this further due to the low number of minority ethnic participants recruited. In Chapter 3, I found evidence that suggested usage of review websites is not distributed evenly across demographics at the area level. In this individual level study I have been able to demonstrate the nature and extent of these variations.

#### **4.4.2 Comparison with previous work (Survey attitudes)**

The findings on awareness of choice in the NHS show a consistent trend with earlier findings. Our figure of 66% awareness is higher than both the last NHS Choice survey in 2010 (where 54% of patients knew they had a choice) (73), work by the King's Fund conducted in 2009 (where 45% were aware) (56) and a figure of 29% in 2006 for the first national choice survey (73). This may be due to differences in sampling or may suggest that national awareness of choice is slowly but steadily increasing. The variation between knowledge of choice are also in keeping with earlier finding, with higher awareness in older and more educated groups, and these with more experience of using healthcare (56).

The results on value of choice are also similar to previous work. Dixon and others for the King's Fund found 89 % of people thought choice was very to somewhat important (56) – compared to our 87% for the same question. The British Social Attitude survey found 72% of patients thought they should have at least some choice about the hospital they went to (74). Dixon found older people, those with non-white ethnicity and those with no qualifications valuing it more – similar to our findings, although I did not observe the same effect with age.

I am not aware of comparable survey data about health behaviours online for the English population, but they are similar to the work done by the Pew Centre in the US. The data from the US in 2013 suggest that 71% of adults who are online have used the internet to look for health information in the last year (the same question I asked) (152), very similar to our figure of 70% and suggesting similar adoption of behaviours in the two countries.

#### **4.4.3 Comparison with previous work (Discrete Choice experiments)**

Our findings add new information in terms of how patients incorporate online reviews into a decisions making process to choose hospitals. These results demonstrate that the effect is roughly the same size as the effect of friend or family recommendation, but less than that of mortality rating. This suggests that the new online rating phenomenon, while not defining the choices people make, could be a real factor that can drive patient behaviour (although it has not yet been tested in real choice situations). This shows that hospital organisations should take this information seriously – because their patients are. An interesting subtlety demonstrated here is that the effect of recommendation online has a more positive effect than other sources. For example with GP or friends/family recommendation, the majority of the effect is on the basis of people being less likely to pick a hospital if a negative view is expressed. In contrast, for online ratings, it is positive reviews that seem to more strongly drive positive choice, rather than negative views deterring. The reason for this is not clear, but might be because negative reviews online are discounted as being embittered customers. People may have a sophisticated way of looking at reviews from their experiences in other domains.

Berwick described two pathways by which improvement through public reporting might occur, selection or internal improvement (17). Hibbard later suggested the motivation to protect reputation also drives improvement (20). The evidence so far has suggested that selection by patients is, in practice, a relatively weak phenomenon (7). The introduction of online reviews might act as a further driver to encourage patients to more actively choose their hospitals. In addition, Schlesinger and colleagues have highlighted the potential for the public to become swamped by the complexity and amount of information about

hospitals in public reporting systems, preventing the expected rational decision making approach. Instead, they make decisions on the basis of heuristics, intuition, and their limited understanding – which has been described as ‘bounded rationality’ (79). It is possible that these rating and reviews are simple enough that they are incorporated into these decision-making models, because reviews and ratings are easier to understand than a complex and esoteric concept such as a standardised mortality ratio. However, as the effect demonstrated seems to be smaller than that of GP recommendation and infection rates, it seems less likely that this new information alone will be enough to re-energise selection as a mechanism to drive improvement.

#### **4.4.4 Comparison with previous work**

The findings of our choice experiments are in line with previous work on stated (hypothetical) preference in the NHS, demonstrating that many members of the public will consider choosing a hospital that is not their closest, and this is influenced by a number of metrics of clinical quality (particularly infection and mortality rates), as well as recommendation from trusted sources (GPs and family and friends).

Early work looking at choice in the NHS, the *London Patient Choice Project*, studied how patients responded to choice between their local provider (where there was a wait for treatment) or travelling to an more distant alternative provider, where treatment could be obtained more quickly (253). They found that 30% of patients always stuck with the closest hospital, 5% would always travel further for a shorter waiting time, 10% tended to opt out of treatment, and the remaining 55% switched between closest and a more distant hospital

on the basis of various hospital characteristics. A subsequent report, *Understanding patients choice at the point of referral*, looked at behaviour of members of the general public who had recently been referred to hospital in England (251). They found GP advice was the most important factor driving choice, but did not override other concerns. I saw a strong effect from GP recommendation, but that infection rates were at least as important. They found that negative advice from GPs had a stronger effect than positive advice – which I saw here. Those with lower educational attainment used information about quality less – something I did not detect in building the model incorporating individual characteristics.

Further, Dixon and colleagues in 2010 found 25% of people always chose their local hospital, regardless of the attributes on display, but still leaving a majority of people (around 70%) who can be swayed by the evidence of performance (56). They found patients were willing to travel for 2 hours to attend a hospital with a low rather than a high infection rate. I found they would be willing to travel 2.1 hours for the same, which is reassuringly similar.

However, our findings differed in some areas. They found those more likely to choose their local hospital were more likely not to have internet access, have lower education attainment, did not travel by car, or lived in a city or town. I found little variation in these subgroups. One possible explanation is that in previous experiments, the choice was framed as your ‘local’ hospital or one more distant, whereas in this experiment I simply listed a hospital 30 minutes away as the closest. This may suggest that even beyond travel time, there are other aspects of loyalty to local hospitals that enter people’s choice decisions.

There are some other differences between our work and previous studies. Previous work focused on a group of patients who had been referred to hospital (56,252,253), and

therefore had recently practical experience of the referral process and decision making (if it was offered). This work asks the public in general. In addition, previous work in 2005, demonstrated a nonlinear relationship with waiting time, while ours appeared to show a more linear relationship

#### **4.4.5 Strengths and limitations**

Limitations to this study include the sampling approach taken. In this study I used an Internet polling panel. This has advantages, in that I was able to sample nationally, across a range of geographies, and get better returns in younger age groups than one might from a paper survey approach (where these groups are consistently poorly sampled). However, I am only sampling those with Internet access, and I know that this is not symmetrically distributed across age and social economic status, with the elderly and poor less likely to use it.

At the same time, this approach has a low response rate. The sampling approach deliberately used a stratified approach, and this lowered the response rate, as it was harder to recruit young males. This potentially introduces a sampling bias in terms of who is likely to respond. I hoped to mitigate part of this by not disclosing the purpose of the survey in the initial invitation – which may have encouraged people with particular views or experiences of online rating to be more likely to respond. In addition, the stratified sampling approach resulted in a pool of respondents that was close to representative of the English population in terms of age and gender, and also included a substantial variety of people across socio-

economic status as measured by income and educational status (although the sample and population proportions were significantly different).

Despite the stratified sampling approach, the participating group is unrepresentative of the English public in most aspects – notably ethnicity, where our sample had a lower proportion of people from an ethnic minority than the national population. These low numbers may limit our ability to be certain about the effects of ethnicity on understanding online health behaviours.

In this work, I measured hypothetical rather than real choice. This is for practical reasons. It is harder to elucidate the decision-making process in real world scenarios where the information on display is not controlled. I am also aware from the previous literature that real and hypothetical choice behaviours do not match up completely. In particular, people are more likely to move away from their local hospital in hypothetical than real choices. Dixon found that in real life patient picked their local provider 69% of the time when offered choice, but in hypothetical situations only picked their local example 61% of the time (56). Burge found 35% always stuck with a local provider in real life, but only 30% in hypothetical examples (251).

I am also aware of the inherent limitations of a discrete choice experiment of this sort. Bryan and Dolan note the psychological issues with the method, in that the process of elicitation in a DCE may have effects on the construction of preferences, as people make use of a variety of framing effects and heuristics in their choices (260) and the design of the experiment may well influence the results obtained. The act of presenting participants with

complex information across a number of attributes and providers in an idealised format is very far from the real life experiences of most people when picking hospitals. Although this experiment tried to mimic the information that a patient might have in front of them if they were choosing a hospital while looking at the government's publically reported data, I am aware that this would be a rare and cognitively burdensome event. Similarly, Dolan and Bryan also note normative issues in experimental construction, as experiments are sometimes not ideally placed to answer policy relevant questions because the population and attributes chosen are not representative to the policy context. I tried to mitigate this by deliberately sampling the whole population, rather than people who have recently used the health service, and by picking realistic attributes that reflected the choices offered to members of the public (for example using waiting times set around government targets and hospital performance attributes that mimic publicly reported measures).

As Louvier and Lascar note (261), future experiments might be improved by incorporating ideas such as best/worst scaling – where the participant gives their top and bottom choice from those presented – to maximise the information obtained from each participant.

Some of the survey results in this study seem large. 5% report having left a comment on a healthcare-rating site. This seems unlikely given their current usage levels (see Chapter 3). This might be the result of selecting an Internet savvy group (who take part in polling panels), or the result of misunderstanding the question, as half said they had used similar websites for hotel choice.

Strengths of this study include the comparatively large sample size compared to other work and the representative nature of the sample (despite selection bias and response rates). In this case, the demonstration of a consistent linear effect for both time to travel and time to wait for an appointment in the experimental model is reassuring, and allows meaningful comparison using waiting time as alternative comparative metric. In addition, the results of both the survey and discrete choice aspects, where they duplicate previous work, have similar findings to predecessors.

## **Chapter 5: Use of sentiment analysis for capturing patient experience from free-text comments posted online**

### **5.1 Background**

Understanding patients' experience of healthcare is central to the process of providing care and a fundamental pillar of healthcare quality (15,103). Traditional measures of patient experience are derived from surveys, and more recently, structured patient reported outcome measures. As described in Chapter 1, such approaches ask specific and limited questions, are conducted infrequently and are often expensive to administer. Today's patients have begun to report their experience of healthcare on the Internet in blogs, social networks and wikis, as well as on the healthcare rating websites I have described. However, as this is largely unstructured, non-standardized free-text information, it is not captured in a systematic way. This represents a missed opportunity for understanding patient's experience in an increasingly 'connected' world.

Outside healthcare, natural language processing, including sentiment analysis and opinion mining, of large data sets has been decisive in understanding consumer attributes and behaviours – for example in election forecasting (221). Sentiment analysis enables the content of natural language – the words we write and speak – to be examined for positive and negative opinion and emotion (262). If applicable to healthcare, these analytical methods could permit interpretation of textual information about patient experience on a huge scale. This information, because of its prose nature has avoided the analytical spotlight of conventional quantitative analysis. Alemi and colleagues have proposed the use of sentiment analysis of comments as real time patient surveys (229). They have shown that

patient comments about specific doctors could be attributed with reasonable accuracy to positive and negative sentiment. They further suggest that capture of sentiment analysis should be compared to traditional methods of assessing patient experience.

The Information Strategy for the NHS in England states that sentiment analysis of data could be a novel source of information that would be valuable for patients in facilitating choice of hospitals (89). I test this assertion, and extend the work undertaken by Alemi and colleagues, by analysing a large number of free-text comments on the NHS Choices website. This website allows patients to describe their experiences of treatment at all hospitals in England (and all other NHS provided services). Crucially, these reviews include both free text descriptions of general experience and Likert scale ratings of particular aspects of care. This presents the opportunity for a natural experiment to assess the accuracy of sentiment analysis techniques (applied to the free text comments) against the patients' own quantitative ratings. The NHS also measures patient experience via a national survey of hospital inpatients. If sentiment analysis techniques are to be considered as useful tools for assessing care quality, it is important to see whether there is an association with traditional measures of patient experience. I therefore compare our sentiment analysis findings to the national patient survey, at the hospital level.

## 5.2 Methods

### 5.2.1 Machine learning from patient comments

We applied data processing techniques to all the online free-text comments about hospitals on the NHS Choices website in 2010. Our purpose was to test whether I could automatically predict patients' views on a number of topics from their free-text responses. A machine learning classification approach was chosen in which an algorithm 'learns' to classify comments into categories from a given set of examples, using open-source Weka data mining software. This software has been extensively used in previous research and provides accurate classification results, including in healthcare (263–265). To test the accuracy of the prediction, I compared our results to quantitative ratings provided by the same individual patients on a Likert scale. Free-text comments were examined in response to the questions: 'What I liked', 'What could have been improved' and 'Any other comments'. A prediction was then made about whether the patient would recommend the hospital or not, whether the hospital was clean or dirty and whether they were treated with dignity and respect. The algorithm was trained using all comments and ratings about hospitals left on the NHS Choices website from 2008, 2009 and 2011 (13,802 in total) as a learning set. Data from 2010 were used to test the predicting accuracy of the process (6,412 comments) because comparable patient experience survey data were available for that year. All comments left on the NHS Choices website were provided directly by the Department of Health in England, but have subsequently been made publically available (219).

### 5.2.2 Technical aspects of the machine learning approach

The machine learning approach had two components. First, pre-processing, in which data from patient comments are split into manageable units to build a representation of the data. Second, classification, in which an algorithm decides which category each comment falls into. A consistent set of methodologies was applied in our machine learning process, including a 'bag-of-words' approach, 'prior polarity' and 'information gain'.

In the 'bag-of-words' approach, the total body of words analysed (known as the corpora) is represented as a simplified, unordered collection of words (266). For this analysis, unigrams (single elements or words) and bigrams (two adjacent elements in a string of tokens, in this case two-word phrase) were used as the basic units of analysis. 5695 n-grams in total were extracted. Higher n-grams (longer phrases) could have been used but the constraints were computer power and processing time. I also included our own classification of certain words in the machine learning approach, known as 'prior polarity'. The 1000 most common single words, and the 1000 most common two word phrases were extracted from the complete set of comments in the corpora. Two researchers independently rated the sentiment of each as positive, negative, either or neutral separately for each of the three domains under consideration: 1) overall recommendation, 2) cleanliness and 3) dignity. Where disagreements occurred, the sentiment was discussed and resolved between the two researchers. Kappa statistics for overall rating were 0.76 for 1 word and 0.71 for 2 word phrases. For rating of dignity they were 0.71 for 1 word and 0.70 for 2 words. For rating of cleanliness they were 0.52 for 1 word and 0.48 for 2 words. For all of these calculations  $P < .001$ ).

A technique called ‘information gain’ was used to reduce the size of the bag-of-words by identifying those words with the lowest certainty of belonging to a given class, and then removing them – this is an approach to feature selection. This improved the computation time, and also demonstrates the words with highest predictive accuracy, shown in Table 26. A number of different technical approaches can be taken to classification in machine learning. I applied four different methods, to see which gave the quickest and most accurate results: 1) naïve Bayes multinomial (NBM) (267), 2) decision trees (268), 3) bagging (269) and 4) support vector machines (270). Decision trees and bagging were carried out with REPTree in the Weka package. Support vector machines used an RBF Kernel. The accuracy of the prediction was compared with the patient’s own quantitative rating by calculating, for each method, the accuracy (the percentage of correctly predicted observations from the total number of observations), the F-measure (the harmonic mean of precision and recall), the Receiver Operating Characteristic (ROC), and the time taken to complete the task were calculated. To reduce computing processing time of the classification, I limited the words in the learning process to the top 10,000 words by frequency. All text was converted to lower case. I removed all punctuation. Typographical errors and misspellings were not corrected.

### **5.2.3 Testing prediction accuracy**

To obtain a score to predict sentiment analysis against, patient ratings left on the NHS Choices website on a Likert scale were converted into simple categories, either positive or negative about cleanliness and dignity, to simplify the prediction task. The website presents patients with five options to rate the cleanliness of a hospital: ‘exceptionally clean’, ‘very clean’, ‘clean’, ‘not very clean’, ‘dirty’ and ‘does not apply’. In this analysis, the first three

options were grouped into a 'clean' class and the 'not very clean' and 'dirty' into a 'dirty' class. The website also asks patients to rate whether they were treated with dignity and respect by the hospital staff, with the options being: 'all of the time', 'most of the time', 'some of the time', 'rarely' and 'not at all'. Again, the first three options were grouped, in this case into a 'more dignity' class and the 'rarely' and 'not at all' into a 'less dignity' class. Finally, the NHS Choices website asks all patients whether they would recommend the hospital or not.

#### **5.2.4 Comparing sentiment analysis with the national inpatient survey**

Having calculated the accuracy of our prediction algorithm, the results of the sentiment analysis were then compared with the national inpatient survey results for 2010. This is an annual, national survey of randomly selected patients admitted to NHS hospitals in England, similar to the HCAHPS survey in the US. The 2010 survey covered all 161 acute hospitals with adult services in England, involving 60,000 respondents nationally (response rate 50%). Patients were contacted via post between September 2010 and January 2011, if they had received overnight care in hospital in 2010 (271). This survey includes both general and specific questions. In this study, I used only areas similar to the specific themes predicted from the NHS Choices data. The questions used were: In your opinion, how clean was the hospital room or ward that you were in? (marked on a four point scale- very to not at all); Overall, did you feel you were treated with respect and dignity while you were in the hospital? (marked on a three point scale- very to not at all); Overall, how would you rate the care you received? (marked on a five-point scale- excellent to poor). The sentiment analysis ranking was compared with the patient survey ranking for each question at the hospital

level, applying Spearman's test for rank correlation, using Stata SE11 statistical software. I compared all 161 adult acute trusts in England.

### **5.3 Results**

There was agreement between the patients' own quantitative rating of whether they would recommend their hospital and our prediction from sentiment analysis between 80.8% and 88.6% of the time (Table 25), depending on the classification method used. Similarly, sentiment analysis agreed with whether the patient was treated with dignity and respect between 83.7% and 84.5% of the time, and agreed with whether the hospital was clean or not between 81.2% and 89.2% of the time. Each classification approach produced similar precision, but the NBM algorithm, a first-order probabilistic model that uses word frequency information, performed the calculation faster (less than 0.2 seconds compared to hundreds of seconds for the other analysed approaches). This accuracy of precision for the NBM approach was reflected as Kappa statistics for inter-rater reliability of between 0.74 and 0.40 ( $p < 0.001$  for all). Table 26 shows the 10 words or two word phrases with the highest predictive accuracy.

Table 25: Accuracy of different approaches to machine learning

| Question            | Naïve Bayes Multinomial |           |              |          | Decision Trees |           |              |          | Bagging |           |              |          | Support Vector Machine |           |              |          |
|---------------------|-------------------------|-----------|--------------|----------|----------------|-----------|--------------|----------|---------|-----------|--------------|----------|------------------------|-----------|--------------|----------|
|                     | ROC                     | F-measure | Accuracy (%) | Time (s) | ROC            | F-Measure | Accuracy (%) | Time (s) | ROC     | F-Measure | Accuracy (%) | Time (s) | ROC                    | F-Measure | Accuracy (%) | Time (s) |
| Overall rating      | 0.94                    | 0.89      | 88.6         | 0.11     | 0.84           | 0.81      | 80.8         | 552      | 0.89    | 0.82      | 82.5         | 4871     | 0.79                   | 0.84      | 84.6         | 612      |
| Cleanliness         | 0.88                    | 0.84      | 81.2         | 0.05     | 0.76           | 0.86      | 88.4         | 206      | 0.83    | 0.87      | 89.2         | 2018     | 0.53                   | 0.84      | 88.5         | 305      |
| Dignity and respect | 0.91                    | 0.85      | 83.7         | 0.06     | 0.79           | 0.80      | 83.0         | 332      | 0.87    | 0.85      | 84.5         | 3164     | 0.60                   | 0.80      | 84.1         | 520      |

**Table 26: The 10 words or two word phases with the most predictive accuracy**

| Prediction category            | Overall      | Cleanliness | Dignity   |
|--------------------------------|--------------|-------------|-----------|
| Top ten most important phrases | 1) told      | dirty       | rude      |
|                                | 2) thank you | floor       | told      |
|                                | 3) left      | left        | left      |
|                                | 4) rude      | the floor   | thank you |
|                                | 5) excellent | thank you   | friendly  |
|                                | 6) the staff | filthy      | excellent |
|                                | 7) hours     | bed         | rude and  |
|                                | 8) asked     | patients    | asked     |
|                                | 9) was told  | friendly    | the staff |
|                                | 10) friendly | hours       | staff     |

There were significant, weak to moderate associations between machine learning predictions using NBM and patient survey quantitative responses from the national inpatient survey for the three categories examined: cleanliness, dignity and overall recommendation (Spearman correlation coefficients between 0.37 and 0.51,  $P < .001$  for all) (Table 27). Correlations for overall recommendations of hospitals are displayed in Figure 26.

**Figure 26: Comparison of the proportion recommending a hospital using sentiment analysis and traditional paper based survey measures**

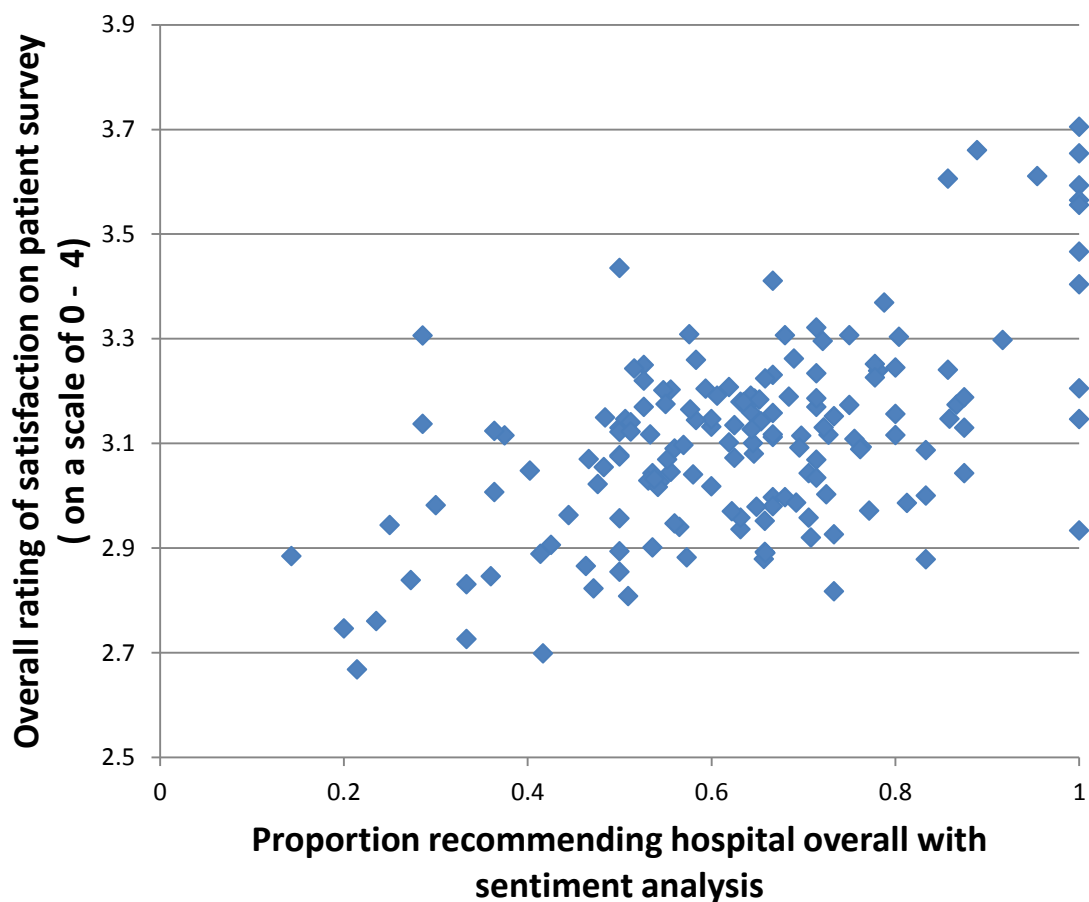


**Table 27: Comparison of patient survey responses and machine learning prediction of comments at hospital trust level**

| Patient Survey Question                                                                         | Machine learning prediction                                                                            | Spearman Correlation coefficient | Probability |
|-------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|----------------------------------|-------------|
| In your opinion, how clean was the hospital room or ward that you were in?                      | Machine learning prediction of comments about standard of cleanliness                                  | 0.37                             | P<0.001     |
| Overall, did you feel you were treated with respect and dignity while you were in the hospital? | Machine learning prediction of comments about whether the patient was treated with dignity and respect | 0.51                             | P<0.001     |
| Overall, how would you rate the care you received?                                              | Machine learning prediction of comments about whether the patient would recommend                      | 0.46                             | P<0.001     |

## 5.4 Conclusions

Our results reinforce earlier findings that sentiment analysis of patients' online comments is possible with a reasonable degree of accuracy (229). I have also shown that unstructured comments in free-text, if processed appropriately, are associated with patient experience results from paper based surveys conducted annually across all hospitals in England. This is in keeping with previous work that has demonstrated that there is a significant association between structured online ratings of care left on healthcare review websites and conventional surveys of satisfaction (see Chapter 3). These results suggest a potential mechanism to make use of the large amounts of text on the Internet in which people describe their care, and that further work exploring the information contained with the free-text comments may be an important avenue for understanding patient experience, providing an additional source of information alongside traditional survey methods.

### 5.4.1 Strengths and limitations

Sentiment analysis via a machine learning approach is only as good as the learning set that is used to inform it. By taking advantage of a complete national rating system over several years I have been able to use many more ratings in this learning set than in other studies. Indeed, our learning set was more than 10 times larger than earlier work [9]. Moreover, in applications of sentiment analysis to healthcare data, researchers have had to train the system themselves by reviewing comments and ascribing characteristics to them, to allow the algorithm to learn. I used a large set of data that permitted us to directly compare free text comments and quantitative ratings posted by the same patients, thus eliminating potential biases of reviewer assignment of comments. Similarly, the consistency of

questions across the NHS inpatient survey and the questions asked on the NHS Choices website about health services allows for a direct comparison of patient comments and surveys that I believe has not previously been reported.

Online comments left without solicitation on any website are likely to have a natural selection bias towards examples of both good and bad care, as I noted in our earlier experiments. It is likely that these online reviews are contributed more by those in particular demographic groups including younger and more affluent people (see Chapter 3). Further, there are aspects of patient's comments that are very hard for sentiment analysis to process. Irony, sarcasm and humour, frequently adopted by English speakers when talking about their care, cannot be easily detected using this process. The use of prior polarity improved the results, and mitigated some colloquial phrasing, but there were difficulties understanding those that depend on context. For example, phrases that cropped up repeatedly, such as 'stank of urine' or 'like an angel', could be easily characterised as negative or positive. The meaning of other frequently used phrases, however, was hard to establish without an understanding of their context. The best example of this was the phrase a 'cup of tea'. It was referred to in many different comments in these data, but without knowing the context, is it impossible to allocate it a direct sentiment. "They didn't even offer me a cup of tea" is very different to "The nurse even made me a cup of tea". Our current algorithm could not yet make use of references to cups of tea or similar phrases that would be clear and obvious looked at by eye on a case-by-case basis. Future attempts to improve a natural language processing ability for patient experience would have to develop the capacity to accurately interpret this level of context-specific and idiomatic content.

### **5.4.2 Policy implications**

Large amounts of data about the use of services are collected in digital form. An important strand of this is consumer opinion and experience. Today, many people express their views and share their experience of goods and services via the Internet and social media. Such data, converted to information, are essential in improving services, facilitating consumer choice and, in some sectors, exploring public accountability and value in the use of taxpayer's money.

By its nature, the information is highly personalised, idiosyncratic, and idiomatic. However, if it is to be useful, it must be analysed in ways that are not solely reliant on someone reading individual contributions (although this is valuable to consumers) nor on pre-structured responses necessary to allow aggregation. A solution to the challenge of 'Big Data' is to find automated methods for analysing unstructured narrative comment that is a potentially rich source of learning. In this respect, healthcare is no different to many other industries although it has perhaps been slower than other sectors to recognise the importance of it.

As our confidence in techniques of data mining and sentiment analysis grow, information of this sort could be routinely collected, processed and interpreted by healthcare providers and regulators to monitor performance. What is more, information could be taken from a number of different text sources online, such as blogs and social media. If this information could be harvested from these locations, and then processed into timely and relevant data, it could be a valuable tool for quality improvement. In Chapter 2, I hypothesized that as the usage of rating websites, social networks and micro-blogs increases, this free-text

information represents a growing and largely untapped source of data that could be considered a 'cloud of patient experience'. Others, including Alemi have discussed similar ideas, with Cambria and colleagues describing a notion of 'crowd validation' of a health service (229,231). This has the potential to provide up to date information about patient experience at lower cost than the traditional survey route. It might also allow the views of younger, more tech savvy groups - who are often poor responders to paper based surveys - to be sampled.

In the initial hypothesis, I described two essential components of this task: harvesting and data processing. The results here suggest that the data processing part is possible with a reasonable amount of accuracy, although there is still much to learn to improve and strengthen the process.

## **Chapter 6: Analysis of patients' comments about the English NHS via Twitter, and comparison with patient surveys.**

### **6.1 Background**

Twitter is a prominent social media outlet and 'micro-blogging service'. Its members can post text-based messages of up to 140 characters about any topic. Twitter and other social media provide users with the opportunity to share opinions widely and to interact with one another. It is well used, with more than 500 million accounts and more than 350 million messages sent a day (272) (145) (273). People talk about all aspects of their lives. Some people are using it to talk about healthcare.

A wide and increasing body of work has attempted to draw meaningful signals from the information being transmitted in the messages left on social networks. In the introduction, I mentioned some of this research in other fields, including the use of Twitter to try to predict Hollywood blockbusters (222), election results (221) or the price of the stock market (223). In healthcare, people have looked to see if they can detect patterns of disease using the digital signals produced on these websites, for example detecting and mapping outbreaks of influenza or cholera (225,274), by measuring how often and where terms relating to the disease are mentioned.

People have also sort to understand not just the frequency of tweets and the patterns they reveal, but to study the content of them in more detail using qualitative approaches. In healthcare, studies have used techniques such as content analysis to look social media. Rui and colleagues used this approach to examine the types of social support provided by health

organisations on Twitter (275), Eysenbach used it to look at what people were saying about flu outbreaks (274) and Lyles used it to look at tweets about cancer screening (276) (277). Simple sentiment analysis of tweets related to health topics has also been used by Love (looking at tweets about vaccine attitudes) (278) and King (looking at tweets about healthcare reform) (279).

In the hypothesis stated earlier in the thesis, I suggest that it might be possible to use the 'cloud of patient experience' that is generated online to measure the quality of health care providers. To do this we would have to demonstrate that platforms such as Twitter contain useful information for this purpose, and that it is collectable and measurable. While people have studied what healthcare organisations and physicians are saying on Twitter (280) (281), I am not aware of any study that has systematically measured tweets aimed at healthcare providers, looking to see what the patients and the public are saying about the quality of care they are being provided

This study examines the content of tweets directed at acute NHS hospitals to see what proportions of tweets are related to care quality, what aspects of care people talk about, and to investigate whether the information that people post on Twitter could have the potential to be used as metric of clinical quality as a part of the 'cloud of patient experience' described in Chapter 2. As well as this qualitative examination of what people say, I also aim to compare tweets at the hospital level with survey measure of satisfaction, using an automated sentiment analysis technique, to see if there is any relationship between commentary on Twitter and more traditional approaches to measuring patient experience such as surveys.

## 6.2 Method

### 6.2.1 Identification and collection of tweets

I set out to capture the tweets that patients sent to NHS hospitals over one year. There is no centrally available database of tweets. Twitter makes the data produced in tweets available publically via its API (application programming interface) – sometimes known as the ‘fire hose’, but only transiently. Therefore, to investigate patterns of tweets over a prolonged period of time, they need to be collected and stored. I prospectively collected tweets about hospitals from the API.

I identified tweets that were aimed at NHS hospitals by using ‘Twitter mentions’ – this is when a tweet included the ‘@’ username of another Twitter user, automatically alerting the account mentioned that they are being discussed. This is normal behaviour on Twitter for deliberately including someone in a conversation. In order to identify which NHS trusts (hospital organisations) were on Twitter (and had @usernames), I took the complete list of hospital trusts from the NHS Health and Social Care Information Centre. For each trust, I searched their web page, and their contact information online. I also searched for the name of the trust on the main Twitter website. In April 2012, I identified 75 (of 166) Trusts as being on Twitter. Data collection for all tweets mentioning these trusts began in May 2012 and continued for one year.

In order to understand the characteristics of hospitals on Twitter, compared to other hospitals not using the platform, I collected data on hospital size (bed number), activity (total admission rates) and performance. Performance was measured as risk adjusted

mortality rate (Summary Hospital Level Morality Indicator July 2012 – 2013) and patient experience (overall rating of satisfaction from the NHS inpatient survey (282) ). Hospital size data were obtained from NHS England (283) and activity data were obtained from the NHS Information centre (284), both for 2012. Mortality rates were obtained from the Health and Social Care Information Centre indicator portal (285).

### **6.2.2 Measuring volume and frequency of tweets**

Having collected the tweets, I performed a simple descriptive analysis of the total set collected. This included measuring the frequency of tweets by hospital trust, the frequency of tweets by day over the study period, and the frequency of tweets by hour of the data, to see if there were observable patterns of activity.

### **6.2.3 Content analysis**

I used a directed content analysis approach to look at the key themes people were talking about in tweets mentioning hospitals. Qualitative content analysis – as described in a healthcare context by Elo and Kynga (286) – and building on the earlier foundational work on qualitative analysis by Glaser and others (287) – seeks to explore the themes that are contained within a body of text. A ‘directed’ approach to content analysis makes use of previous theory to consider the themes (288). I used the initial theory that tweets about care quality could be divided into the component parts described in the NHS quality definition: patient experience, effectiveness and safety (103).

In order to develop a thematic codebook, a second person and I analysed a sample of 250 random tweets, rating them manually for sentiment, and coding them thematically. The unit

of analysis was individual tweets. A codebook was developed by iterative discussion between both reviewers. Although we used the NHS's quality definition as a starting point, we broadened out the coding to include a number of other topics discovered, as only a proportion of the tweets related to quality. Kappa scores for both sentiment and the primary theme coding were calculated between the first and the second reviewer for the 250 jointly coded tweets, after final development of the codebook. Once the codebook had been defined, it was then used to code a further 750 random tweets (by one reviewer). No further themes emerged in the subsequent analysis of the remaining group, suggesting saturation of themes had been reached (288).

### **6.2.3 Automated sentiment analysis**

We also performed sentiment analysis of the collected tweets using a commercially available sentiment analysis product from TheySay Ltd. I applied this to all of the collected tweets to produce a sentiment score for each tweet of positive, negative or neutral. Using this information, I calculated the average sentiment of tweets per trust, expressed as a proportion of positive tweets compared to all tweets. I compared this to the results of the national inpatient survey for 2012, obtained from the UK data service using the overall rating of experience question (Question 68) (282). Some hospital trusts were not included in the general inpatient survey, including children's hospitals and specialist hospitals, so the analysis presented is limited to 59 trusts. Comparison was done with Spearman's rank correlation coefficient, to see if there was a relationship between sentiment on Twitter and large random survey measures of patient experience. Analysis was conducted with Stata SE software.

## 6.3 Results

### 6.3.1 Descriptive analysis of tweet volume and frequency

I collected tweets from 17 April 2012 to 26 June 2013. Over the period there were a number of outages due to technical reasons, including changes to the API, failures of servers and lack of memory (accounting for 64 days in the first year period), so an additional period of measurement was added, to ensure a 365 day collection period in total. I collected 187424 tweets over the 365-day period. The total number of retweets was 88297.

### Characteristics of hospital trusts on Twitter

No significant difference was observed between hospitals that were on Twitter and those that were not, according to the four comparative features used.

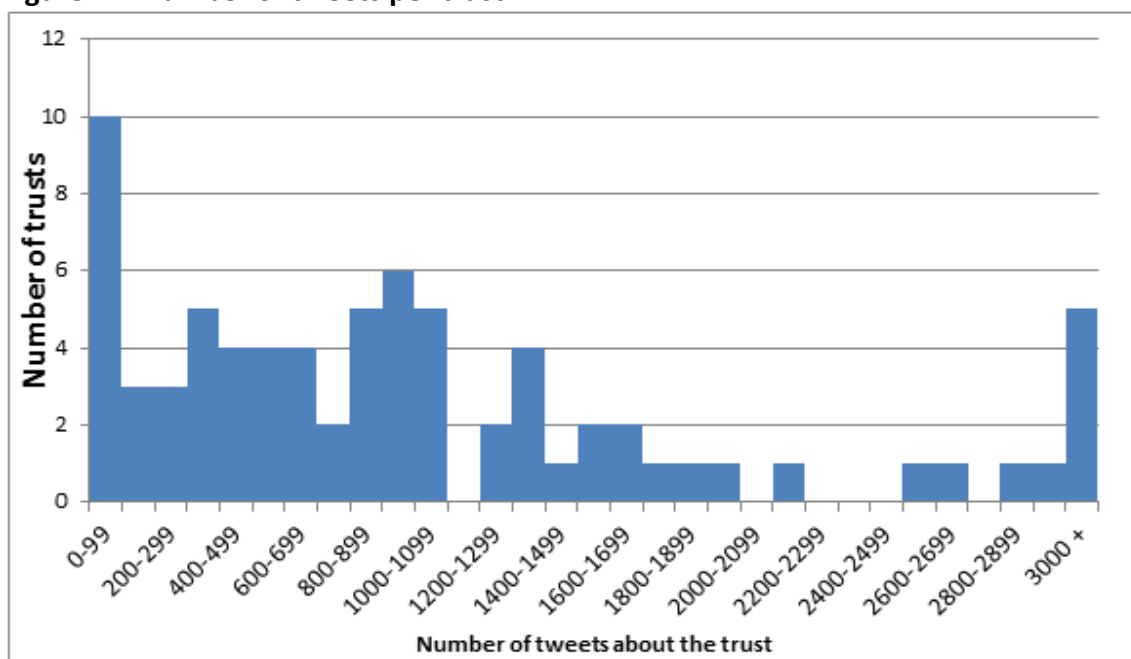
**Table 28: Characteristics of hospital trusts on Twitter**

| Hospital Characteristic                 | Hospitals on Twitter | Hospitals not on Twitter | P value |
|-----------------------------------------|----------------------|--------------------------|---------|
| Total number of Beds                    | 672                  | 682                      | 0.86    |
| Total number of Admissions              | 88444                | 88027                    | 0.96    |
| Overall satisfaction (inpatient survey) | 7.75                 | 7.91                     | 0.15    |
| Risk adjusted mortality (SHMI)          | 0.99                 | 1.01                     | 0.12    |

### Tweets per trust

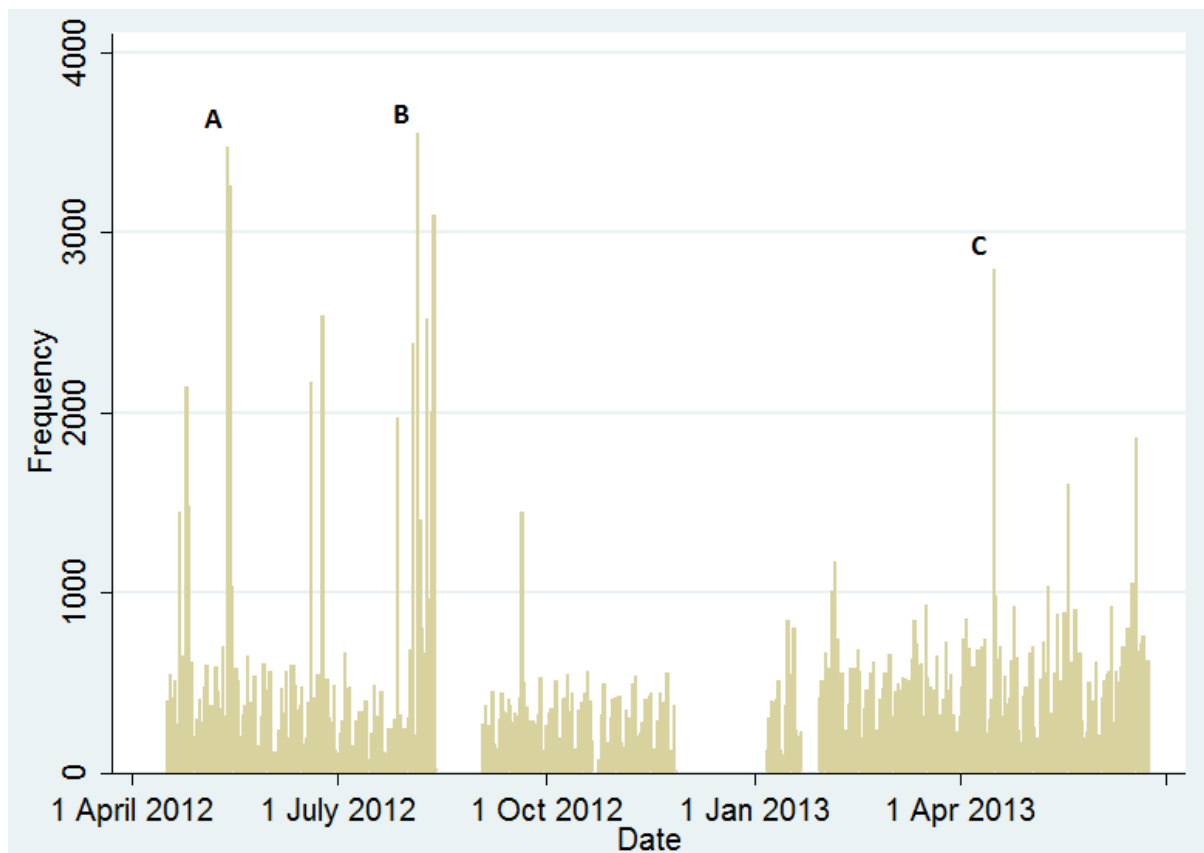
The mean number of tweets per trust was 2499 and the median was 742. The range was 1 to 84110 per trust. There were three large outliers, Great Ormond Street Hospital (a specialist children's hospital) with 84110 tweets, Birmingham Children's Hospital with 18533 tweets and the Royal Marsden Hospital (a specialist cancer hospital) with 14504 tweets. The distribution is shown in Figure 27.

**Figure 27: Number of tweets per trust**



We examined the number of tweets by day, in order to see whether there was a steady stream of information via Twitter. The mean number was 508 and the median was 405 per day. The range was 65 – 3549 tweets per day. The frequency of tweets by day is shown in Figure 28.

**Figure 28: Frequency of tweets by day over the period studied**



On a number of days the number of tweets was substantially higher than usual. I looked at three of the most active days in more detail to find the most frequent message on those days. On 13<sup>th</sup> May (Point A) there were 3489 tweets. This message was retweeted 2260 times:

*RT @Piersmorgan: Most painful tweet of my life: well done @samnasri19 - I am sending a cheque to @greatormondst for 10k today.*

For context, Piers Morgan (television presenter) was discussing a bet about the result of the football season with Alan Sugar (Business tycoon and star of 'The Apprentice'), with the proceeds going to Great Ormond Street.

On 8<sup>th</sup> August (Point B), there were 3549 tweets in total. The most common message was retweeted 813 times. Piers Morgan was discussing his bet at the Olympics, with proceeds going to Great Ormond Street.

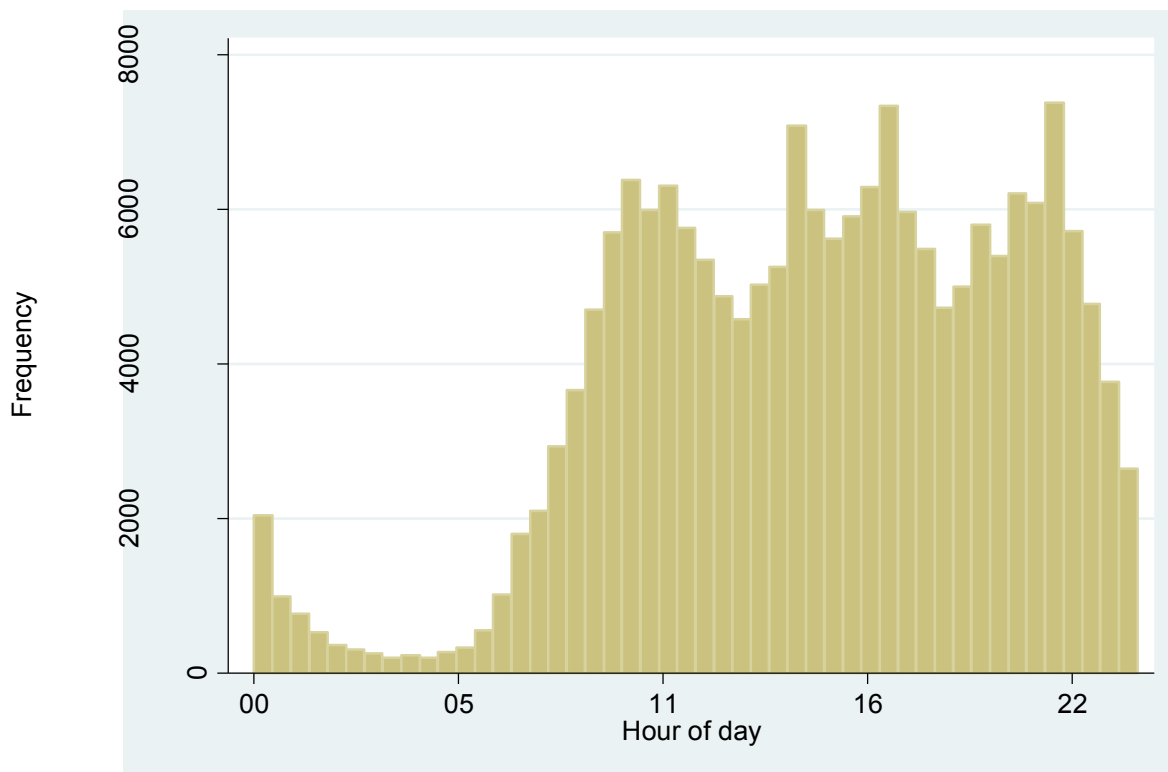
*RT @Piersmorgan: BOOM! BOOM! Murray, a Scot, wins Gold & sings GSTQ. For that, he's earned £2k for @GreatOrmondSt - total now £14k.*

On 16<sup>th</sup> April 2013, there were 2800 tweets. The most common was a retweet of a fundraising message by Nathan (member of popular boy-band 'The Wanted'), which was retweeted 1771 times:

*RT @NathanTheWanted: Guys!! I'm using my bday to raise money for @royalmarsden! Give a gift and donate to help cancer patients here [weblink] :D x*

We also recorded the number of tweets at different times of day these are shown in Figure 29, with peaks in activity at 10:30AM, 2: 00 PM, 5:50PM and 9:30PM, and a lull overnight.

**Figure 29: Frequency of tweets about NHS hospitals by hour of day**



### 6.3.2 Content analysis

In the process of developing the codebook for tweets, we identified six key themes that tweets were about, of which 'quality' was only one. The others were 1) fundraising activities, 2) health information, 3) organisational or practical information about the hospital, 4) promotional messages and 5) messages to patients receiving care. Within these themes, some were divided into further categories and sub-categories. The final codebook, and the frequency of the codes, is shown in Table 29. The inter-rater reliability for the main theme level between the two human raters for each tweet was 0.82 ( $P < 0.001$ ). Each tweet could be assigned more than one theme (although in practice due to their brevity, this was rare). Of the 1000 tweets, 47% were positive, 47% were neutral and 5.6 % were negative.

Inter-rater reliability for sentiment between human raters was as 0.76 ( $P<0.001$ ). Examples of each code are shown in Table 30.

**Table 29: Codebook of tweet classification including frequency of occurrence**

| Main theme                           | n   | %    | Minor theme                             | n  | %   | Sub theme                       | n  | %   |
|--------------------------------------|-----|------|-----------------------------------------|----|-----|---------------------------------|----|-----|
| Care Quality (1)                     | 114 | 11.4 | Patient centred (1.1)                   | 98 | 9.8 | Staff (1.1.1)                   | 55 | 5.5 |
|                                      |     |      |                                         |    |     | Environment/ Facilities (1.1.2) | 11 | 1.1 |
|                                      |     |      |                                         |    |     | Access/ Timeliness (1.1.3)      | 14 | 1.4 |
|                                      |     |      | Effective (1.2)                         | 29 | 2.9 |                                 |    |     |
|                                      |     |      | Safe (1.3)                              | 4  | 0.4 |                                 |    |     |
| Health Information (2)               | 44  | 4.4  | Requesting information (2.1)            | 10 | 1   |                                 |    |     |
|                                      |     |      | Giving information (2.2)                | 34 | 3.4 |                                 |    |     |
| Organisational Information (3)       | 71  | 7.1  | General organisational (3.1)            | 54 | 5.4 |                                 |    |     |
|                                      |     |      | Service configuration (3.2)             | 17 | 1.7 |                                 |    |     |
| Fundraising (4)                      | 387 | 38.7 | Fundraising involving a celebrity (4.1) | 96 | 9.6 |                                 |    |     |
| Promotion/ advertising (5)           | 88  | 8.8  | Individual promotion (2.1)              | 23 | 2.3 |                                 |    |     |
|                                      |     |      | Organisational promotion (2.2)          | 65 | 6.5 |                                 |    |     |
| Related to a patient in hospital (6) | 105 | 10.5 |                                         |    |     |                                 |    |     |
| Non-specific content (7)             | 207 | 20.7 |                                         |    |     |                                 |    |     |

**Table 30: Examples of tweets in each of the categories**

|              |                    |                                            |                                                                                                                                                                |
|--------------|--------------------|--------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Care Quality | Patient Experience | Staff interaction                          | [@named hospital] Want to say thanks to [named staff members] in resus.Were superb in their care overnight of my mum in law.so compassionate x                 |
|              |                    |                                            | Home from [@named hospital] after a weeks stay...we feel blessed to have been cared for by such an amazing team. Thank you [named ward] x'                     |
|              |                    |                                            | [@named hospital] Please tell your staff that their administrative mistakes are not the patient's fault, and not to treat them like idiots.                    |
|              |                    |                                            | [@named hospital] My complaints are not primarily of a clinical nature, more a lack of care. i will mail the team today. thx for your response.                |
|              |                    | Environment/<br>Facilities:<br>Cleanliness | [@named hospital] Your a&e department is absolutely filthy it makes the hospital visit even more unpleasant. #unsatisfactory                                   |
|              |                    |                                            | [@named hospital] Look dirty floors already at 8.30 [weblink]                                                                                                  |
|              |                    |                                            | Q Which [@named hospital] hospital has hand dryer as dirty as this? [weblink]                                                                                  |
|              |                    |                                            | [@named hospital] Shit on floor wet sheets, visitors having to change bedding, shit in toilet, ignorant staff- [named ward]!! Stay away !                      |
|              |                    |                                            | Be nice if this room had been cleaned before we got it. Blood filled cap from an iv on the bedside cabinet, unflushed toilet [@named hospital]                 |
|              |                    |                                            | Credit to [@named hospital]: friendly staff, efficient clinic and everywhere spotlessly clean #proudoftenhs #grateful #freeatthepointofneed                    |
|              |                    | Environment/<br>Facilities:<br>Food        | Don't suppose there is any chance of full english [@named hospital] Been here since 3 yesterday no hot food or drink #poor                                     |
|              |                    |                                            | Back from another stay [@named hospital] and have to agree generally the food is excellent - very impressed                                                    |
|              |                    |                                            | [@named hospital] Cannot believe I have been served this for my 18 month old. Tastes disgusting and hardly nutritional! [weblink]                              |
|              |                    | Environment/<br>Facilities:<br>Parking     | [@named hospital] Poor poor parking....sort it out, shud av designed a hospital with adequate facilities in first place.#notsuprised                           |
|              |                    |                                            | Changes in disabled parking [@named hospital] Outrageous !!!                                                                                                   |
|              |                    |                                            | Positive support from residents. [@named person] is delighted to hear engagement from [@named hospital] on parking.                                            |
|              |                    |                                            | [@named hospital] Ridiculously difficult to park today anyway near physio dept. People are using drop off and double parking in the disabled'                  |
|              |                    |                                            | [@named hospital] where the waiting time is ridiculous waited 3hr yesterday, 3lots of bloods took, 2hrs so far today for a blood test again!                   |
|              |                    |                                            | [@named hospital] Well it's 8:44 and I am already on the ECP machine and I've seen the Dr too... Super efficient service from [named ward] ...                 |
|              |                    |                                            | still waiting at [@named hospital] for a 10.40 appointment with [named department]. #gettingridiculous                                                         |
|              |                    | Effective care                             | @worldmalariaday a huge thank you to [@named hospital] for their speedy and effective diagnosis of malaria yesterday. What a coincidence.                      |
|              |                    |                                            | Horrific exp [@named hospital].. Nurse had no respect for me, didn't speak to me & i'm discharged home with the same thing i came in with!!                    |
|              |                    |                                            | [@named hospital] [Named chief executive ] should come down onto the wards n see what's really goin on under her nose. I wish my nan was in [another hospital] |
|              |                    |                                            | [@named hospital] my nan is on [named ward]. I'm appalled at the care! I'm a nurse & would never treat my patients like that. The CQC will enjoy my            |

|                            |                |                                                                                                                                                    |
|----------------------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
|                            |                | complaint                                                                                                                                          |
|                            | Safe care      | [@named hospital]Yes pls. Main concern now is the doctor overprescribing. We worked out the error but vulnerable patient might not                 |
|                            |                | Getting pretty fed up of the lack of care [@named hospital] n mistakes made.First i got told i was cured "looked at someone else's scan"????       |
|                            |                | #tweetsfromhospbed [@named hospital] I think I'll have to report him, needs more training on drugs before serious mistake hate to do it though'    |
| Charity                    | General        | Laughterhouse comedy night in aid of [@named hospital] 8 acts for an amazing cause! [weblink] [named radio station] RT pls                         |
|                            |                | Looking forward to running the Royal Parks Half Marathon for [@named hospital] this wknd'                                                          |
|                            | Celebrity      | RT@piersmorgan: Just a reminder, Shuggsy, of our £20,000 bet (to [@named hospital] ) on who, Arsenal or Spurs, finishes highest in League.. Wow!'  |
|                            |                | RT @NathanTheWanted: Guys!! I'm using my bday to raise money for [@named hospital]! Give a gift and donate to help cancer patients here http ...   |
| Promotion                  | Individual     | Was a privilege to be exec champion for the excellence in customer care award at last night's [@named hospital] staff awards. Early night tonight! |
|                            |                | [@named hospital] our very own [named staff member] presenting #LOPS13 Improving patient safety [weblink]                                          |
|                            | Organisational | [@named hospital] Named One Of The UK's 40 Top Hospitals for the Fifth Year Running [weblink]                                                      |
|                            |                | Such a fabulous night recognising all the hard work and dedication that goes on at [@named hospital] hospital!                                     |
| Health Information         | Giving         | Patients - don't be afraid to remind staff [@named hospital] of the #5Moments for hand hygiene –[weblink]                                          |
|                            | Asking         | [@named hospital] how much do you think dairy products are a contributory factor towards asthma'                                                   |
| Organisational information |                | RT[@named hospital]: Most outpatient clinics running today We'll be reviewing afternoon clinics & if any cancelled, we'll contact patients '       |
|                            |                | 'RT [namedperson] Over 30 local organisations join up to give views on draft Joint Health & Wellbeing Strategy [@named hospital] '                 |
| Support for patient        |                | 'RT [named person]: He will remain in the ITU [@named hospital] he's very poorly but he's a little fighter #staypositive #thankyou'                |
|                            |                | 'This Tuesday back [@named hospital] for Will's Cath. We should find out if he has pulmonary hypertension. If so, hopefully stent can help...'     |

### 6.3.3 Tweets about care quality

Tweets about care quality accounted for only 11.4% of those that were examined. For those tweets in this group, the existing approach to describing quality in the NHS (dividing it into patient experience, effectiveness and safety) was feasible for classifying the tweets.

However, tweets were not distributed evenly between these different classifications – 87% (98/113) mentioned patient experience, 26% perceived effectiveness and 4% safety.

Sentiment was also not evenly distributed. For all tweets related to quality 77% (87/113) were positive, and (24/113) 21% were negative, and 2% (2/113) were neutral.

#### *Patient experience*

We found tweets about patient experience could often be further broken down into sub-divisions around 1) relationships with staff, 2) the environment and facilities within the hospital, and 3) issues of access and timeliness of services. Many tweets referred to interactions with staff, and were often full of praise - 96% (53/55) were positive. Specific wards, teams and named members of staff were often named.

At the [@named hospital] just had an operation on me foot. Outstanding care as usual, & the nurse has just made me a cracking cup of tea :-)

[@named hospital] Want to say thanks to [named staff members] in resus. Were superb in their care overnight of my mum in law.so compassionate x

There were examples of negative comments, but these were rare.

[@named hospital] [named ward] - Disgusted with your treatment of my mother. Will be making huge complaints.'

10% (11/114) of the quality related tweets were about facilities. These were relatively balanced, some full of praise for the standards of hospital accommodation, but some negative and occasionally combative - 43% (6/14) were positive and 43% (6/14) were negative. Cleanliness was a topic that came up more frequently.

"Be nice if this room had been cleaned before we got it. Blood filled cap from an iv on the bedside cabinet, unflushed toilet [@named hospital]"

"Spent a night in [named hospital] with my son. Excellent care - spotlessly clean. Thank you [@named hospital]"

People also spoke about practical issues of access such as parking.

Positive support from residents. [named person] is delighted to hear engagement from [@named hospital] on parking.

[@named hospital] Ridiculously difficult to park today anyway near physio dept. People are using drop off and double parking in the disabled'

Tweets about food were present. Again people were willing to both praise and criticise.

*Don't suppose there is any chance of full english [@named hospital] Been here since 3 yesterday no hot food or drink #poor*

*Back from another stay [@named hospital]and have to agree generally the food is excellent - very impressed*

### ***Timeliness/access***

The real time nature of Twitter allowed people to make comments about waiting, often while they were waiting. Again, these revealed both positive and negative experiences, and accounted for 12% (14/114) tweets about quality.

*[@named hospital] Thanks for squeezing me in with orthoptist [named person] today. Great service just so sad that waiting list for [named surgeon] so long :-('*

*Waiting at [@named hospital]- appointment was over 2 hours ago. can we get takeout delivered??*

### **Effectiveness**

Tweets about effectiveness were hard to identify. The majority classified here made an explicit comment about whether a treatment had worked or not, or some aspect of the technical performance of the hospital in the process of provide care. Of the tweets about effectiveness 69% (20/29) were positive, 31% (9/29) were negative.

*'@worldmaliaday @willscomms a huge thank you to [@named hospital].. for their speedy and effective diagnosis of malaria yesterday. What a coincidence.'*

*Horrific exp [@named hospital].. Nurse had no respect for me, didn't speak to me & I'm discharged home with the same thing i came in with!!*

## **Safety**

Tweets that related to safety were even less common, often referring to perceived medical error. All (4/4) were negative. However, they did contain useful information for an organisation seeking to improve care.

*[@named hospital]Yes pls. Main concern now is the doctor overprescribing. We worked out the error but vulnerable patient might not*

*[@named hospital] Also looking at a scan from 2010 when u didn't get scanned until 2011 not good, wrong person, terrible, disgusting*

### **6.3.4 Other themes found in the qualitative analysis**

#### ***Fundraising***

The most common use of Twitter related to NHS hospitals in our study was fundraising and charitable activities. This accounted for 39% (387/1000) of all tweet. This included people seeking funding, describing their experiences of charitable activity, supporting others and advertising fundraising events. 55% of these tweets had a positive sentiment. Many of these

tweets also involved celebrities – either taking part in the discussion or asking for their support and endorsement.

*'Looking forward to running the Royal Parks Half Marathon for [@named hospital] this wknd'*

*'@piersmorgan: Just a reminder, Shuggsy, of our £20,000 bet (to [@named hospital]) on who, Arsenal or Spurs, finishes highest in League.. Wow!'*

### ***Promotion/Advertising***

There were many examples of people using Twitter for promotion (accounting for 9% of tweets), both of themselves, their colleagues, and organisations. People have previously suggested that Twitter is widely used for self-promotion, and our study certainly found evidence of this. From an analytical point of view, this makes attempts to examine quality harder. Promotional tweets often look similar to praise from the public – but they may represent a biased view that might not be useful for monitoring quality performance. 50% of these tweets had a positive sentiment, and 50% neutral.

*'Was a privilege to be exec champion for the excellence in customer care award at last night's [@named hospital] staff awards. Early night tonight!'*

*[@named hospital] Named One Of The UK's 40 Top Hospitals for the Fifth Year Running [weblink]*

### ***Health information***

There were occasional examples of people seeking medical guidance, and of health advice being provided (4.4% of tweets). We did not observe specific medical information being provided for individuals, instead the guidance tended to be generic health promotion advice that might be relevant for the public. 86% of these tweets had a neutral sentiment. The questions asked by the public could be more specific:

*'The sun has its hat on but have You? Remembering sun rules has to be better than a visit to [@named hospital] ED..ouch! [weblink]'*

*'[@named hospital] how much do you think dairy products are a contributory factor towards asthma'*

### ***Organisational information***

There were also examples of tweeting functional information about health services, to keep patients up to date about the health services - accounting for 7% of tweets. Within the organizational information, there were examples of hospitals keeping their patients up to date about their services and community activities. These were often tweets by hospital staff, or retweets of hospital tweets. 76% of these tweets had a neutral sentiment. There was also a subgroup of people talking about healthcare reform.

*'[@named hospital] Hourly Update! 11 cases this hour, 162 cases in total, 17 currently in department #[named hospital]'*

*'RT[@named hospital]: Most outpatient clinics running today We'll be reviewing afternoon clinics & if any cancelled, we'll contact patients'*

### ***Support for patients***

Some tweets (11%) were simple messages of support for patients in hospital. These were often personal, individual messages. Some however were large organised campaigns that had grown virally and contained numerous retweets with celebrity endorsements. 60% of these tweets had a positive sentiment. This tweet is an example from this sort of campaign:

*'RT { @namedperson}: He will remain in the ITU [ @named hospital] he's very poorly but he's a little fighter #staypositive #thankyou'*

### ***No theme***

It was notable that a moderate proportion of the tweets (20.1%) contained no identifiable theme. This is because Twitter messages are short and often part of longer conversations which cannot be understood out of context. Most of these tweets were neutral in sentiment (63%), but some however contain positive (31%) and negative in tone (6%), although it was not possible to tell what it was about.

### ***Multimedia approach***

People are creative in how they used Twitter messages - there were examples of people including other media in their tweets. Twitter allows users to embed images and videos - and some patient had adopted this approach. This was particularly common when talking about the facilities. For example, a concerned mother tweeted about the food that had been given to her young son – accompanied by photo:

*[@named hospital] Cannot believe I have been served this for my 18 month old.*

*Tastes disgusting and hardly nutritional! [link to photo]*

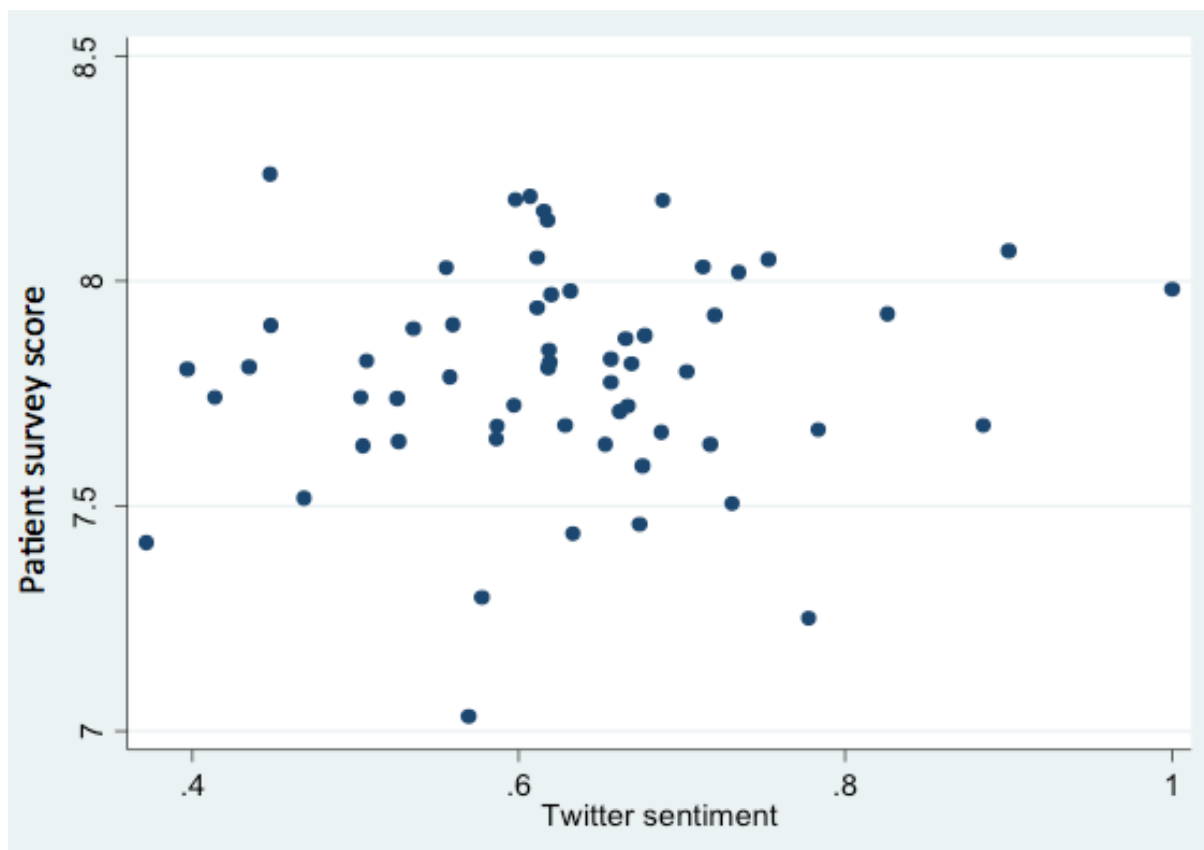
**Figure 30: Example of a tweeted picture**



### **6.3.5 Associations with conventional patient experience**

The Spearman rank correlation coefficient between the NHS inpatient survey and the automated Twitter sentiment analysis score was 0.08 ( $P = 0.56$ ), demonstrating no significant association (see Figure 31). The agreement between manually rated sentiment and automated sentiment analysis was 71% (for classification as positive, negative or neutral). This represents a Kappa statistic of 0.39 ( $P < 0.001$ ).

**Figure 31: Trust level association between Twitter sentiment and the NHS patient survey**



## 6.4 Discussion

These results show that people do talk about hospitals on Twitter, and they sometimes talk about care quality too. However, tweets that describe care are in the minority (only a little above 10%), and a wide variety of other topics are being discussed as well - not least charitable and fundraising activity. That said, when patients do talk about care, they talk about many of its aspects, and provide information that most hospitals would want to know about – particularly their experiences of staff, facilities and processes of care. In addition, the volume of messages (a mean of almost 2500 per trust - far more than the number of responses to the national patient survey) demonstrate that there is a substantial potential

body of information to be listened to. The scale of use is also increasing. At the end of the period, I re-performed the search for NHS trusts on Twitter and found that 133 (80% of 166 Trusts) were now using a Twitter account (up from 75 at the start of the experiment). This represents a 77 % increase in trusts using Twitter over the course of one year

Despite this volume, the distribution of tweets was not even between hospitals. A limited number received far more than most, with just three trusts accounting for more than half of all tweets. In particular, children's hospitals and specialist cancer hospitals received many more, which may be related to their more active fundraising efforts. Tweets were also distributed unevenly across days. It is noticeable that on those days where Twitter volume was particularly high, there was some form of celebrity endorsement going on.

The results show that the majority (77%) of tweets about quality were positive, similar to the figures for online reviews on websites in Chapter 3. I did however find the many of the tweets on non-quality related topics were positive also, potentially skewing the effect of looking for sentiment about quality as other topics are included too. If Twitter is to become a useful sensing device for quality issues, then systems will need to be developed that can tell the conversations about quality apart from the other ones. We will also need to develop improved sentiment analysis techniques for examining Twitter data, as this was not as accurate as the sentiment analysis of free text comments in Chapter 5. The difficulty of untangling the tweets about quality from the others is illustrated in the quantitative analysis, where there was no link between sentiment of tweets and patient survey measures of experience at the trust level.

#### **6.4.1 Comparison with previous work**

There is comparatively little work to compare these findings with. Lagu and colleagues, looking at the longer comments about hospital on the NHS Choices website is perhaps the closest (192). There are some similarities with their findings: they also found comments about staff were common (mentioned in 90% of comments) and that comments tended to be positive (83% favourable). They however found that comments about technical aspects of care (what I have measured as effectiveness and safety) were included in 62% of reviews, far more than I found in tweets. Similarly they found that 36% of comments identified perceived medical errors – in Twitter this was 4%. They saw comments about facilities in 52% of comments – on Twitter it was 10%. This might suggest that in the limited format required by Twitter, people talk about the thing that matters most – and this most commonly is the human relationship with staff. This is in keeping with other work looking at patient surveys where interpersonal relationships seem to be the key factor behind overall perception of quality in healthcare (120).

#### **6.4.2 Unmoderated comments**

One of the peculiarities of comments left on Twitter when compared to physician or hospital rating websites is that there is no moderation. The consequences can be seen in some of the comments left. Many of the comments refer directly to staff members – with both good and bad comments aimed at particular individuals. People are willing to be direct, sometimes even rude:

*[@named hospital] [Named chief executive] should come down onto the wards n see what's really going on under her nose. I wish my nan was in [another hospital]*

*[@named hospital] Shit on floor wet sheets, visitors having to change bedding, shit in toilet, ignorant staff- Saunders ward!! Stay away !*

These comments would be unlikely to have made it past the human moderators on NHS Choices – being too close to the ‘poison pen’ that many physicians fear. However, it true, this is useful and important knowledge for a health service provider to learn from and would not be heard on other more filtered platforms.

#### **6.4.3 The @piersmorgan bias**

The data here demonstrate the frailty of social media as a sensitive source of real time information. External and irrelevant things can have an effect. In the analysis, I found that almost 30,000 tweets referred to Piers Morgan (@piersmorgan) – a celebrity and television presenter. Almost all of these appeared to be related to charitable fundraising for great Ormond Street Hospital. Similarly, other spikes were related to activity around celebrities discussing hospitals.

Away from celebrities, some comments also addressed issues that the broader NHS might consider frivolous:

*[@named hospital] A cappuccino should be 1/3 espresso, 1/3 milk, 1/3 foam not 2/3 milk. If I wanted a latte I'd ask for one'*

Although unrepresentative, this comment seems to push the idea of consumerism further than many working in the NHS would be willing to accept. Such comments have little to do with the clinical services provided by the hospitals. However they still counted in our overall results. These comments that relate to less relevant aspects of care, taken together with celebrity driven activity and the large group of self-promotional tweets demonstrate the fragility of social media as a data collection system, vulnerable to amplification of potentially low relevance topics.

#### **6.4.4 Strengths and limitations**

With a comparatively new technical approach, there are limitations. I am uncertain about the completeness of the information from the Twitter API. This is the only publically available central Twitter source, but the way the information is published is not entirely transparent, and there may be filtering methods in place that I am not aware of.

We looked only at messages sent one way, to the hospitals. Twitter enables people to have conversations. A more wide reaching study might attempt to look at the nature of the interaction across multipart conversations, and could also potentially look into whether comments about quality lead to quality improvement activity.

The study may also have missed many important conversations. I have been limited to these situations where people have deliberately targeted messages at hospitals. Many people using social networks might not think to do this. In many cases, people may refer to hospitals by their informal or local names, rather than the formal names of their local

hospital trusts (many trusts contain more than one hospital, while the public may be more inclined to think in terms of individual hospitals).

There were biases in the information I collected. There are demonstrable biases in the people who use Twitter. Its demographics have been shown to be larger in younger groups and minority ethnic populations (289). However, these groups have often been hard to engage in other participatory healthcare activity, so this might represent an opportunity for wider population engagement with healthcare services.

Strengths of this study include its novelty, the prospective nature of the data collection and the attempt to gather data over a prolonged period of time for all the trusts on Twitter in the country. The relatively high kappa coefficients for human inter-rater reliability for thematic and sentiment measurement also provide some reassurance about the qualitative analysis approach. The similarity of hospitals using Twitter compared to those who were not using it (at least according to the four characteristics we measured) is reassuring, in that it implies we were not dealing with a significantly biased sample.

#### **6.4.5 Conclusion**

Twitter is certainly a useful tool for signal detection in many fields as varied as marketing, film and even disease detection. However in this case I have demonstrated some of the risks and limitations of the approach when looking for information about care quality. Contained in these tweets there appears to be information that would be useful for monitoring performance of hospitals, and possibly for an improved dialogue between the hospital and their patients, but they are in the minority. Some patients are willing to talk openly, and in

the most frank terms. At the same time, using current techniques, I could find little correlation between tweet sentiment and traditional patient surveys. The information contained within tweets is not distilled down, but rather in a rough, unrefined format that will need to be adequately processed and filtered before it can be useful.

## **Chapter 7: Discussion**

This thesis has looked at the topic of online feedback in healthcare from a number of different angles, with a variety of different observational and experimental approaches. To summarise and consider the findings, it is useful to consider what I found from each of the analyses in turn, and then consider how these findings relate to one another, and where this directs us in terms of advice for policy, practice and future research. My reflections on study limitations and relationship with existing literature are included in the individual chapters.

### **7.1 Key findings of the thesis**

#### **7.1.1 Analysis of comments of the NHS Choices website and comparison with other quality metrics**

In Chapter 3, I showed that online rating websites are being used – with comments about all hospitals in the country and the majority of general practices left on the NHS Choices website. The ratings and reviews are largely favourable (with 68% of hospital reviews and 69% of GPs positive), and are associated with other measures of patient experience and care quality. That said, they remain relatively rare activity, with only 0.04% of hospital admissions and 0.005% of GP consultations being reviewed online. These findings, when first published, represented the first national review of an online ratings system. These results suggest that discretionary patient ratings, solicited through a website, may be a more useful tool than previously considered, both for patients and healthcare workers. If patients are making choices based on this information, they can be reassured that the

ratings are not entirely misleading, and may be providing relevant information about healthcare quality. The signal that they produce is not just noise.

Patient ratings in primary care, where the numbers of rating per practice are low, may be less immediately useful as a measure of quality in general practice but they may provide a novel route for organisational learning. However, as numbers of ratings rise, it is likely that they will become a more useful tool for patients to make informed choices about where to receive their care.

Although I appreciate that there are inherent risks in the use of ratings from a small group of self-selecting patients, I also see that there are advantages of using this new form of information compared to using the traditionally survey approach -- in particular its lower cost and its ability to detect episodes of poor care that a random survey might miss.

Although online feedback mechanisms should not replace patient surveys, they may provide information on care quality that is relevant, complementary to survey data, much more timely and potentially useful for patients when making choices about their healthcare.

There is a background history to asking the '*uneducated masses*' about their views. Crowd sourcing, while very much an idea of our current time has its roots in figures such as the Victorian polymath, Sir Francis Galton. Not content with inventing the weather map, or doing much of the early work in establishing the notion of correlation, Sir Francis stalked the village fairs of 1860s England. He was no agricultural expert, but he won numerous 'guess the weight of the bull' competitions by simply asking as many people as he could at the fair to take a guess at the weight, then entering the arithmetic mean of these guesses as his

answer. Similarly, in his book *The Wisdom of Crowds* (2004), James Surowiecki argues that a diverse collection of ‘independently deciding individuals’ is likely to make better predictions and decisions than single individuals or even experts. In this case, at least to an extent, the crowd of patients appears to be wise.

### **7.1.2. Discrete Choices Experiments**

In Chapter 4, the discrete choice experiment showed that online reviews affect the way people pick their hospital in hypothetical experiments if they see this information. However, this effect may be mitigated in reality through lack of awareness of online feedback systems. These results show that, when placed side by side with other metrics, the effect of online recommendation is as large as that of recommendation by friends and family, with an average member of the public willing to wait 5 weeks to go to a hospital with good reviews compared to bad ones. What is more, the effect seems to be a positive driver to pick a hospital if reviews are good, unlike many other factors that tend to make a patient less likely to pick a hospital if they are bad. Online reviews remain subject to many criticisms from professionals, and our hypothetical experimental design limits our confidence in extrapolating to real world effects, but it is possible that at least some patients are using them to select hospitals.

The survey results show, however, that online reviewing in healthcare is not a common activity yet. Only 30% were aware of these websites, 12% had looked for information on them and 5% of our sample had provided a review of their own. I also demonstrated that behaviours of this sort do vary with individual characteristics, in particular they are more

likely to be used by younger groups, those with higher educational attainment and those with non-white ethnicity.

If practices from other industries— where consulting online reviews is a behaviour practiced by the majority – extend over to healthcare, it can reasonably be expected that their influence in healthcare to grow. The new and sometimes disruptive way we behave on the Internet has changed the way we do many things in life. These results add to a conclusion that our approach to hospital choice is no exception. I also produced survey results on attitudes towards choice (87% think it important) and knowledge about choice (66% are aware) that renew our understanding of these issues at a national level.

### **7.1.3. Sentiment analysis of free text comments**

The data presented in Chapter 5 demonstrates that sentiment analysis of patients' comments about their experience of healthcare is possible (we were able to predict attitudes towards three aspects of care with at least 85% accuracy), and that this novel approach is associated with patient experience measured by traditional methods such as surveys. Although at an early and experimental stage, it presents future possibilities to understand healthcare system performance in close to real time. Our work in this area was not ground breaking from a technical perspective – others have used more sophisticated algorithms for machine learning tasks – but for the first time I have shown that this approach produces results that map to those seen in the real world. Bates and colleagues have described the confluence of patient centred care and social media as a 'perfect storm'

that is likely to be of major value to the public and to healthcare organisations (291). These early findings hint at how that might occur.

#### **7.1.4. Analysis of Twitter information about hospitals**

Our analysis of tweets about NHS hospital in England shows that people are talking about their care on Twitter. I measured almost 200,000 over the course of a year. This is more than the number of responses to the NHS inpatient survey over the time period – and I only captured a fraction of what people were saying because I only looked at conversations that had been explicitly labelled. The results show that some tweets contained fascinating insights about care quality (around 11% were on this topic). However, the majority did not, but instead reflect other activities that relate to the hospital – and sometimes only tangentially. The most substantial proportion was about fundraising (39%), and other aspects of supporting patients receiving care. Activity by external factors – not least celebrities – can significantly skew results of sentiment around a hospital. The results demonstrate sentiment analysis of tweets is possible (but less accurate than for longer free text comments), and finding the right tweets to put through the process is even harder. The limited length and odd dialect and grammar of tweets make understanding their meaning harder. In other settings, for example trying to map flu outbreaks compared to real world data from surveillance systems, Twitter has proved itself to be timely and accurate. However, when looking at issues of care quality it does not appear to fare so well. There was very little correlation between tweets and national patient surveys. Finding the correct information from Twitter about hospital performance is not quite finding a needle in a haystack, but it will be a difficult task. If this information is to become more useful, we need to improve the tools we have to process and understand it.

## **7.2 Future policy and practice: Online ratings and reviews**

The findings in this thesis have practical implications. Online descriptions of care have the opportunity to provide a new way to look at quality, a new impetus to drive care improvement, and a new stimulus to hospital selection. However there are also real barriers to their use and benefit.

### **7.2.1 The policy enthusiasm for online rating by patients is set to continue**

The NHS in 2013 has gone through the most sustained set of criticism and controversy about the quality and safety of care provided since its foundation. The call for new and better ways to understand and measure quality are manifest in several reports, including Darzi (103), Francis (133) and Keogh (292). A desire for better quality is at the top of the political agenda (104) .

At the same time, there is no shortage of data with which to make judgements about performance of health systems. New data sources are becoming available: the Friends and Family Test started to report monthly patient experience in July 2013 (129), PROMS data has been collected in some disciplines for more than 2 years in England (132), granular data at the level of the surgeon's performance is making its way into the public domain (94). However there is a risk of being swamped. Meyer and colleagues have described the confusion this plurality introduces, and argued the need for parsimony in the quality measurement approach (107). The question for the policy practitioner - whether the civil servant, regulator, hospital manager or concerned clinician - becomes how to synthesise this multitude of important, but often contradictory information together.

Darzi and Donaldson describe a core triad of characteristics that quality indicators must have if they are to be usable. They must be 1) acceptable to clinicians, 2) collectable and 3) understandable by the public (209). They describe this as “a simple triad that is hard to reconcile”. The information provided online on rating sites studied here is not perfect as a quality indicator. No single quality indicator is. It still bumps into professional resistance, but is certainly collectable, and it seems to be understandable to the public – possibly at a uniquely engaging level compared to other indicators. It also provides a broad based view of care, covers the issue that patients care about – particularly communication with care workers - and it does this in a cheap and quick way.

### **7.2.2 Potential for online rating to improve quality**

#### ***Online rating is a potential new way to drive selection***

In the introductory chapters, I described the increasing role of public reporting in many health systems, not least the English NHS. We saw that while a great deal of hope had been entrusted in the idea, the size of the effect it has, and the route by which this effect occurs, is less certain. Our findings raise questions about whether online reviews and ratings deserve a more prominent place in public reporting systems. The findings from the discrete choice experiments show that online reviews have an effect on decision-making when put alongside other data. The associations with other quality and patient experience measures hint that the information contained within ratings and reviews is valid. However, the survey results also show that the public is not widely aware of them yet.

Our results suggest that the ratings are more likely to illustrate patient satisfaction with their care than dissatisfaction, and that they do correlate with some generally accepted aspects of the technical quality of care provided and the consequent outcomes achieved. Structure ratings and comments online may be a more useful tool than previously considered for both patients and healthcare workers. I suggest that unsolicited patient comments may have an important place in national quality and safety measurement systems.

We know that hospital choice is complex (6). The way we pick hospitals may have more in the common with the way we pick our football team, than the way we pick our new washing machine. It is complex, emotional and sometimes-even tribal – depending on our own experiences and those of others around us. The selection pathway from Berwick's drivers to improved quality has been lying dormant – or at least underpowered (7) - particularly in the less commercial and less choice defined English healthcare marketplace. Despite the uncertainty around the effect of choice and competition, the NHS in England has embarked on a course that places these two ideas as central to its future. Can a focus on online reviews and commentary reignite the selection pathway as a mechanism to improve quality? At present, this remains unknown, but the evidence hints that it might have a role – as a simple, user generated and understandable metric for comparison. It certainly has done in other industries (203), where the TripAdvisor model is often criticised, but its power to persuade the customer is clear.

### ***Online rating is a potential new way to drive improvement***

There is a profound problem in quality improvement in terms of converting measured poor or mediocre quality into action to improve care. In Chapter 1, I described the work done looking at the national inpatient survey, and how often it has often failed as tool to drive improvement. In the UK, crisis and scandals seem to occur which, in retrospect, had been seen and noticed – as exemplified by the events at Stafford Hospital. Just not by the people running the organisation, or in good time. The effect was that the public's voice was not acted upon. Only later, when the scandal reached the newspaper front page, does any action occur. Online comments provide the potential for a genuine dialogue with patients. The setting is less formal or accusatory than a complaint. It is less rigid, delayed and academic than a survey. It is more in keeping with the new behaviours we have developed in a more connected world online. Organisations and individual clinicians need to learn to use this online commentary as a part of a constructive feedback loop. In this way it might have the opportunity to drive targeted, focussed and local improvement activity. This is aspirational. We need to understand the mechanism through which online information can be transformed into quality improvement action in hospitals or other settings – and this is no easy thing.

#### **7.2.3 Do advantages outweigh disadvantages?**

These findings add to the limited evidence base in the field of online feedback. In the introductory chapters, I noted a series of arguments both for and against the use of this type of information – many of which were poorly evidenced. I hope that our findings

demonstrate useful new information to support or refute these arguments, and that a number of the criticisms levelled against online healthcare initially are mitigated (Table 31).

Our results do not upend all of the arguments against the use of patient website reporting, including that such reports come from a biased selection of patients, and that they carry the risk of harming doctor-patient relationships. I also note that hospitals may try to encourage particular patients to rate their care online - this adds an element of management to the self-selecting group of raters and may lead to a more complex selection bias.

**Table 31: New information about advantages and disadvantages of online feedback**

| <b>Disadvantages</b>         |                                                                                                                                                                       |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Selection bias               | Evidence that a selection bias does exist, but that online reporting is better used by young and minority ethnic groups. It is less likely to be used by the elderly. |
| No relationship with quality | Evidence that comments online are associated with clinical quality and other measures of patient experience, at both primary and secondary care levels.               |
| Poison pen mentality         | Evidence that this is rarely the case. The majority comments are positive.                                                                                            |
| Risk of gaming               | Not examined here.                                                                                                                                                    |
| <b>Advantages</b>            |                                                                                                                                                                       |
| More easily understandable   | Online ratings and reviews are at least as well used as other measures such as friend and family recommendation in choosing hospitals.                                |
| Already being used           | Evidence that usage of ratings is relatively widespread, if not a common activity.                                                                                    |
| Feedback drives improvement  | Not examined here.                                                                                                                                                    |

### **7.2.5 Issues to be overcome: Resistance, engagement and comparison**

Despite the various potential advantages I have considered to online feedback systems, there is much that needs to be corrected or improved. In particular, there are issues about continued professional resistance, wider engagement of the public to highlight the potential benefits thus encouraging them to contribute more of their own data, and making information across these websites comparable.

Even though the development of NHS Choices was aligned with high-level government policy across administrations (89)(91), there were continued voices of challenge at various stages of implementation of the various platform that have developed in the UK. In particular, there was hostility from the medical profession about the threat of the 'poison pen' and as a challenge to their notions of accountability and autonomy. Any attempt to build on the success of NHS Choices, either to improve the system in England, or to adapt a similar approach elsewhere, should learn from some of its lessons so far by expecting this challenge. Further attempts will have to tread a perilous tightrope, cautiously balancing the concerns of clinicians with a drive for transparency. They will need to work carefully with clinicians to explain the benefits and mitigate concerns, and develop strong systems to minimize the risk of defamation and gaming.

Although NHSC is used by a substantial number of people, a person choosing to contribute an anecdotal report about their care is still a relatively rare event. In comparison to other industries, healthcare has been slow to pick up on this phenomenon, and it has not yet become normal consumer behaviour in England. More focus, and possibly financial resource, may need to be placed on raising the profile of online reporting functions. This

might encourage both more and a broader variety of people to leave their comments. It might also result in more people using the information to help make decisions about care. As with any 'crowdsourced' resource, the larger the crowd, the more useful the information. NHSC has found that by making its data available publically (via their website, through APIs and by providing the data at health hack days), they have spawned a number of independent academic analyses in both the US and UK, as well as apps and data 'mashups' designed to increase public use (293). This suggests that sharing the content allows others to creatively use the information and maximize benefit beyond its initial aims – and may present a useful behaviour for others attempting similar systems.

There is a need to standardise aspects of methodology and approach. The data about care reported online is at its most usable, by both the patient and the organisational audience, when it can be compared and contrasted between providers. The government's new Friends and Family test is being widely criticised because there is no consistent approach to collection of data (130). NHS Choices provides a useful platform for measurement and comparison precisely because it is rule based – with a consistent set of questions asked nationally. In many other settings, particularly the unregulated online physician review market, structures of information are not set – leading to complexity and uncertainty in comparison. By setting fixed questions or approaches, or by drawing in views from as wide a variety of sources as possible and standardising the answers, a broader mix of voices can be incorporated into a system and heard. The new customer insight platform, to replace NHS Choices, will apparently build on these ideas to allow comments from multiple sources to be incorporated (216). The existing plurality of online review sites in the market, and the

variety of questions they ask, shows that trying to achieve consistency will be hard.

However, there may be simple overall questions, such as 'would you recommend this provider to other users', which could be suggested as a top level comparison metrics to be adopted by all those working in this space.

There is still a potential problem of gaming. Staff may write positive reviews about their own workplace, or deliberately harm the reputation of a competitor with false negative reviews. Little evidence of this in healthcare has been demonstrated so far, but I am aware of its prevalence in other industries – and as usage of online reviews increases, so does this risk. It is therefore important that those providing review services do what they can to minimise this. This might include using an automated process to identify inappropriate reviews. Although this has been demonstrated as theoretically possible for hotel reviews (208), it is not at a practical level in healthcare yet. Other potential technical solutions include preventing multiple comments from a single IP address, or simple manual checking. None of these are going to be perfect. At the moment many systems rely on human checking of comments prior to detect suspicious patterns of behaviour such as multiple similar reviews of one location, leaving it susceptible to potential abuse. As this risk is real, both sets of users – patients and organisations - will have to retain a degree of caution in interpreting the data available.

#### **7.2.6 The consequence of selection bias**

Perhaps the central limitation to online feedback as it stands, with relatively low levels of use, is the effect of the selection bias from those choosing to use these feedback systems. In

this work I have demonstrated that systematic biases in use do exist, and that there is lower usage in traditionally disadvantaged groups - notably those with lower socio-economic status and those who are older. This presents a real threat to equity in the use of the information, as there is a risk that certain privileged views will be heard louder than others, creating a service responsive to the of only part of the population. If we are to build a health service that is appropriate and responsive to the weakest in society, and that will act to reduce inequality not increase it, those people learning from online feedback will need to tread a cautious line – ensuring that as many and as diverse a set of patient voices are heard. This might require particular action to publicise this online feedback route to marginalised communities, or to provide other routes for feedback via conventional, less technology-focused methods that might have wider reach.

#### **7.2.7 Use and adoption of online feedback**

As momentum develops around the need to capture the patient's experience, these ratings represent a potentially valuable source of information about quality of care, when taken with other more conventional measures. I suggest they can be used as a complementary lens to look at the issue. Therefore, as such ratings become more common, health care providers and purchasers should promote the development of such online reporting schemes and encourage their use. Similarly, hospitals, family physicians and practices should develop strategies to respond to these comments in constructive ways.

Currently online reviews are not incorporated in many performance comparison systems or report cards around the world (although they are in England on the NHS Choices system).

Some have called for the adoption of an NHS Choices style system for collecting and reporting online feedback – particularly in the US where it has been argued that patients' comments and ratings should be included on the Hospital Compare website. Lagu described this as 'putting the public back into public reporting' (2). The findings in this thesis do little to support the arguments against this approach, and go somewhere towards endorsing those in favour. In those environments where online feedback systems have already been adopted, future policy should focus on how to take this information and make sure it is accurate, understandable, comparable and usable to its various audiences.

### **7.3 Drawing all this information together to create a 'Cloud of Patient Experience'**

The second part of this thesis has been more experimental. In Chapter 2 I set out a hypothesis - that comments left by patients on the Internet across a wide variety of platforms could be collected and analysed. However, the results of the experiments in Chapters 5 and 6 show that trying to collate and measure the 'cloud of experience' is difficult. I have demonstrated here that some of the ideas are possible and feasible, but that there is still some way to go before we have usable data. As I mentioned in the introduction, the process of harnessing this cloud will be divided into different tasks: 1) finding and collecting the right information and then 2) processing it to distil it down to a usable form.

In Chapter 5, I have seen that the processing part may be easier. Sentiment analysis can be done, and reasonably accurately. Chapter 6 showed us that the task of selecting the right information from the huge canvas of the Internet is harder. Our analysis of Twitter data

suggests that there is fascinating and relevant information about care quality contained within the data, but that it is not abundant. There is a lot of information there, but it is unstructured, poorly labelled and difficult to attribute to a particular care provider. It is also hidden within a larger set of less relevant information. The vagaries of what is happening elsewhere on the Internet can also have an effect on what people are saying and bias results.

Even if both these tasks could be achieved, testing the hypothesis as described – that aggregated social media data could be used to spot hospitals providing poor care - would be hard. I doubt that aggregated real time patient feedback would be an immediately effective test of clinical performance for an organisation. However, considering the lack of engagement with, and the potential inaccuracy of, the current metrics - such as standardised mortality ratios - I believe there is a need to explore alternatives. The soft intelligence provided by this proposed approach – capturing and processing the cloud of patient experience - could provide a further way to look at the patchwork quilt of health quality. In a time where regulatory organisations are stretched and struggling to complete their basic workload, (294) the ability to biopsy the patient voice in real time and without expensive new infrastructure is appealing. Furthermore, taken in conjunction with measurement of other metrics of patient outcome, it is possible to imagine a national early warning system that would highlight poorly performing hospitals faster than is done at present.

Eventually, there might even be the potential to develop a close to real-time early warning system for poor clinical care, if large enough amounts of data can be collected and

prediction accuracy can be reliably reproduced. Ultimately, this early warning system could become part of a national, real-time monitoring tool for hospital performance.

I appreciate there are substantial technical and logistic problems with an idea like this. The analytic component is also a technical challenge: machines struggle to read and understand comments accurately; software finds comments preceded by negatives difficult to interpret. Sarcasm and irony are almost impossible to process. As with any sensing system, it will have biases. For example, there is a considerable potential for a selection bias as patients who choose to talk about their care online may not be representative of those attending health care facilities. Data suggest that use of rating websites and Twitter tend to be associated with higher socio-economic status and younger age groups (289) (Chapters 3 and 4)- and therefore hospitals serving certain populations may receive less attention on social media than others. Some sources of data will be easier to automatically process than others. The comments left on ratings websites, and those left on patient discussion fora, are rich in material directly relevant to their healthcare, but tend to be posted infrequently. In comparison, Tweets are necessarily brief, containing less contextual information about health care and are composed in their own particular grammar. While these characteristics present real challenges for auto-processing, the growing popularity of Twitter means it is essential that novel methods are developed to adequately capture information on health care from this sort of source. A further concern is the autocatalytic quality of some social media, where ideas are repeated on the basis of their popularity, reiterating and reinforcing views, rather than providing new information. I have included a list of potential data sources and their relative merits and drawbacks in Table 32.

**Table 32: Potential sources of information for the ‘Cloud of Patient Experience’**

| <b>Type of source</b>                       | <b>Examples</b>                                                | <b>Information that could be used</b>                                                                         | <b>Advantages</b>                                                               | <b>Disadvantages</b>                                                                                          |
|---------------------------------------------|----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Rating and feedback websites                | RateMDs (295)<br>Patient Opinion (210)<br>Iwantgreatcare (211) | Ratings and free text descriptions of healthcare providers and individual clinicians                          | Comments usually directly relate to care experience                             | Comparatively low usage, possibility of deliberate gaming                                                     |
| Patient networks, discussion fora and blogs | Patientslikeme (154)<br>Mumsnet (296)<br>Epatients.net (297)   | Patients and carers’ shared descriptions of their care and experiences                                        | Authentic voice of the patient, often well used in specific patient communities | May be a selection bias towards particular demographics (with higher socioeconomic status) or interest groups |
| Micro-blogs                                 | Twitter (298)                                                  | Tweets (short messages) directed towards hospitals or care providers                                          | High volume of traffic, often tagged with service they relate to                | Short, unstructured messages may contain minimal information about care quality                               |
| Social networks                             | Facebook (299)<br>Google+ (300)                                | Comments left on hospital or friends pages about care or specific signals of appreciation (e.g. likes, ‘+1’s) | High membership and usage by the public                                         | Public rarely talks about healthcare on these platforms<br>Content may be from employees rather than patients |

There is also a risk that any system like this will throw up false positives- and precautions would have to be taken to prevent wrongful reputational damage. It may require the bar for an alert to be raised to be set high initially, or that hospitals with low amounts of 'chatter' about them online are excluded until a certain amount of engagement is reached. Caution should also be taken before placing too much faith in a quantitative approach, as the qualitative analysis of information of this sort is known to provide useful insights (192).

Qualitative and quantitative approaches should be seen as complementary. However the

ubiquity of this data suggests that we should attempt to harvest it and translate it into something more meaningful - to see if we can pick out a signal from the noise.

Social media analytics and 'Big Data' are ideas that have been growing in the commercial management mind-set in recent years. The quality improvement movement in healthcare has often been at its innovative best when adopting ideas from other industries - be they the safety processes of aviation or the rigour and consistency of manufacturing production lines. Perhaps this field presents another opportunity to borrow and adapt the best ideas from other areas, and potentially develop a system to make the current wave of quality and regulatory failures less likely in the future. The fact that the idea of sentiment analysis of NHS comments is now being suggested at the national policy level (89) (92) suggests that this work may be put to a practical test sooner than anticipated.

## **7.4 Future research and questions to answer**

As I noted in the introductory Chapters, the research literature in this field is limited and undeveloped. As such, there are many other questions still to be answered, many of them surprisingly simple. Some questions are those that build on the work I have presented here. Others relate to important aspects that were beyond the scope of the thesis but are important. In the next section, I describe a number of the areas where research in this field could and maybe should go next.

#### **7.4.1 Comment and rating websites**

NHS Choices allows patients to rate their care at the hospital level. This is different to many websites, which rate individual physicians. It may be more difficult to develop insights into care quality when rating individuals, as the smaller number of ratings may preclude robust estimates of patient experience. More research comparing performance of individuals' physicians in terms of ratings online and through technical measures (for example the individual level surgical outcomes in NYC and now in England) would be interesting. However, this would be difficult to do in England, where individual rating is not permitted on NHSC.

Our work added to the complex and occasionally contradictory body of research that looks at the relationship between care quality and patient experience and satisfaction. Further research is needed to dissect the relationship between bad clinical outcomes and patient satisfaction in the context of appropriate medical care. In healthcare, as opposed to most other services with online rating, you can do all the right things, but still have a bad outcome. Having a clearer understanding of the interaction between clinical outcomes and satisfaction measures would be useful.

As a system that relies on patients volunteering to come forward to rate services, the nature of responses will also depend on how active the process of solicitation is. It would therefore be useful to understand why people choose to come forward, and what effect increasing the promotion of such websites would have on the nature of responses. This could include focus group work with patients who have chosen to leave comments, and those who have not. Quantitative experiment trial experiments could also be done to examine response

rates associated with different forms of reminder to leave commentary. Professional feedback industries in other sectors have maximised levels of feedback by sending out reminders at particular points after a purchasing decision is made – often 2 weeks. It would be interesting to see whether this might improve usage of a similar system in healthcare, or even change the nature of the responses.

A further question concerns that best way to screen comments prior to them being made public. If screening is to be undertaken, there are a number of ways this could be done. New comments could be published without review, reviewed by a person, or by an automated algorithmic approach, or a combination of both. NHSC uses a human approach, whereby every comment must be observed and approved by a moderator before it is published. On NHS Choices, the majority of comments (88% in 2013) are accepted. The moderator removes some comments (and the associated ratings) that are the most negative or personal. The disadvantage of the human approach is cost, and the delay in processing which means a comment posted does not come up on the website immediately. Popular social media sites – such as Facebook - tend to allow anything to be posted, but have a system for flagging content as inappropriate, which then moves it into a process including human review. However, given the potential for defamation or inappropriate reputational damage, this was considered too high risk despite the low cost. Even in the US where defamation laws are less stringent than in the UK, the need to prevent offence and defamation is clear and it would seem that having no screening would be a high-risk strategy. Further understanding of the effect of moderation policy and whether allowing un-moderated comments would produce stronger associations would be useful.

There has also been discussion about whether there should be a minimum number of comments required before they are displayed publically, to improve reliability and avoid bias. Anecdotal reports are potentially more useful when there are enough to present a clear pattern of performance, reducing the risk of single extreme comments producing an unrepresentative description. On the other hand, if you hold back patient comments until a set number have been gathered you effectively censor early commenters for an unknown time frame, something that was not seen appropriate in a 'patient first' healthcare system.

NHSC has no minimum number required, but others have suggested that there should be a minimum of 5-10 comments in general for health comparison websites (180). This is an issue for NHSC where all hospitals are covered, but volumes of comments are not always high (our work found a mean of 62 comments per NHS hospital organization (median 46, range 1-290)) over 24 months on NHSC. For primary care, looking at data for 15 months from the start of reporting, there was an average of 2.1 reviews per practice with 61% of 8089 practices nationally having a comment. Results from the US described previously show a similar partial coverage (only 20-40% of individual clinicians reviewed(188) (190) ). If a minimum number of comments were required, the majority of practices would not have their reviews made public. As many practices have few or no ratings online, work is needed to determine the numbers of ratings required over a defined period of time for a patient to obtain a reasonably accurate appreciation of the strengths and weaknesses of a family practice. New platforms may need to carefully consider their usage levels before deciding if a minimum level is viable.

In the discussion I have emphasized the role of online feedback as a quality improvement tool. It would be useful to do qualitative work with clinicians and managers of healthcare organisations to understand more about what dictates their response to online commentary, and what drives their actions. Earlier work qualitatively analysing the content of review by Lagu and Lopez (192,193) clearly demonstrated that these reviews identify actionable areas for improvement – in particular the finding that more than a quarter of reviews described perceived medical errors. We know that response rates to online comments are variable, even though all are offered the opportunity to respond. This qualitative work might include exploring the attitudes and concerns that providers and clinicians have, for instance their views on how it challenges autonomy. This area might benefit from case studies of organisations that are already doing this, and identifying this who are more hesitant or event resistant towards the idea. A survey of clinician and manager attitudes, to find the proportion of people in favour and against would also be useful. Understanding the level of clinical resistance will be crucial to considering the suitability and rate of roll out of the ideas in England and other health systems

#### **7.4.2 How do patients make choices with this information?**

If comments are to be displayed, one of their key purposes must be to inform the public and change the way they behave. The results of our discrete choice experiment give the first clear indication of this effect. However, there is still a lot we do not know about. More research is needed to understand how web-based patient ratings are used by patients in making choices about health care -- including how they reconcile them with objective clinical information such as outcomes and processes of care in real world settings- not just

the hypothetical choices I was able to measure. Further research might go beyond the hypothetical nature of stated preference, and look at revealed preference – the real life decision made. Normally this is made difficult because the circumstance they make their decision in- and the information they have inform of them at that point – are not controlled. However, the choose and book framework, which is currently being redesigned, could provide an opportunity to measure actual behaviour, together with information about what comparative performance data the patient looked at on their website. This could usefully be supplemented by qualitative research exploring patients' views about online content. This could look at some of the more nuanced ideas and expectations about rating care, and better understanding their decision making processes. For example, are people already using the highly nuanced approach to online commentary that is observed in travel decision making- and why it is that online ratings seem to act to 'pull' towards good reviews, rather than 'push' away from bad reviews? It would also be useful to repeat the questions asked in our discrete choice experiment with a nationally representative sample that were not all internet users, as this might give a more balanced picture.

How to present the ratings and reviews to the public is another question. Rating can be condensed into composite scores, but narratives are by their nature unstructured and hard to compare. They are therefore hard to display alongside quantitative data, particularly when data is presented in a comparison table between providers. The theory of bounded rationality suggests having too much information might make rational choice difficult (79). In addition, there is a risk that compelling anecdotal information might undermine more representative and validated information from other sources. NHSC experimented with a number of different ways to present the unsolicited patient comments and ratings with

other comparative data, such as hospital mortality rates and choose to deliberately simplify the information given on first view. Experimental work has demonstrated that ordering of information is important within the NHSC context, with the public paying more attention to the information presented first (301). This work also suggested that display of numerical information in simple evaluative labels (e.g. 'excellent', 'good' or 'poor') was preferred by patients, and that patients benefited from simple, easily recognizable symbols (e.g. ticks and exclamation marks), both of which have been adopted in the NHSC comparison display. It would be useful to conduct further experiment about how the presentation of information affects patient choice. For example, does ordering or framing of the information affect the various possible cognitive biases, and alter choice behaviour.

Further work could also explore the role of individual narrative comments compared to quantitative ratings. There is a difference between seeing a score out of five on TripAdvisor and reading the individual comments left by another user. It might be possible that the human, narrative element of reviews has a more powerful effect than a quantitative score. It would be useful to flesh out this difference by placing narrative comments as their own separate attribute in a discrete choice framework.

It would be interesting to understand whether information from trusted sources is more likely to be effective in influencing choices. For example, does a review by someone you know and like count more than reviews by an unknown stranger? Work in other social networks has suggested that this is the case, and that having a known peer make a choice makes others in their social network more likely to make the same purchasing decision (150). Does this phenomenon carry over to healthcare choices?

On a more commercial note (and outside of a price fixed system similar to the UK), it would be interesting to see whether online comments and reviews via social media allow hospitals to maintain market share at higher prices, as the traffic drives increased demand – as has been demonstrated in other industries (203).

### **7.4.3 The Cloud of experience**

There are a separate set of further research questions that relate to the more experimental task of trying to harness the ‘cloud of patient experience’ that was hypothesized earlier.

#### ***Improve sentiment analysis***

Further research is needed to improve the performance of sentiment analysis and opinion mining tools, extending the process to other forms of free-text information on the Internet and exploring the relationships between views expressed by patients' online and clinical healthcare quality. For example, several technical components might be added to improve the process, including the consideration of higher number n-grams (longer phrases). From a technical perspective, this might also include improved algorithms to detect the specific parts of speech, the ideas and concepts contained within them. This might allow refining contextual of polarity (understanding what a word or phrase means given its context in a sentence), and reduce the risks misinterpreting sarcasm and irony, often used by English speakers. It might even be the case that wholly new dictionaries of sentiment dependent words would need to be developed for different countries or regions. It is unlikely that words used to describe care in England and America have exactly the same meaning.

### ***Improve collection and filtering techniques***

If accurate sentiment analysis and opinion mining can be achieved, the other vital task is to collect and filter the correct information to put into this process. Our findings here show that this is difficult. Although useful information is there, it needs to be identified and sorted – the wheat from the chaff – if useful information is to be found. In this case, this will require the development of further intelligent scraping processes that are able to collect information, and particularly identify whom the information is targeted at. If the commentary online is to be useful at any local level, this will be necessary. It may not be as simple as identify '@ messages' which I choose to do. People describe their caregivers in many ways, and with many different names. Is the John Radcliffe hospital better known as the JR, JRH, the 'oxford hospital', '@OUhospitals', Oxford University Hospitals Trust, or any other name? For a detection system to work, this identification process will be needed, which may require the development of an entire lexicon of hospital and service names.

### ***Better understanding of Internet behaviours.***

There is a need to better understand the Internet, and how it is used for health. In the US, the Pew Centre measures online health behaviours regularly across a representative range of people, and the on-going nature of the study allows for changes over time to be looked at (302). In the UK we have to rely on intermittent samples from the social trends team (146) and ad-hoc surveys such as the one done in this thesis. It would be useful to incorporate some of these factors in to a regular national polling process. The Oxford Internet Survey has taken a lead in measuring other behaviours online, including information seeking; the addition of further regular health questions in this or some other survey would be beneficial for this field.

### ***Proof of concept: a social media 'smoke detector' for poor quality care***

In the introduction of the thesis I set out a hypothesis: that it might be possible to detect institutional poor performance via social media. To test if this is correct, one would need to identify future cases of poorly performing hospitals, and collect corresponding data on how patients described these hospitals on the Internet. One could then test the capacity of this novel data source as a predictive measure of poor quality care, calculating the sensitivity and specificity of the alert process in the same way as one might test the capability of a new diagnostic test. In addition, it might be possible to compare the timings at which the various signals of poor quality started to emerge, to see if this approach might give us a head start against conventional measures such as mortality rates. Retrospective analysis of previous care quality scandals is not possible, as use of social media has grown so rapidly and data would not be available for them. However, an analysis of the next set of scandal hit health organisations might prove a more constructive experiment.

## **7.5 New contributions from this thesis**

We hope that much of the work in this thesis has provided original knowledge. In particular, the demonstration of an association between online ratings and quality (Chapter 3), the usage and awareness of online rating websites, and their role in influencing hospital decisions (Chapter 4), the comparison of sentiment analysis results with patient surveys (Chapter 5), and the analysis of tweets for care quality information (Chapter 6) have not been demonstrated before elsewhere to our knowledge. Those parts of the thesis already published have been cited by others, including further studies comparing online reviews with quality measures (192) (245), attempts to develop better patient experience metrics (303) (304) and wider discussion of the role of patients in understanding the role of quality

(305) (194) (291). Subsequent work by others citing ours has also considered the role of online feedback in changing governance structures in public services (306), the challenges it poses to professional practice and ethics (307) (308) and the nature of the new patient economy that is emerging (153). Early findings were also included as evidence for broader attempts to encourage patient engagement online in the English NHS's national information strategy (89) and government ministers have cited the findings in public statements (309). Parts of the work have also featured in the national press – notably the relationship between ratings and mortality (309–311) but also the idea of sentiment analysis of online comments (312) .

## 7.6 Conclusions

The changes that I described in the first chapters are substantial. Each of the three drivers I considered— public reporting of performance data, a renewed focus on patient experience and the increasing reliance we have on the Internet - look as if they will be permanent and will affect the practice of medicine. The public is demanding transparency in all areas of government and the public sector – healthcare is no exception. Patient experience is increasingly important as we move away from mechanistic measures of structure and process, and think more about the end outcomes of the treatments that healthcare provides. The Internet is profoundly changing the way we live our lives. The health system of the future will have to respond to all of these cultural and societal shifts, and the topic of this thesis - patients describing their care online – may well be one of the first and most obvious exemplars of these changes.

I hope that I entered into these analyses with a sense of equipoise, keen to assess a new idea and understand new behaviours. What I have seen is clear evidence that this phenomenon is real, and taking a global view, evidence that this activity is increasing. The use of web-based patient ratings has become common in other industries such as hotels and restaurants, and consumers value these rankings in making choices. I believe that the information provided by these online feedback websites, although potentially flawed, represents an important new development in the measurement of healthcare quality. Anecdotal comments and ratings might represent an untapped seam of valuable information about service quality, and a potential hook on which to encourage patients to

engage with wider comparative data. They also provide patients with an immediate voice – something important in of itself in a healthcare system that aims to put patients first

The comparison with ‘TripAdvisor’ is easily made, but overly simple. While the relationship between patients and their health care providers is different than a simple consumer-purchaser relationship, it is clear that patients put stock in this information online and likely that it will be a feature of the healthcare landscape in the future. However, to expect patients to use this information in the same way as they might when choosing a hotel is to ignore a body of literature that suggests many patients are almost tribal in their allegiance to hospitals, and in which a consumerist attitude is far from the mainstream. Although we can learn many lessons from other industries, we must also accept the limitations of the comparison.

These findings show that despite the criticism, there are positive and hopeful aspects around rating care online. Comments online are overwhelmingly positive. And although the sample is biased, the bias is not quite what might be anticipated. They were well used by the young – traditionally a hard group to engage – and by minority ethnic populations- who have too often been at the wrong end of health inequalities. These approaches present new optimism in terms of broader engagement. Whether this engagement is enough to reignite Berwick’s selection pathway is another matter. Similarly we still have much to learn about how we can use feedback to drive improvement and close the feedback loop.

Rating and review sites are not without substantial risk. They present a profound challenge to existing notions of autonomy, hierarchy and have shaken many physicians, who struggle

to deal with a lack of control. They are exposed to potential manipulation. Online care reviews emphasise certain parts of quality over others. In particular the aspects related to communication and interpersonal relationships. But these have been hard to measure in the past, and provided they are looked at in parallel with other more technical information, this might present a strength of this new approach. Online care reviews also may neglect the elderly – although the use of other novel technology appears to be at least slowly broadening in older age groups.

The act of patients rating and reviewing their care online on structured websites appears here to stay – and to have immediate uses. That said, we are still some way from having an immediately useable ‘cloud of patient experience’. The work presented here is early and experimental. It might hopefully provide a part of a foundation for further exploration of this notion. These untapped pools of patient experience online will require considerable technical skill and ability to be corralled and processed into useable information. At the moment elements of the noise drown out this signal. But there is a signal there. And techniques borrowed from other industries give us clues about how it might be extracted, refined, processed- and eventually used as a channel for patients to become engaged with their health providers, as a source of valuable information for organisations and clinicians about how to improve and potentially as a ‘smoke detector’ for bad care.

Much technology is a solution looking for a problem. In this area, there is a large and clear problem. For once, the technology may have a real problem to address. Online patient feedback represents a possible new important source of information about quality that is too often uncaptured and ignored. It is quick, understandable to the public, potentially low

cost and already being adopted. The question now becomes one of how to maximise its usefulness and practical application, and less arguing about whether it should even be allowed. There is less to fear and much to learn about this increasing process of transparency in healthcare and the new technology that enables it.

## References:

1. Cabinet Office. Public Services: On your side. London: The Stationary Office; 2009.
2. Lagu T, Lindenauer PK. Putting the public back in public reporting of health care quality. JAMA. 2010 Oct 20;304(15):1711–2.
3. Kadry B, Chu LF, Kadry B, Gammas D, Macario A. Analysis of 4999 Online Physician Ratings Indicates That Most Patients Give Physicians a Favorable Rating. J Med Internet Res. 2011 Jan;13(4):e95.
4. McCartney M. Will doctor rating sites improve the quality of care? No. BMJ. 2009;338:b1033.
5. Bacon N. Will doctor rating sites improve standards of care? Yes. BMJ. 2009;338:b1030.
6. Marshall M, McLoughlin V. How do patients use information on health providers? BMJ. 341:c5272.
7. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med. 2008;148(2):111–23.
8. Public Reporting as a Quality Improvement Strategy: A Systematic Review of the Multiple Pathways Public Reporting May Influence Quality of Health Care - Research Protocol. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
9. Cacace M. How Health Systems Make Available Information on Service Providers: Experience in Seven Countries. RAND Heal Quaterly. 2012;1(1):11.
10. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The Public Release of Performance Data. JAMA. 2000;283(14):1866–74.
11. Davies H, Marshall MN. Public disclosure of performance data: does the public get what the public wants? Lancet. 1999 May;353(9165):1639–40.
12. Colmers JM. Public Reporting and Transparency. New York: Commonwealth Fund; 2007.
13. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003 Jun 26;348(26):2635–45.
14. Institute of Medicine. To Err Is Human: Building a Safer Health System. Washington, DC: National Academies Press; 1999.
15. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.

16. Atlas of Variation 2011. London: Department of Health; 2011.
17. Berwick DM, James B, Coye MJM. Connections Between Quality Measurement and Improvement. *Med Care*. 2003;41(1):30–8.
18. Hibbard JH. What Can We Say about the Impact of Public Reporting? Inconsistent Execution Yields Variable Results. *Ann Intern Med*. 2008 Jan 15;148(2):160.
19. Hibbard JH, Stockard J, Tusler M. Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts? *Heal Aff*. 2003 Mar 1;22(2):84–94.
20. Hibbard JH, Stockard J, Tusler M. Hospital performance reports: impact on quality, market share, and reputation. *Heal Aff*. 2005 Jan 1;24(4):1150–60.
21. Hibbard JH, Stockard J, Tusler M. Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts? *Heal Aff*. 2003 Mar 1;22(2):84–94.
22. Hsu J. The relative efficiency of public and private service delivery. Geneva: World Health Organization; 2010.
23. Weintraub ER. General Equilibrium Analysis: Studies in Appraisal. University of Michigan Press; 1985.
24. Vladeck BC, Rice T. Market failure and the failure of discourse: facing up to the power of sellers. *Heal Aff*. 2009 Jan 1;28(5):1305–15.
25. Nightingale F. Notes on Hospitals. 3rd ed. London: Longman, Green, Longman, Roberts, and Green; 1863.
26. Codman E. A study in hospital efficiency : as demonstrated by the case report of the first five years of a private hospital. Boston: Todd; 1918.
27. Chassin MR. Achieving And Sustaining Improved Quality: Lessons From New York State And Cardiac Surgery. *Heal Aff*. 2002 Jul 1;21(4):40–51.
28. Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA*. 1994 Mar 9;271(10):761–6.
29. Hannan EL, Kumar D, Racz M, Siu AL, Chassin MR. New York State’s Cardiac Surgery Reporting System: four years later. *Ann Thorac Surg*. 1994 Dec;58(6):1852–7.
30. Mukamel DB, Mushlin AI. Quality of care information makes a difference: an analysis of market share and price changes after publication of the New York State Cardiac Surgery Mortality Reports. *Med Care*. 1998 Jul;36(7):945–54.
31. Jha AK, Epstein AM. The predictive accuracy of the New York State coronary artery bypass surgery report-card system. *Heal Aff*. 2006 Jan 1;25(3):844–55.

32. Werner RM, Konetzka RT, Stuart EA, Norton EC, Polsky D, Park J. Impact of public reporting on quality of postacute care. *Heal Serv Res*. 2009 Aug;44(4):1169–87.
33. Mukamel DB, Spector WD, Zinn JS, Huang L, Weimer DL, Dozier A. Nursing homes' response to the nursing home compare report card. *J Gerontol B Psychol Sci Soc Sci*. 2007 Jul;62(4):S218–25.
34. Jin GZ, Sorensen AT. Information and consumer choice: the value of publicized health plan ratings. *J Heal Econ*. 2006 Mar;25(2):248–75.
35. Lamb GC, Smith MA, Weeks WB, Queram C. Publicly reported quality-of-care measures influenced Wisconsin physician groups to improve performance. *Heal Aff*. 2013 Mar;32(3):536–43.
36. Street A. The resurrection of hospital mortality statistics in England. *J Heal Serv Res Policy*. 2002 Apr;7(2):104–10.
37. Tabor LM, Torda P, Thomas SS, Zutz JL. Measuring Quality in the Early Years of Health Insurance Exchanges. *Am J Manag Care*. 2013;19(3):220–7.
38. Lubalin JS, Harris-Kojetin LD. What do consumers want and need to know in making health care choices? *Med Care Res Rev*. 1999 Jan;56 Suppl 1:67–102.
39. Vladeck BC, Goodwin EJ, Myers LP, Sinisi M. Consumers and hospital use: the HCFA “death list”. *Heal Aff*. 1988 Jan;7(1):122–5.
40. Mennemeyer ST, Morrissey MA, Howard LZ. Death and reputation: how consumers acted upon HCFA mortality information. *Inquiry*. 1997 Jan;34(2):117–28.
41. Longo DR, Land G, Schramm W, Fraas J, Hoskins B, Howell V. Consumer reports in health care. Do they make a difference in patient care? *JAMA*. 1997 Nov 19;278(19):1579–84.
42. Rosenthal GE, Hammar PJ, Way LE, Shipley SA, Doner D, Wojtala B, et al. Using hospital performance data in quality improvement: the Cleveland Health Quality Choice experience. *Jt Comm J Qual Improv*. 1998 Jul;24(7):347–60.
43. Bentley JM, Nash DB. How Pennsylvania hospitals have responded to publicly released reports on coronary artery bypass graft surgery. *Jt Comm J Qual Improv*. 1998 Jan;24(1):40–9.
44. Luce JM, Thiel GD, Holland MR, Swig L, Currin SA, Luft HS. Use of risk-adjusted outcome data for quality improvement by public hospitals. *West J Med*. 1996 May;164(5):410–4.
45. Romano PS, Zach A, Luft HS, Rainwater J, Remy LL, Campa D. The California Hospital Outcomes Project: using administrative data to compare hospital performance. *Jt Comm J Qual Improv*. 1995 Dec;21(12):668–82.

46. Schauffler HH, Mordavsky JK. Consumer reports in health care: do they make a difference? *Annu Rev Public Heal.* 2001 Jan;22:69–89.
47. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med.* 2008;148(2):111–23.
48. Mannion R, Davies H, Marshall M. Impact of star performance ratings in English acute hospital trusts. *J Heal Serv Res Policy.* 2005 Jan;10(1):18–24.
49. Closing the Quality Gap: Revisiting the State of the Science (Public Reporting as a Quality Improvement Strategy). Rockville, MD: Agency for Healthcare Research and Quality; 2012.
50. Berger ZD, Joy SM, Hutfless S, Bridges JFP. Can public reporting impact patient outcomes and disparities? A systematic review. *Patient Educ Couns.* 2013 [Epub ahead of print]
51. Rothberg MB, Morsi E, Benjamin EM, Pekow PS, Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Heal Aff.* 2008;27(6):1680–7.
52. Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D, Lytle B, et al. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation.* 1996 Jan 1;93(1):27–33.
53. Moscucci M, Eagle KA, Share D, Smith D, De Franco AC, O'Donnell M, et al. Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases. *J Am Coll Cardiol.* 2005 Jun 7;45(11):1759–65.
54. Joynt KE, Blumenthal DM, Orav EJ, Resnic FS, Jha AK. Association of public reporting for percutaneous coronary intervention with utilization and outcomes among Medicare beneficiaries with acute myocardial infarction. *JAMA.* 2012 Oct 10;308(14):1460–8.
55. Bridgewater B, Grayson AD, Brooks N, Grotte G, Fabri BM, Au J, et al. Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25,730 patients undergoing CABG surgery under 30 surgeons over eight years. *Heart.* 2007 Jun;93(6):744–8.
56. Dixon A, Robertson R, Appleby J, Burge P, Devlin N, Magee H. Patient choice: How Patients choose and how providers respond. London: King's Fund; 2010.
57. Bevan G, Hood C. What's Measured is What Matters: Targets and Gaming in the English Public Health Care System. *Publ Admin.* London; 2005;84(3):517–38.
58. The NHS Constitution the NHS belongs to us all. London: Department of Health; 2013.

59. Coulter A. Do patients want a choice and does it work? *BMJ*. 2010 Jan 14;341(oct14\_2):c4989.
60. Smith P. Market Mechanisms and the Use of Health Care Resources. *Achieving Better Value for Money in Health Care*. Paris: OECD Publishing; 2009.
61. Gaynor M, Town RJ. Competition in Health Care Markets. Cambridge, MA: NBER Working Paper No. 17208; 2012.
62. Mays N. Is there evidence that competition in healthcare is a good thing? No. *BMJ*. 2011 Jul 5;343:d4205.
63. Cooper Z, Gibbons S, Jones S, McGuire A. Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms. *Econ J*. Blackwell Publishing Ltd; 2011;121(554):F228–F260.
64. Bloom N, Hoxby C, Huckman R, Kessler D, McConnell J. The Impact of Competition on Management Quality. Cambridge, MA: NBER Working Paper; 2010.
65. Gaynor M, Moreno-serra R, Propper C. Death by Market Power Reform, Competition and Patient Outcomes in the National Health Service. Bristol: The Centre for Market and Public Organisation Working Paper; 2011.
66. Siciliani L, Martin S. An empirical analysis of the impact of choice on waiting times. *Heal Econ*. 2007 Aug;16(8):763–79.
67. Cooper ZN, McGuire A, Jones S, Le Grand J. Equity, waiting times, and NHS reforms: retrospective study. *BMJ*. 2009 Jan;339:b3264.
68. Pollock A, Macfarlane A, Kirkwood G, Majeed FA, Greener I, Morelli C, et al. No evidence that patient choice in the NHS saves lives. *Lancet*. 2011. p. 2010–2.
69. Bloom N, Cooper Z, Gaynor M, Gibbons S, Jones S, McGuire A, et al. In defence of our research on competition in England's National Health Service. *Lancet*. 2011. 378(9809):2064-5
70. Stark J, Gallivan S, Lovegrove J, Hamilton JR, Monro JL, Pollock JC, et al. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet*. 2000 Mar 18;355(9208):1004–7.
71. Cheraghi-Sohi S, Hole AR, Mead N, McDonald R, Whalley D, Bower P, et al. What Patients Want From Primary Care Consultations: A Discrete Choice Experiment to Identify Patients' Priorities. *Ann Fam Med*. 2008;6(2):107–15.
72. Priorities for the NHS [Internet]. Ipsos Mori. 2013 [cited 2013 Jul 1]. Available from: <http://www.ipsos-mori.com/researchpublications/researcharchive/3165/Priorities-for-the-NHS.aspx>
73. Report on the National Patient Choice Survey - February 2010 England. London: Department of Health; 2010.

74. British Social Attitudes 25th Report [Internet]. NatCen. 2009 [cited 2013 Jun 30]. Available from: <http://www.natcen.ac.uk/study/british-social-attitudes-25th-report/our-findings>
75. Fotaki M, Roland M, Boyd A, McDonald R, Scheaff R, Smith L. What benefits will choice bring to patients? Literature review and assessment of implications. *J Heal Serv Res Policy*. 2008 Jul;13(3):178–84.
76. Vaiana ME, McGlynn EA. What cognitive science tells us about the design of reports for consumers. *Med Care Res Rev*. 2002 Mar;59(1):3–35.
77. Birk HO, Henriksen LO. Which factors decided general practitioners' choice of hospital on behalf of their patients in an area with free choice of public hospital? A questionnaire study. *BMC Heal Serv Res*. 2012 Jan;12(1):126.
78. Lavery AA, Smith PC, Pape UJ, Mears A, Wachter RM, Millett C. High-profile investigations into hospital safety problems in England did not prompt patients to switch providers. *Heal Aff*. 2012 Mar;31(3):593–601.
79. Schlesinger M, Kanouse DE, Rybowski L, Martino SC. Consumer Response to Patient Experience Measures in Complex Information Environments. *Med Care*. 2012;50(11):56–64.
80. Boyce T, Reutskaja E. Choosing a high quality hospital: The role of nudges , scorecard design and information. London: The King's Fund;
81. Cameron D. Letter to Cabinet ministers [Internet]. Cabinet Office. 2011. Available from: <http://www.number10.gov.uk/news/letter-to-cabinet-ministers-on-transparency-and-open-data/>
82. Police.uk [Internet]. Ministry of Justice. 2013. Available from: [www.police.uk](http://www.police.uk).
83. Making local councils more transparent and accountable to local people [Internet]. Department for Communities and Local Government. 2013 [cited 2013 Jul 7]. Available from: <https://www.gov.uk/government/policies/making-local-councils-more-transparent-and-accountable-to-local-people>
84. Performance Tables [Internet]. Department for Education. 2013. Available from: <http://www.education.gov.uk/schools/performance>
85. Cabinet Office. Excellence and fairness: Achieving World Class Public Services. London: The Stationery Office; 2008.
86. Maude promises new era of data transparency [Internet]. BBC News. 2011 [cited 2013 Apr 24]. Available from: <http://www.bbc.co.uk/news/uk-politics-14389709>
87. Hospital guide - directory of UK private & NHS hospitals [Internet]. Dr Foster. 2012 [cited 2013 Jul 7]. Available from: <http://www.drfoosterhealth.co.uk/hospital-guide/>
88. Health and Social Care Act 2012. London: House of Commons; 2012.

89. The power of information: Putting all of us in control of the health and care information we need. London: Department of Health; 2012.
90. NHS Commissioning Board. Everyone counts: Planning for Patients 2013/14. London: Department of Health; 2013.
91. NHS England. Putting patients first. The NHS England Business plan for 2013/14 - 2015/16. London: NHS Commissioning Board; 2013.
92. NHS England Board Meeting Minutes– 12 April. London: NHS England; 2013.
93. Equity and Excellence: Liberating the NHS. London: Department of Health; 2010.
94. The Vascular Society. Surgeon Level Public Reporting [Internet]. 2013 [cited 2013 Jul 1]. Available from: <http://www.vsqip.org.uk/surgeon-level-public-reporting/>
95. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. *Soc Sci Med*. 2000 Oct;51(7):1087–110.
96. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. *Qual Saf Heal Care*. 2006 Oct 1;15(5):307–10.
97. Mayer TA, Cates RJ, Mastorovich MJ, Royalty DL. Emergency department patient satisfaction: customer service training improves patient satisfaction and ratings of physician and nurse skill. *J Healthc Manag*. 43(5):427–40; discussion 441–2.
98. The World Health Report 2000: Health Systems: Improving Performance. Geneva: World Health Organization; 2000.
99. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. *BMJ*. 2002 Oct 26;325(7370):961–4.
100. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q*. 2005 Jan;83(4):691–729.
101. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Heal Care*. 2003 Dec 1;15(6):523–30.
102. Maxwell RJ. Dimensions of quality revisited: from thought to action. *Qual Health Care*. 1992 Sep;1(3):171–7.
103. Darzi A. High quality care for all: NHS Next Stage Review final report. London: Department of Health; 2008.
104. Jeremy Hunt. Making quality of care as important as quality of treatment [Internet]. The King's Fund. 2012 [cited 2013 Jul 9]. Available from: <http://www.kingsfund.org.uk/audio-video/jeremy-hunt-making-quality-care-important-quality-treatment>

105. Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD Health Care Quality Indicators Project. *Int J Qual Heal Care*. 2006 Sep;18(suppl 1):5–13.
106. Manary MP, Boulding W, Staelin R, Glickman SW. The Patient Experience and Health Outcomes. *N Engl J Med*. 2012 Dec 26;368(3):201–3.
107. Meyer GS, Nelson EC, Pryor DB, James B, Swensen SJ, Kaplan GS, et al. More quality measures versus measuring what matters: a call for balance and parsimony. *BMJ Qual Saf*. 2012 Nov 1;21(11):964–8.
108. Berenson RA, Pronovost PJ, Krumholz HM. Measure patient experience with care and patient-reported outcomes as ends in themselves [Internet]. Available from: <http://thehealthcareblog.com/blog/2013/05/26/4-measure-patient-experience-with-care-and-patient-reported-outcomes-as-ends-in-themselves/>
109. Krumholz HM. Patient-Centered Medicine: The Next Phase in Health Care. *Circ Cardiovasc Qual Outcomes*. 2011 Jul 19;4(4):374–5.
110. Care Quality Commission. Briefing Note: 2010 inpatient survey [Internet]. 2011. Available from: <http://www.nhssurveys.org/survey/1016>
111. Airey C, Brewster S, Erens B, Lilley S, Pickering K PL. National Surveys of NHS Patients General Practice. London: NHS Executive; 1998.
112. Ipsos Mori. The GP Patient Survey: January - September 2012 Summary Report [Internet]. 2012. Available from: [http://www.gp-patient.co.uk/results/download/\\_y7q1/Y7W1 National Summary Report.pdf](http://www.gp-patient.co.uk/results/download/_y7q1/Y7W1%20National%20Summary%20Report.pdf)
113. Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev*. 2009/07/30 ed. 2010;67(1):27–37.
114. Robert G, Cornwell J. Rethinking policy approaches to measuring and improving patient experience. *J Heal Serv Res Policy*. 2013 Apr 1;18(2):67–9.
115. Inpatient survey 2012 | Care Quality Commission [Internet]. [cited 2013 Jul 9]. Available from: <http://www.cqc.org.uk/public/reports-surveys-and-reviews/surveys/inpatient-survey-2012>
116. Mountford J, Shojania KG. Refocusing quality measurement to best support quality improvement: local ownership of quality measurement by clinicians. *BMJ Qual Saf*. 2012;21(6):519–23.
117. DeCourcy A, West E, Barron D. The National Adult Inpatient Survey conducted in the English National Health Service from 2002 to 2009: how have the data been used and what do we know as a result? *BMC Heal Serv Res*. 2012 Jan;12(1):71.
118. Paddison C, Elliott M, Parker R, Staetsky L, Lyratzopoulos G, Campbell JL, et al. Should measures of patient experience in primary care be adjusted for case mix?

- Evidence from the English General Practice Patient Survey. *BMJ Qual Saf.* 2012 Aug;21(8):634–40.
119. Lyratzopoulos G, Elliott MN, Barbiere JM, Staetsky L, Paddison CA, Campbell J, et al. How can health care organizations be reliably compared?: Lessons from a national survey of patient experience. *Med Care.* 2011 Aug;49(8):724–33.
  120. Paddison CAM, Abel GA, Roland MO, Elliott MN, Lyratzopoulos G, Campbell JL. Drivers of overall satisfaction with primary care: evidence from the English General Practice Patient Survey. *Heal Expect.* 2013 May 30;
  121. Evans D, Powell J, Cross T. Patient Advice and Liaison Services: results of an audit survey in England. *Heal Expect.* 2008 Sep;11(3):304–16.
  122. South J. Bridging the gap? A critical analysis of the development of the Patient Advice And Liaison Service (PALS). *J Heal Organ Manag.* 2007 Jan;21(2):149–65.
  123. Data on Written Complaints in the NHS - 2011-12 [Internet]. Health and Social Care Information Centre. 2012 [cited 2013 Jun 26]. Available from: <http://www.hscic.gov.uk/catalogue/PUB07197>
  124. Trigg N. Written complaints about care in the NHS rise by 8 per cent. *Nurs Manag.* 2012 Oct;19(6):6–7.
  125. Tingle J. The NHS needs to listen harder and learn more from complaints. *Br J Nurs.* 19(21):1372–3.
  126. Wise J. Handling complaints poorly makes problems worse, says NHS ombudsman. *BMJ.* 2010 Jan;341:c5874.
  127. Parliamentary and Health Services Ombudsman Annual Report. London: The Stationery Office; 2012.
  128. British Prime Minister's Office. Hospital Friends and Family Test Announced [Internet]. London: Cabinet Office; 2012. Available from: <http://www.number10.gov.uk/news/friends-and-family-test/>
  129. NHS Friends and Family Test: Implementation Guidance. London: NHS England; 2013.
  130. Lynn P. The friends and family test is unfit for purpose [Internet]. *Guardian.* 2013 [cited 2013 Jul 2]. Available from: <http://www.guardian.co.uk/healthcare-network/2013/apr/09/friends-family-test-unfit-for-purpose>
  131. Patientexperience.co.uk. Criticism of the Friends and Family Test [Internet]. Patientexperience.co.uk. 2013 [cited 2013 Jul 2]. Available from: <http://www.patientexperience.co.uk/friends-and-family-test/criticism-friends-and-family-test/>
  132. Timmins N. NHS goes to the PROMS. *BMJ.* 2008 Jun 28;336(7659):1464–5.

133. Francis R. Independent Inquiry into care provided by Mid Staffordshire NHS Foundation Trust January 2005 – March 2009. London: The Stationery Office; 2010.
134. NHS Patient Feedback Challenge [Internet]. NHS Institute for Improvement and Innovation. Leeds; 2012. Available from: [http://www.institute.nhs.uk/innovation/spread\\_and\\_adoption/nhs\\_patient\\_feedback\\_challenge.html](http://www.institute.nhs.uk/innovation/spread_and_adoption/nhs_patient_feedback_challenge.html)
135. Richard Simmons, Carol Brennan. Grumbles , Gripes and Grievances The role of complainTs in Transforming public services. London: NESTA; 2013.
136. Cracknell R. Key Issues for the New Parliament 2010: The ageing population. London: House of Commons Library; 2010.
137. Moore A. Older people. We can work it out. *Heal Serv J*. 2007 Jan 11;117(6038):24–6.
138. Appleby J. Databriefing. How satisfied are we with the NHS? *BMJ*. 2011 Jan 21;342(mar21\_2):d1836.
139. Politics and satisfaction with the NHS [Internet]. The King’s Fund, 2013 [cited 2013]. Available from: <http://www.kingsfund.org.uk/blog/2013/04/politics-and-satisfaction-nhs>
140. Statistical Bulletin Internet Access - Households and Individuals , 2012 [Internet]. Office for National Statistics. 2013. p. 1–19. Available from: <http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2012/stb-internet-access--households-and-individuals--2012.html>
141. Joel M. Ctrl Alt Delete: Reboot Your Business. Reboot Your Life. Your Future Depends on It. San Francisco, CA: Business Plus; 2012.
142. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. *Heal Aff*. 2000 Nov 1;19(6):255–65.
143. Edmonds R. Newspaper advertising was down 7.3 percent, almost \$2 billion, in 2011 | Poynter. [Internet]. Poynter. 2012 [cited 2013 Jun 25]. Available from: <http://www.poynter.org/latest-news/business-news/the-biz-blog/166454/newspaper-advertising-was-down-7-3-percent-almost-2-billion-in-2011/>
144. Facebook’s Monthly Active Users Up 23% to 1.11B; Daily Users Up 26% To 665M [Internet]. TechCrunch. 2012 [cited 2013 Jul 2]. Available from: <http://techcrunch.com/2013/05/01/facebook-sees-26-year-over-year-growth-in-daus-23-in-maus-mobile-54/>
145. Twitter Passed 500M Users In June 2012, 140M Of Them In US [Internet]. TechCrunch. 2012 [cited 2013 Jul 2]. Available from: <http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/>

146. Seddon C. Social Trends: Lifestyles and social participation [Internet]. Office for National Statistics. 2011 [cited 2013 Apr 12]. Available from: [http://data.gov.uk/dataset/social\\_trends/resource/ba22c47d-8225-4010-ad97-ecd2a4ff30a1](http://data.gov.uk/dataset/social_trends/resource/ba22c47d-8225-4010-ad97-ecd2a4ff30a1)
147. Randall C. Social Trends 41: e-society [Internet]. Office for National Statistics. 2011. Available from: <http://www.ons.gov.uk/ons/rel/social-trends-rd/social-trends/social-trends-41/index.html>
148. Giustini D. How Web 2.0 is changing medicine. *BMJ*. 2006 Dec 23;333(7582):1283–4.
149. O'Reilly T. What is web 2.0 [Internet]. Oreilly.com. 2005 [cited 2013 Apr 11]. Available from: [http://www.ttivanguard.com/ttivanguard\\_cfmfiles/pdf/dc05/dc05session4003.pdf](http://www.ttivanguard.com/ttivanguard_cfmfiles/pdf/dc05/dc05session4003.pdf)
150. Iyengar R, Han S, Gupta S. Do Friends Influence Purchases in a Social Network? Cambridge: Harvard Business School; 2009.
151. Pires GD, Stanton J, Rita P. The internet, consumer empowerment and marketing strategies. *Eur J Mark*. 2006 Jan 9;40(9/10):936–49.
152. Fox S, Duggan M. Health Online 2013. Washington, DC: Pew Research Center; 2013.
153. Lupton D. The Commodification of Patient Opinion: the Digital Patient Experience Economy in the Age of Big Data [Internet]. 2013 [cited 2013 Jul 2]. Available from: [http://ses.library.usyd.edu.au/bitstream/2123/9063/1/Commodification of Patient Opinion - Working Paper No. 3.pdf](http://ses.library.usyd.edu.au/bitstream/2123/9063/1/Commodification%20of%20Patient%20Opinion%20-%20Working%20Paper%20No.%203.pdf)
154. PatientsLikeMe [Internet]. 2013 [cited 2013 Jul 7]. Available from: <http://www.patientslikeme.com/>
155. Health Unlocked: Condition and wellness communities from HealthUnlocked [Internet]. [cited 2013 Jul 7]. Available from: <http://www.healthunlocked.com/>
156. Richards T, Montori VM, Godlee F, Lapsley P, Paul D. Let the patient revolution begin. *BMJ*. 2013 Jan;346:f2614.
157. Muir Gray. The third health care revolution [Internet]. The King's Fund. 2013 [cited 2013 Jun 24]. Available from: <http://www.kingsfund.org.uk/audio-video/muir-gray-third-health-care-revolution>
158. deBronkart D. How the e-patient community helped save my life: an essay by Dave deBronkart. *BMJ*. 2013 Jan;346:f1990.
159. Dave deBronkart. Meet e-Patient Dave [Internet]. TED.com. 2011 [cited 2013 Jul 8]. Available from: [http://www.ted.com/talks/dave\\_debronkart\\_meet\\_e\\_patient\\_dave.html](http://www.ted.com/talks/dave_debronkart_meet_e_patient_dave.html)
160. George A. Crowdsourcing a cure for my brain cancer [Internet]. New Scientist. 2012 [cited 2013 Jul 7]. Available from:

<http://www.newscientist.com/article/mg21628880.300-crowdsourcing-a-cure-for-my-brain-cancer.html#.UdnGQvmW98F>

161. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. *J Gen Intern Med.* 2010 Sep;25(9):942–6.
162. Isaac T, Zaslavsky AM, Cleary PD, Landon BE. The Relationship between Patients' Perception of Care and Measures of Hospital Quality and Safety. *Heal Servi Res.* 2010;45(4):1024–40.
163. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' Perception of Hospital Care in the United States. *N Engl J Med.* 2008;359(18):1921–31.
164. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *Am J Manag Care.* 2011 Jan;17(1):41–8.
165. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Arch Int Med.* 2012 Mar 12;172(5):405–11.
166. Chang JT, Hays RD, Shekelle PG, Maclean CH, Solomon DH. Patient Care Patients' Global Ratings of Their Health Care Are Not Associated with the Technical Quality of Their Care. *Ann Intern Med.* 2006;144(9):665–72.
167. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open.* 2013 Jan 3;3(1):e001570–.
168. Emmert M, Sander U, Esslinger AS, Maryschok M, Schöffski O. Public reporting in Germany: the content of physician rating websites. *Methods Inf Med.* 2012 Jan;51(2):112–20.
169. Kontopantelis E, Roland M, Reeves D. Patient experience of access to primary care: identification of predictors in a national patient survey. *BMC Fam Pr.* 2010/08/31 ed. 2010;11:61.
170. Elwyn G, Buetow S, Hibbard J, Wensing M. Respecting the subjective: quality measurement from the patient's perspective. *BMJ.* 2007;335(7628):1021–2.
171. Hastings A, Matthews P. “Sharp Elbows”: Do the Middle-Classes have Advantages in Public Service Provision and if so how? Glasgow: University Of Glasgow; 2011.
172. Burt J, Hooper R, Jessopp L. The relationship between use of NHS Direct and deprivation in southeast London: an ecological analysis. *J Public Heal.* 2003;25(2):174–6.

173. Mead N, Varnam R, Rogers A, Roland M. What predicts patients' interest in the Internet as a health resource in primary care in England? *J Heal Serv Res Policy*. 2003;8(1):33–9.
174. Poon DCH, Leung L. Effects of Narcissism, Leisure Boredom, and Gratifications Sought on User-Generated Content Among Net-Generation Users. *Int J Cyber Beh Psy Leatn*. 2011 Jan;1(3):1–14.
175. Tudor Hart J. The Inverse Care Law. *Lancet*. 1971 Feb 27;297(7696):405–12.
176. BBC. Patients to rate GPs on website [Internet]. BBC; 2008. Available from: <http://news.bbc.co.uk/1/hi/health/7804125.stm>
177. Boffey D. GPs' fury over "poison pen" online tirades [Internet]. *Guardian*. London; 2011. Available from: <http://www.guardian.co.uk/society/2011/dec/03/gps-poison-pen-online>
178. Mackay B. RateMDs.com nets ire of Canadian physicians. *CMAJ*. 2007 Mar 13;176(6):754.
179. Mueller M a, Glassett TS, Burke SM, Nichter LS. Protecting a Plastic Surgeon's Reputation: Damage Control for Dishonest Online Reviews. *Plast Reconstr Surg*. 2013 Apr;131(4):670e–1e.
180. Strech D. Ethical principles for physician rating sites. *J Med Internet Res*. 2011 Jan 6;13(4):e113.
181. McKinsty B, Colthart I, Walker J. Can doctors predict patients' satisfaction and enablement? A cross-sectional observational study. *Fam Pr*. 2006 Apr 1;23(2):240–5.
182. Shannon SE, Mitchell PH, Cain KC. Patients, Nurses, and Physicians Have Differing Views of Quality of Critical Care. *J Nurs Sch*. 2002 Jun;34(2):173–9.
183. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance *Med Teach*. 2006;28(2):117–28.
184. Caldwell G, Hospital W, Sussex W. Real-time assessment and feedback of junior doctors improves clinical performance. *Clin Teach*. 2006;3(3):185–8.
185. Reinders ME, Ryan BL, Blankenstein AH, van der Horst HE, Stewart MA, van Marwijk HWJ. The effect of patient feedback on physicians' consultation skills: a systematic review. *Acad Med*. 2011 Nov;86(11):1426–36.
186. Jain S. Googling Ourselves — What Physicians Can Learn from Online Rating Sites. *N Engl J Med*. 2010;362(1):6–7.
187. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. *J Med Internet Res*. 2013 Jan;15(2):e24.

188. Gao G, McCullough J, Agarwal R, Jha A. A Changing Landscape of Physician Quality Reporting: Analysis of Patients' Online Ratings of Their Physicians Over a 5-Year Period. *J Med Internet Res*. 2012;14(1):e38.
189. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online Reviews of 500 Urologists. *J Urol*. 2012 Dec 7;
190. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. *J Gen Intern Med*. 2010/05/14 ed. 2010;25(9):942–6.
191. Black EW, Thompson L, Saliba H, Dawson K, Black NMP. An analysis of healthcare providers' online ratings. *Inf Prim Care*. 2009;17:249–53.
192. Lagu T, Goff SL, Hannon NS, Shatz A, Lindenauer PK. A mixed-methods analysis of patient reviews of hospital care in England: implications for public reporting of health care quality data in the United States. *Jt Comm J Qual Patient Saf*. 2013 Jan;39(1):7–15.
193. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. *J Gen Intern Med*. 2012 Jun;27(6):685–92.
194. Galizzi MMM, Miraldo M, Stavropoulou C, Desai M, Jayatunga W, Joshi M, et al. Who is more likely to use doctor-rating websites, and why? A cross-sectional study in London. *BMJ Open*. 2012 Jan;2(e001493).
195. TripAdvisor reviews increase 50% year-on-year to 100 million [Internet]. *Hotelmktg.com*. 2013 [cited 2013 Jul 8]. Available from: [http://www.hotelmarketing.com/index.php/content/article/tripadvisor\\_reviews\\_increase\\_50\\_year\\_on\\_year\\_to\\_100\\_million/](http://www.hotelmarketing.com/index.php/content/article/tripadvisor_reviews_increase_50_year_on_year_to_100_million/)
196. Rowling JK. *Harry Potter and the Deathly Hallows*. London: Bloomsbury; 2007.
197. Social shopping 2011 [Internet]. [cited 2013 Jul 3]. Available from: [http://www.powerreviews.com/assets/download/Social\\_Shopping\\_2011\\_Brief1.pdf](http://www.powerreviews.com/assets/download/Social_Shopping_2011_Brief1.pdf)
198. Laura McCarthy, Debra Stock RV. How Travelers Use Online and Social Media Channels to Make Hotel-choice Decisions [Internet]. *Cornell Hospitality Report*. 2010 [cited 2013 Jun 23]. Available from: [http://www.hotelschool.cornell.edu/chr/pdf/showpdf/1247/chr/research/socialmediaverma.pdf?t=CHR&id=1247&my\\_path\\_info=chr/research/socialmediaverma.pdf](http://www.hotelschool.cornell.edu/chr/pdf/showpdf/1247/chr/research/socialmediaverma.pdf?t=CHR&id=1247&my_path_info=chr/research/socialmediaverma.pdf)
199. Racherla R, Connolly D, Christodoulidou N. Unscrambling the Puzzling Matter of Online Consumer Ratings: An Exploratory Analysis. *Cornell Hosp Rep*. 2012;11(16).
200. Melián-González S, Bulchand-Gidumal J, González López-Valcárcel B. Online Customer Reviews of Hotels: As Participation Increases, Better Evaluation Is Obtained. *Cornell Hosp. Q*. 2013;29: 23.

201. Stuart E. Levy WD and SB. An Analysis of One-Star Online Reviews and Responses in the Washington, D.C., Lodging Market. *Cornell Hosp Q.* 2013;54: 49.
202. Pinch T, Kesler F. How Aunt Ammy Gets Her Free Lunch: A Study of the Top-Thousand Customer Reviewers at Amazon.com. Ithaca, NY: Cornell Univeristy; 2011. p. 1–90.
203. Anderson C. The Impact of Social Media on Lodging Performance. *Cornell Hosp Rep.* New York; 2012;12(15):4–11.
204. Streitfield D. Ferreting Out Fake Reviews Online - NYTimes.com [Internet]. New York Times. 2011 [cited 2013 Jul 3]. Available from: [http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?\\_r=2&](http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?_r=2&)
205. Streitfield D. In a Race to Out-Rave, 5-Star Web Reviews Go for \$5 [Internet]. New York Times. 2011 [cited 2013 Jul 4]. Available from: [http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?\\_r=2&](http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?_r=2&)
206. Fernandez C. Tripadvisor: Hotel owners bribe guests in return for good reviews [Internet]. Daily Mail. 2011 [cited 2013 Jul 4]. Available from: <http://www.dailymail.co.uk/travel/article-2013391/Tripadvisor-Hotel-owners-bribe-guests-return-good-reviews.html>
207. Main Street Hub [Internet]. [cited 2013 Jul 4]. Available from: <http://www.mainstreethub.com/>
208. Hancock JT, Cardie C. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, OR; 2011. p. 309–19.
209. Donaldson LJ, Darzi A. Quality measures: bridging the cultural divide. *BMJ Qual Saf.* 2012 Nov 1;21(11):969–72.
210. PatientOpinion. Patient Opinion [Internet]. 2013. Available from: <https://www.patientopinion.org.uk/>
211. iWantGreatCare. iWantGreatCare [Internet]. 2013. Available from: <https://www.iwantgreatcare.org/>
212. Hodgkin P. Power to the people [Internet]. E-Health Insider. 2013 [cited 2013 Jul 8]. Available from: <http://www.ehi.co.uk/insight/analysis/993/power-to-the-people>
213. NHS Choices [Internet]. NHS Choices. 2013. Available from: <http://www.nhs.uk/>
214. NHS Choices. NHS Choices performance and statistics [Internet]. NHS Choices. 2013. Available from: <http://www.nhs.uk/aboutNHSChoices/professionals/developments/Pages/Preformancesstatistics.aspx>

215. NHS Choices. NHS Choices user survey [Internet]. Department of Health; 2012 [cited 2013 May 6]. Available from:  
<http://www.nhs.uk/aboutNHSChoices/professionals/healthandcareprofessionals/otherresources/Pages/research-findings.aspx>
216. Gann B, Grant MJ. From NHS Choices to Integrated Customer Service Platform. *Health Info Libr J*. London; 2013 Mar p. 1–3.
217. NHS Choices. NHS Choices comments policy [Internet]. NHS Choices. 2013. Available from:  
<http://www.nhs.uk/aboutNHSChoices/aboutnhschoices/termsandconditions/Pages/commentspolicy.aspx>
218. NHS Choices. NHS Choices privacy policy [Internet]. NHS Choices. 2013. Available from:  
<http://www.nhs.uk/aboutNHSChoices/aboutnhschoices/termsandconditions/Pages/Privacyacypolicy.aspx>
219. NHS Choices. NHS Choices datasets: Hospital comments and responses [Internet]. 2013. Available from:  
<http://www.nhs.uk/aboutNHSChoices/professionals/developments/Pages/NHSChoicesDatasets.aspx>
220. McAfee A, Brynjolfsson E. Big Data: The Management Revolution. *Harv Bus Rev*. 2012;90(10):60–8.
221. Tumasjan A, Sprenger T, Sandner P, Welp I. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*. Washington DC: AAAI Press; 2010.
222. Schlesinger S. Using analytics to predict Hollywood blockbusters [Internet]. *Harvard Business Review Blog*. 2012 [cited 2013 Aug 13]. Available from:  
[http://blogs.hbr.org/cs/2012/10/using\\_analytics\\_to\\_predict\\_hollywood\\_blockbusters.html](http://blogs.hbr.org/cs/2012/10/using_analytics_to_predict_hollywood_blockbusters.html)
223. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *J Comput Sci*. 2011;2(1):1–8.
224. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012–4.
225. Signorini A, Segre AM, Polgreen PM. The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS One*. 2011;6(5):e19467.
226. Chunara R, Andrews JR, Brownstein JS. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. *Am J Trop Med Hyg*. 2012;86(1):39–45.

227. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. *J Med Internet Res*. 2009;11(1):e11.
228. Fox S. *The Social Life of Health Information*, 2011. Washington, DC: Pew Research Center; 2011.
229. Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. *Qual Manag Heal Care*. 2012;21(1):9–19.
230. Cambria E, Hussain A, Durrani T, Havasi C, Eckl C, Munro J. Sentic Computing for patient centered applications. *Signal Processing (ICSP)*, 2010 IEEE 10th International Conference on. 2010. p. 1279–82.
231. Cambria E, Hussain A, Havasi C. Towards the Crowd Validation of the British National Health Service. *Proceedings of WebSci10: Extending the Frontiers of Society On-Line*. Raleigh, NC: US.; 2010.
232. Patient Safety Indicators Overview. AHRQ Quality Indicators. [Internet]. Rockville: Agency for Healthcare Research and Quality; 2006. Available from: [http://www.qualityindicators.ahrq.gov/psi\\_overview.htm](http://www.qualityindicators.ahrq.gov/psi_overview.htm)
233. Hospital Quality Reports [Internet]. London: Dr Foster Health; 2011. Available from: <http://www.drfoosterhealth.co.uk/quality-reports/>
234. NHS Comparators [Internet]. Leeds: NHS Information Centre; 2011. Available from: [www.nhscomparators.nhs.uk](http://www.nhscomparators.nhs.uk)
235. Secondary Uses Service [Internet]. London: NHS Connecting for Health; 2011. Available from: <http://www.connectingforhealth.nhs.uk/systemsandservices/sus>
236. Results from the mandatory surveillance of MRSA bacteraemia [Internet]. Health Protection Agency; 2011. Available from: [http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb\\_C/1233906819629](http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1233906819629)
237. Mandatory surveillance of *Clostridium difficile* [Internet]. London: Health Protection Agency; 2011. Available from: [http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb\\_C/1179746015058](http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1179746015058)
238. Acute Trusts: Adult Inpatients Survey [Internet]. UK Data Service. London: UK Data Service; 2010. Available from: <http://www.esds.ac.uk/findingData/snDescription.asp?sn=6503&key=adult+inpatients+2009>
239. NHS Inpatient Survey briefing note [Internet]. Care Quality Commission. London: Care Quality Commission; 2010. Available from: [http://www.cqc.org.uk/publications.cfm?fde\\_id=15550](http://www.cqc.org.uk/publications.cfm?fde_id=15550)

240. Quality and Outcomes Framework: Online GP practice results database [Internet]. Leeds: NHS Information Centre; 2012. Available from: <http://www.qof.ic.nhs.uk/>
241. GP Patient Survey Summary of Weighting Strategy for year 2011-2012 [Internet]. Available from: <http://www.gp-patient.co.uk/faq/weighting/>
242. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-Performance Programs in Family Practices in the United Kingdom. *N Engl J Med*. 2006;355(4):375–84.
243. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet*. 2008 Aug 30;372(9640):728–36.
244. Saxena S, Car J, Eldred D, Soljak M, Majeed A. Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. *BMC Heal Serv Res*. 2007;7:96.
245. Bardach N, Asteria-Peñaloza R, Boscardin WJ, Dudley RA. The relationship between commercial website ratings and traditional hospital performance measures in the USA. *BMJ Qual Saf*. 2013 Mar 1;22(3):194–202.
246. Salisbury C, Wallace M, Montgomery AA. Patients' experience and satisfaction in primary care: secondary analysis using multilevel modelling. *BMJ*. 2010;341.
247. Hall JA, Dornan MC. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Soc Sci Med*. 1990;30(7):811–8.
248. Penfold RB, Dean S, Flemons W, Moffatt M. Do Hospital Standardized Mortality Ratios Measure Patient Safety? HSMRs in the Winnipeg Regional Health Authority. *Heal Pap*. 2008;8(4):8–24.
249. Jarman B. In Defence of the Hospital Standardized Mortality Ratio. *Heal Pap*. 2008;8(4):37–42.
250. Sung L, Alibhai SM, Ethier M-C, Teuffel O, Cheng S, Fisman D, et al. Discrete choice experiment produced estimates of acceptable risks of therapeutic options in cancer patients with febrile neutropenia. *J Clin Epidemiol*. 2012 Jun;65(6):627–34.
251. Burge P, Devlin N, Appleby J, Gallo F, Nason E, Ling T. *Understanding Patients' Choices at the Point of Referral*. London: Rand; 2006.
252. Robertson R, Burge P. The impact of patient choice of provider on equity: analysis of a patient survey. *J Heal Serv Res Policy*. 2011 Apr 1;16(suppl 1):22–8.
253. Burge P, Devlin N, Appleby J, Rohr C, Grant J. *London Patient Choice Project Evaluation*. London: RAND Corporation; 2005.

254. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy*. 2003; 2(1): 55-64
255. Train K. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge University Press; 2009.
256. Office for National Statistics. Population Estimates by Ethnic Group [Internet]. 2011. Available from:  
<http://www.ons.gov.uk/ons/taxonomy/index.html?nscl=Population+Estimates+by+Ethnic+Group#tab-data-tables>
257. Office for National Statistics. Population Estimates for England and Wales, Mid 2012 [Internet]. 2013. Available from:  
<http://ons.gov.uk/ons/taxonomy/index.html?nscl=Population#tab-data-tables>
258. Office for National Statistics. Official Labour Market Statistics. London; [Internet]. 2012. Available from: <http://www.ons.gov.uk/ons/publications/all-releases.html?definition=tcm:77-21589>
259. Pateman T. Rural and urban areas: comparing lives using rural/urban classifications. London; Office for National Statistics; 2011.
260. Bryan S, Dolan P. Discrete choice experiments in health economics. For better or for worse? *Eur J Health Econ*. 2004 Oct;5(3):199–202.
261. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user’s guide. *Pharmacoeconomics*. 2008 Jan;26(8):661–77.
262. Pang B, Lee L. Opinion Mining and Sentiment Analysis. *Found Trends Inf Retr*. 2008;2(1-2):1–138.
263. Ivanciuc O. Weka machine learning for predicting the phospholipidosis inducing potential. *Curr Top Med Chem*. 2008;8(18):1691–709.
264. Sigurdardottir AK, Jonsdottir H, Benediktsson R. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. *Patient Educ Couns*. 2007;67(1):21–31.
265. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. *Bioinformatics*. 2004;20(15):2479–81.
266. Zhang Y, Jin R, Zhou Z-H. Understanding bag-of-words model: a statistical framework. *Int J Mach Learn Cyber*. 2010 Aug 28;1(1-4):43–52.
267. McCallum A, Nigam K. A comparison of event models for naive bayes text classification. *AAAI-98 Workshop on “Learning for Text Categorization.”* 1998. p. 41–8.
268. Quinlan JR. Induction of Decision Trees. *Mach Learn*. 1986;1(1):81–106.

269. Breiman L. Bagging predictors. *Mach Learn.* 1996;24(2):123–40.
270. Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995;20(3):273–97.
271. NHS Inpatient Survey [Internet]. Picker Europe. 2011. Available from: <http://www.nhssurveys.org/>
272. Twitter turns six [Internet]. Twitter blog. 2012. Available from: <https://blog.twitter.com/2012/twitter-turns-six>
273. Twitter hits half a billion tweets a day [Internet]. CNET News. 2013 [cited 2013 Aug 12]. Available from: [http://news.cnet.com/8301-1023\\_3-57541566-93/report-twitter-hits-half-a-billion-tweets-a-day/](http://news.cnet.com/8301-1023_3-57541566-93/report-twitter-hits-half-a-billion-tweets-a-day/)
274. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One.* 2010 Jan;5(11):e14118.
275. Rui JR, Chen Y, Damiano A. Health Organizations Providing and Seeking Social Support: A Twitter-Based Content Analysis. *Cyberpsychol Behav Soc Netw.* 2013 Jun 20;16(9):669–73.
276. Lyles CR, López A, Pasick R, Sarkar U. “5 mins of uncomfyness is better than dealing with cancer 4 a lifetime”: an exploratory qualitative analysis of cervical and breast cancer screening dialogue on Twitter. *J Cancer Educ.* 2013 Mar;28(1):127–33.
277. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. *BMC Res Notes.* 2012 Jan;5:699.
278. Love B, Himelboim I, Holton A, Stewart K. Twitter as a source of vaccination information: Content drivers and what they are saying. *Am J Infect Control.* 2013 Jun;41(6):568–70.
279. King D, Ramirez-Cano D, Greaves F, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. *Health Policy (New York).* Elsevier; 2013 Mar 12;110(2-3):291–7.
280. Brynolf A, Johansson S, Appelgren E, Lynoe N, Edstedt Bonamy A-K. Virtual colleagues, virtually colleagues--physicians' use of Twitter: a population-based observational study. *BMJ Open.* 2013 Jan;3(7).
281. Desai T, Shariff A, Shariff A, Kats M, Fang X, Christiano C, et al. Tweeting the meeting: an in-depth analysis of Twitter activity at Kidney Week 2011. *PLoS One.* 2012 Jan;7(7):e40253.
282. Acute Trusts: Adult Inpatients Survey, 2012 [Internet]. UK Data Service. 2013. Available from: <http://discover.ukdataservice.ac.uk/catalogue/?sn=7273&type=Data catalogue>

283. Bed Availability and Occupancy. [Internet]. NHS England. 2013.  
<http://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/>
284. Hospital Episode Statistics, Admitted Patient Care, England - 2012-13. Leeds: Health and Social Care Information Centre. 2013.
285. Health and Social Care Information Centre. Indicator Portal [Internet]. 2013. Available from: <https://indicators.ic.nhs.uk/webview/>
286. Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs*. 2008 Apr;62(1):107–15.
287. Glaser BG. The constant comparative method of qualitative analysis. *Soc Probl*. 1965;12(4):436–45.
288. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Heal Res*. 2005 Nov 1;15(9):1277–88.
289. Mislove A, Lehmann S, Ahn YY, Onnela JP, Rosenquist JN. Understanding the Demographics of Twitter Users. Proceedings of the fifth international conference on weblogs and social media. Barcelona, Spain: The AAAI Press, Menlo Park, California; 2011.
290. Surowiecki J. *The Wisdom of Crowds*. New York: Doubleday; 2004.
291. Rozenblum R, Bates DDW. Patient-centred healthcare, social media and the internet: the perfect storm? *BMJ Qual Saf*. 2013;22(3):183–6.
292. Keogh B. Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report. London: Department of Health; 2013.
293. GP Ratings App [Internet]. FineFettle Apps. 2012. Available from: <https://itunes.apple.com/gb/app/gp-ratings/id529897188?mt=8>
294. House of Commons Committee of Public Accounts. *The Care Quality Commission: Regulating the quality and safety of health and adult social care*. London: The Stationery Office; 2012.
295. RateMDs [Internet]. 2013. Available from: [ratemds.com](http://ratemds.com)
296. Mumsnet [Internet]. 2013. Available from: [mumsnet.com](http://mumsnet.com)
297. Epatients.net [Internet]. 2013. Available from: [Epatients.net](http://Epatients.net)
298. Twitter [Internet]. 2013. Available from: [twitter.com](http://twitter.com)
299. Facebook [Internet]. 2013. Available from: [facebook.com](http://facebook.com)
300. Google+ [Internet]. 2013. Available from: [plus.google.com/](http://plus.google.com/)

301. Boyce T, Dixon A, Fasolo B, Reutskaja E. Choosing a high quality hospital. London: King's Fund; 2010.
302. Pew Research Center. Demographics of Internet Users [Internet]. Washington, DC; 2012. Available from: [http://pewinternet.org/Static-Pages/Trend-Data-\(Adults\)/Whos-Online.aspx](http://pewinternet.org/Static-Pages/Trend-Data-(Adults)/Whos-Online.aspx)
303. Xie B, Wang M, Feldman R, Zhou L. Internet Use Frequency and Patient-Centered Care: Measuring Patient Preferences for Participation Using the Health Information Wants Questionnaire. *J Med Internet Res*. 2013;15(7):e132.
304. Lyles C, Sarkar U. Additional considerations for Harnessing the cloud of patient experience. *BMJ Qual Saf*. 2013;22(698).
305. Luxford K. What does the patient know about quality? *Int J Qual Heal Care*. 2012;24(5):439–40.
306. Margetts H, Dunleavy P. The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Phil Trans R Soc A*. 2013;371(20120382,).
307. McKee R. Ethical issues in using social media for health and health care research. *Health Policy (New York)*. 2013;110(2):298–301.
308. Selden N, Ghogawala Z, Harbaugh R. The future of practice science: challenges and opportunities for neurosurgery. *Neurosurg Focus*. 2013;34(1):E8.
309. Mason R. TripAdvisor-style reviews by patients “predict worst hospitals” [Internet]. *Telegraph*. 2012 [cited 2013 Jul 24]. Available from: <http://www.telegraph.co.uk/health/healthnews/9082151/TripAdvisor-style-reviews-by-patients-predict-worst-hospitals.html>
310. Parsons C. Patients' ratings on NHS website “accurately predict which hospitals are cleanest and have lowest death rates” [Internet]. *Daily Mail*. 2012 [cited 2013 Jul 24]. Available from: <http://www.dailymail.co.uk/news/article-2101600/Patients-ratings-NHS-website-accurately-predict-hospitals-cleanest-lowest-death-rates.html>
311. Zorlu G. Patient ratings are a good indicator of hospital performance [Internet]. *Guardian*. 2012. Available from: <http://www.guardian.co.uk/healthcare-network/2012/feb/14/patient-ratings-hospital-pe>
312. Ahuja A. Got the flu? Google will find out long before your GP [Internet]. *Telegraph*. 2013 [cited 2013 Jul 24]. Available from: <http://www.telegraph.co.uk/technology/social-media/9824456/Got-the-flu-Google-will-find-out-long-before-your-GP.html>

## Appendix 1: Outputs from PhD

### Publications relating to PhD:

- **Greaves F**, Pape UJ, King, D., Darzi, A., Majeed, A., Wachter, R. M., & Millett, C. (2012a). Associations Between Web-Based Patient Ratings and Objective Measures of Hospital Quality. *Arch Int Med*, 172(5), 435–436.
- **Greaves F**, Pape UJ, King D, Darz, A, Majeed A, Wachter RM, & Millett C. (2012b). Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Qual Saf*, 21(7), 600–5.
- **Greaves F**, Pape UJ, Lee H, Smith DM, Darzi A, Majeed A, & Millett C. (2012). Patients' Ratings of Family Physician Practices on the Internet: Usage and Associations With Conventional Measures of Quality in the English National Health Service. *J Med Internet Res.*, 14(5), e146.
- **Greaves F**, Ramirez-Cano D, Millett C, Darzi A, & Donaldson L. (2013). Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. *BMJ Qual Saf*, 22(3), 251–5.
- **Greaves F**, Millett C. Consistently increasing numbers of online ratings of healthcare in England. *J Med Internet Res.* 2012 Jun 29;14(3):e94. [Letter]
- **Greaves F**, Ramirez-Cano D, Millett C, Darzi A, & Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. In press. *J Med Internet Res.*

### Submitted papers awaiting decisions:

- England's experience incorporating 'anecdotal' reports from consumers' into their national reporting system: Lessons for what to do or not do? Undergoing revision. *Medical Care Research and Review*.

### Planned papers:

- Impact of online comments about care on hospital choice: A discrete choice study. Target journal: *Health Expectations*.
- A descriptive analysis of tweets about healthcare providers in the English NHS. Target journal: *BMJ Quality and Safety*

### Presentations given relevant to PhD:

#### Invited Policy presentations

- **Greaves F**. Online health reviews. A survey and discrete choice experiment. NHS Choices Information Board. Department of Health. London. August 2013.

- **Greaves F.** Tripadvisor for Health: New ways to measure patient experience. The Scottish patient experience programme board. Scottish Government. Glasgow. June 2012.
- **Greaves F.** Associations Between Web-Based Patient Ratings and Objective Measures of Hospital Quality. NHS Choices Information Board. Department of Health. London. May 2012.

#### Oral Conference Presentations

- **Greaves F.** Ramirez Cano D, Millett C, Darzi A, Donaldson L. Machine learning and sentiment analysis of unstructured free-text information about patient experience online. Public Health Science. London, UK. November 2012.
- **Greaves F.** Patients' ratings of family physicians on the internet: usage and associations with conventional measures of quality in the English NHS. MedicineX. Palo Alto, USA. September 2012.
- **Greaves F.** Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. Medicine 2.0. Boston, USA. September 2012.
- **Greaves F.** Tripadvisor for Health. International Forum on Quality and Safety in Healthcare. Paris, France. April 2011.

#### Grants Received:

- £10,000 for discrete choice experiments from the Department of Health.
- \$4,500 from AHRQ, to provide advice and policy options on the development of narrative reporting platforms in the US.
- Harkness Fellowship to continue further work in this area at Harvard School of Public Health (\$119,000).

#### Grants applied for:

- Co-Investigator on a £1.9M, 4 year NIHR proposal for harnessing online patient experience to improve quality of care. Passed first stage assessment. Currently preparing second round application.

#### Courses taught during PhD:

- Improving Health Services (MPH Module)
  - Organised a 5 day compulsory course, including 10 lectures, exam setting and marking. Taught 3 lectures on the course.
- [Lecture] Measuring Quality in Healthcare. MSc Health Policy. Imperial College London. In Measuring Health Systems Course.
- [Lecture] Seminal papers in public health. MSc Health Policy. Imperial College London. In Health and Society Course.

## Other publications during PhD not related directly to thesis:

- **Greaves F**, Pappas Y, Bardsley M, Harris M, Curry N, Holder H, Blunt I, Soljak M, Gunn L, Majeed A, Car J. Evaluation of complex integrated care programmes: the approach in North West London. *Int J Integr Care*. 2013 Mar 8;13:e006.
- Harris M, **Greaves F**, Gunn L, Patterson S, Greenfield G, Car J, Majeed A, Pappas Y. Multidisciplinary group performance-measuring integration intensity in the context of the North West London Integrated Care Pilot. *Int J Integr Care*. 2013 Feb 15;13:e001.
- Harris M, **Greaves F**, Patterson S, Jones J, Pappas Y, Majeed A, Car J. The North West London Integrated Care Pilot: innovative strategies to improve care coordination for older adults and people with diabetes. *J Ambul Care Manage*. 2012 Jul-Sep;35(3):216-25.
- **Greaves F**, Millett C, Pape UJ, Soljak M, Majeed A. Association between primary care organisation population size and quality of commissioning in England: an observational study. *Br J Gen Pract*. 2012 Jan;62(594):e46-54.
- **Greaves F**, Harris M, Goodwin N, Dixon A. The commissioning reforms in the English National Health Service and their potential impact on primary care. *J Ambul Care Manage*. 2012 Jul-Sep;35(3):192-9.
- **Greaves F**, Donaldson L. Measles in the UK: a test of public health competency in a crisis. *BMJ*. 2013 May 1;346:f2793.
- Hamilton FL, **Greaves F**, Majeed A, Millett C. Effectiveness of providing financial incentives to healthcare professionals for smoking cessation activities: systematic review. *Tob Control*. 2013 Jan;22(1):3-8.
- King D, **Greaves F**, Vlaev I, Darzi A. Approaches based on behavioral economics could help nudge patients and providers toward lower health spending growth. *Health Aff (Millwood)*. 2013 Apr;32(4):661-8.
- King D, Ramirez-Cano D, **Greaves F**, Vlaev I, Beales S, Darzi A. Twitter and the health reforms in the English National Health Service. *Health Policy*. 2013 May;110(2-3):291-7.
- King D, **Greaves F**, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. *J R Soc Med*. 2013 Mar;106(3):76-8.
- Pinder RJ, Greaves FE, Aylin PP, Jarman B, Bottle A. Staff perceptions of quality of care: an observational study of the NHS Staff Survey in hospitals in England. *BMJ Qual Saf*. 2013 Jul;22(7):563-70
- Coltart CE, Cheung R, Ardolino A, Bray B, Rocos B, Bailey A, Bethune R, Butler J, Docherty M, Drysdale K, Fayaz A, **Greaves F**, et al. Leadership development for early career doctors. *Lancet*. 2012 May 12;379(9828):1847-9.
- Finlayson AE, **Greaves F**, Ali FR. Technologies for global health. *Lancet*. 2012 Nov 17;380(9855):1739; author reply 1739. (Letter)

## Appendix 2: Details of regression analysis in Table 14

### 1) Initial Regression (results in main thesis)

Logistic regression of odds of being rated including all practice level variables

|                             |               |   |       |
|-----------------------------|---------------|---|-------|
| Logistic regression         | Number of obs | = | 7723  |
|                             | LR chi2(8)    | = | 659   |
|                             | Prob > chi2   | = | 0.000 |
| Log likelihood = -4792.0251 | Pseudo R2     | = | 0.064 |

| Variable           | Odds Ratio | Std. Err. | Z      | P>z   |
|--------------------|------------|-----------|--------|-------|
| Proportion white   | 0.763      | 0.130     | -0.017 | 0.113 |
| IMD score          | 0.988      | 0.002     | -7.780 | 0.000 |
| Proportion over 65 | 0.113      | 0.063     | -3.940 | 0.000 |
| List size          | 1.000      | 0.000     | 15.410 | 0.000 |
| Singlehanded       | 0.702      | 0.054     | -4.610 | 0.000 |
| PMS                | 0.964      | 0.049     | -0.730 | 0.466 |
| Training practice  | 1.022      | 0.064     | 0.350  | 0.729 |
| Population denisty | 1.006      | 0.001     | 6.720  | 0.000 |
| _cons              | 1.431      | 0.272     | 1.890  | 0.059 |

### 2) Second model

Logistic regression of odds of being rated including all practice level variables, except training practice status

|                             |               |   |        |
|-----------------------------|---------------|---|--------|
| Logistic regression         | Number of obs | = | 7982   |
|                             | LR chi2(7)    | = | 688    |
|                             | Prob > chi2   | = | 0      |
| Log likelihood = -4967.3966 | Pseudo R2     | = | 0.0648 |

| Variable           | Odds Ratio | Std. Err. | Z      | P>z   |
|--------------------|------------|-----------|--------|-------|
| Proportion white   | 0.768      | 0.127     | -1.590 | 0.111 |
| IMD score          | 0.988      | 0.001     | -7.860 | 0.000 |
| Proportion over 65 | 0.144      | 0.073     | -3.820 | 0.000 |
| List size          | 1.000      | 0.000     | 17.290 | 0.000 |
| Singlehanded       | 0.725      | 0.054     | -4.360 | 0.000 |
| PMS                | 0.983      | 0.049     | -0.340 | 0.734 |
| Population Density | 1.006      | 0.001     | 6.880  | 0.000 |
| _cons              | 1.327      | 0.244     | 1.540  | 0.124 |

### 3) Third model

Logistic regression of odds of being rated including all practice level variables, except training practice status and practice contract status

|                             |               |   |   |        |
|-----------------------------|---------------|---|---|--------|
| Logistic regression         | Number of obs | = | = | 7982   |
|                             | LR chi2(6)    | = | = | 688    |
|                             | Prob > chi2   | = | = | 0      |
| Log likelihood = -4967.4544 | Pseudo R2     | = | = | 0.0648 |

| Variable           | Odds Ratio | Std. Err. | Z      | P>z   |
|--------------------|------------|-----------|--------|-------|
|                    |            |           |        |       |
| Proportion white   | 0.768      | 0.127     | -1.590 | 0.111 |
| IMD score          | 0.988      | 0.001     | -7.890 | 0.000 |
| Proportion over 65 | 0.145      | 0.074     | -3.810 | 0.000 |
| List size          | 1.000      | 0.000     | 17.330 | 0.000 |
| Singlehanded       | 0.725      | 0.054     | -4.350 | 0.000 |
| Population Density | 1.006      | 0.001     | 6.880  | 0.000 |
| _cons              | 1.317      | 0.241     | 1.510  | 0.131 |

### 4) Fourth model

Logistic regression of odds of being rated including all practice level variables, except training practice status, practice contract status and population ethnicity

|                             |               |   |   |       |
|-----------------------------|---------------|---|---|-------|
| Logistic regression         | Number of obs | = | = | 7982  |
|                             | LR chi2(5)    | = | = | 686   |
|                             | Prob > chi2   | = | = | 0.000 |
| Log likelihood = -4968.7281 | Pseudo R2     | = | = | 0.065 |

| Variable           | Odds Ratio | Std. Err. | Z     | P>z   |
|--------------------|------------|-----------|-------|-------|
|                    |            |           |       |       |
| IMD score          | 0.989      | 0.001     | -7.73 | 0.000 |
| Proportion over 65 | 0.113      | 0.055     | -4.5  | 0.000 |
| List size          | 1.000      | 0.000     | 17.28 | 0.000 |
| Singlehanded       | 0.726      | 0.054     | -4.34 | 0.000 |
| Population Density | 1.006      | 0.001     | 7.98  | 0.000 |
| _cons 1.054657     | 1.055      | 0.124     | 0.45  | 0.651 |

## Appendix 3: Details of regression analysis in Table 15

### 1) Initial Regression (results in main thesis)

Regression of proportion recommending including all practice level variables

Number of obs = 4312

F( 8, 4303) = 45.80

Prob > F = 0.0000

R-squared = 0.0785

Adj R-squared = 0.0768

Root MSE = .37196

| Variable           | Coefficient | Std. Err. | Z     | P>z   |
|--------------------|-------------|-----------|-------|-------|
| Proportion white   | 0.305       | 0.041     | 7.39  | 0     |
| IMD score          | -0.001      | 0.000     | -3.93 | 0     |
| Proportion over 65 | 0.230       | 0.127     | 1.81  | 0.071 |
| List size          | 0.000       | 0.000     | -4.61 | 0     |
| Singlehanded       | -0.080      | 0.021     | -3.81 | 0     |
| PMS                | 0.007       | 0.012     | 0.64  | 0.523 |
| Population denisty | -0.001      | 0.000     | -4.16 | 0     |
| Training practice  | 0.088       | 0.013     | 6.61  | 0     |
| _cons              | 0.443       | 0.045     | 9.84  | 0     |

### 2) Model 2

Regression of proportion recommending including all practice level variables except contract status

Number of obs = 4312

F( 7, 4304) = 52.29

Prob > F = 0.0000

R-squared = 0.0784

Adj R-squared = 0.0769

Root MSE = .37194

| Variable           | Coefficient | Std. Err. | Z     | P>z   |
|--------------------|-------------|-----------|-------|-------|
| Proportion white   | 0.305       | 0.041     | 7.39  | 0     |
| IMD score          | -0.001      | 0.000     | -3.9  | 0     |
| Proportion over 65 | 0.223       | 0.127     | 1.77  | 0.078 |
| List size          | 0.000       | 0.000     | -4.56 | 0     |
| Singlehanded       | -0.080      | 0.021     | -3.84 | 0     |
| Population density | -0.001      | 0.000     | -4.18 | 0     |
| Training practice  | 0.088       | 0.013     | 6.6   | 0     |
| _cons              | 0.446       | 0.045     | 9.97  | 0     |

### 3) Model 3

Regression of proportion recommending including all practice level variables except

contract status and population age

Number of obs = 4312

F( 6, 4305) = 60.46

Prob > F = 0.0000

R-squared = 0.0777

Adj R-squared = 0.0764

Root MSE = .37203

| Variable           | Coefficient | Std. Err. | Z     | P>z |
|--------------------|-------------|-----------|-------|-----|
|                    |             |           |       |     |
| Proportion white   | 0.328       | 0.039     | 8.37  | 0   |
| IMD score          | -0.002      | 0.000     | -4.01 | 0   |
| List size          | 0.000       | 0.000     | -4.67 | 0   |
| Singlehanded       | -0.083      | 0.021     | -3.97 | 0   |
| Population density | -0.001      | 0.000     | -4.77 | 0   |
| Training practice  | 0.090       | 0.013     | 6.75  | 0   |
| _cons              | 0.465       | 0.043     | 10.72 | 0   |

## Appendix 4: Example survey for discrete choice experiment

### Initial invitation to take part:

*“Dear [Name],*

*You have been invited to take part in a survey for a medical research study. The study looks at how people choose which hospital to go to when they need medical treatment. You will be asked to provide some general information about yourself, and to make some hypothetical choices about which hospitals you might choose to go to. The survey will take about 20 minutes to complete.”*

Are you happy to take part in this study? **Yes / No**”

### Seeking informed consent from those choosing to take part:

*“Thank you for your interest in taking part in this survey. This survey is part of a medical study. The study looks at how people choose their hospitals. You will be asked to provide some general information about yourself, and to make some hypothetical choices about which hospitals you might choose to go to. The survey will take about 20 minutes to complete.*

*The research is being organised by staff at the Department of Primary Care and Public Health at Imperial College in London. You are under no obligation to take part in the survey. If you do choose to take part, your answers will be treated confidentially and stored on secure computer servers. We will not keep any identifiable information such as your name or date of birth.*

Are you happy to give your consent to take part in this study? **Yes / No**”

**Demographic questions:**

|                                             |        |
|---------------------------------------------|--------|
| What gender would you describe yourself as? | Male   |
|                                             | Female |

What was your year of birth? (select from dropdown menu)

**Health status**

How would you rate you own health?

To help people say how good or bad their health state is, we have drawn a scale on which the best state can imagine is marked 100, and the worst state you can imagine is marked 0.

We would like you to indicate on the scale how good or bad your own health is today, in your opinion.

Scale 0-100

|                                                      |                    |
|------------------------------------------------------|--------------------|
| How many times have you visited your GP in last year | None               |
|                                                      | 1                  |
|                                                      | 2-5 times          |
|                                                      | 6-10 times         |
|                                                      | More than 10 times |

|                                                     |                        |
|-----------------------------------------------------|------------------------|
| What is your past experience of your local hospital | Generally good         |
|                                                     | Mixed                  |
|                                                     | Generally poor         |
|                                                     | No previous experience |

|                                                                          |            |
|--------------------------------------------------------------------------|------------|
| Have you been referred to hospital in the last two years for any reason? | Yes        |
|                                                                          | No         |
|                                                                          | Don't Know |

|                                                                      |                      |
|----------------------------------------------------------------------|----------------------|
| How important do you think it is to be offered a choice of hospital? | Very important       |
|                                                                      | Important            |
|                                                                      | Somewhat important   |
|                                                                      | Of little importance |
|                                                                      | Unimportant          |

**[Choice Experiments: each participant will be asked to complete 6 similar choices]**

- Imagine you are in the situation where you have been referred by your GP to hospital for the diagnosis and treatment of a nonlife-threatening condition. For example treatment of joint pain that is interrupting your daily life, which may require an operation to resolve in the future.
- Each question in this section asks you to choose between three hospitals using **only** the information provided. You may also choose not to receive treatment at any hospital.
- The information about each hospital may change from question to question. Sometimes the information may be missing altogether.
- Remember, use the information for each hospital to make your choice, and tick **one** box only in each question.
- There are no right or wrong answers to these choices; we are only interested in your views.

| Hospital Attribute                                           |  | Hospital A                                  | Hospital B                      | Hospital C                             | I would chose not to be referred to any of these hospitals |
|--------------------------------------------------------------|--|---------------------------------------------|---------------------------------|----------------------------------------|------------------------------------------------------------|
| Travel time to hospital                                      |  | 30 mins                                     | 60 mins                         | 90 mins                                |                                                            |
| Waiting time to first appointment                            |  | Within 1 week                               | 4 weeks                         | 8 weeks                                |                                                            |
| MRSA infection rates                                         |  | High infection rate                         | Average infection rate          | Low infection rate                     |                                                            |
| Standardized mortality rate                                  |  | High mortality rate                         | Average mortality rate          | Low mortality rate                     |                                                            |
| Recommended by GP                                            |  | GP actively does not recommend              | GP has no view                  | GP recommends                          |                                                            |
| Recommended by friends or family                             |  | Friends or family actively do not recommend | Friends or family have no view  | Friends of family recommend recommends |                                                            |
| Recommended by patients on a hospital review website         |  | More recommendations than average           | Average level of recommendation | Below average level of recommendation  |                                                            |
| Please select one of the following options (please tick one) |  |                                             |                                 |                                        |                                                            |

**Further questions post choice experiments:**

|                                                                     |                    |
|---------------------------------------------------------------------|--------------------|
| Are you aware that you have a choice of hospital for NHS treatment? | Yes, I am aware    |
|                                                                     | No, I am not aware |

|                                                                            |                                       |
|----------------------------------------------------------------------------|---------------------------------------|
| What sources of information would you use if you were choosing a hospital? | GP's advice                           |
|                                                                            | NHS choices website                   |
|                                                                            | Hospital website                      |
|                                                                            | Friend/family member'                 |
|                                                                            | Own experience                        |
|                                                                            | Patients' ratings on a rating website |
|                                                                            | Local patient organisation            |
|                                                                            | Other internet websites               |
|                                                                            | Other (please describe)               |

|                                           |                  |
|-------------------------------------------|------------------|
| How would you usually travel to hospital? | Walk             |
|                                           | Bike             |
|                                           | Car              |
|                                           | Public transport |
|                                           | Taxi             |
|                                           | Other            |

|                                                                                                                                  |                                           |
|----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| If you were picking a local restaurant or hotel, which of the following approaches would you use? (you may select more than one) | Personal experience                       |
|                                                                                                                                  | Friends recommendation                    |
|                                                                                                                                  | Guide book                                |
|                                                                                                                                  | Review/rating website (e.g. Trip Advisor) |
|                                                                                                                                  | None of these                             |

|                                                                                                                                                                                                                                      |                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|
| Have you looked for information online about health or medical issues, either for yourself or someone else. Specifically, in the last 12 months, have you looked online for information about a specific disease or medical problem? | Yes, I have    |
|                                                                                                                                                                                                                                      | No, I have not |
|                                                                                                                                                                                                                                      | Don't know     |

|                                                                                                             |                 |
|-------------------------------------------------------------------------------------------------------------|-----------------|
| Are you aware of the existence of websites where you can leave reviews or ratings about doctor or hospital? | Yes, I am aware |
|                                                                                                             | No, I not aware |
|                                                                                                             | Don't know      |

|                                                                                                              |                |
|--------------------------------------------------------------------------------------------------------------|----------------|
| Have you looked at a hospital or doctor rating/review website when making a decision about your health care? | Yes, I have    |
|                                                                                                              | No, I have not |
|                                                                                                              | Don't know     |

|                                                                                       |                |
|---------------------------------------------------------------------------------------|----------------|
| Have you ever left a review or comment on a hospital or doctor rating/review website? | Yes, I have    |
|                                                                                       | No, I have not |
|                                                                                       | Don't know     |

|                                                 |                            |
|-------------------------------------------------|----------------------------|
| How would you describe the area where you live? | City/large town            |
|                                                 | Suburbs of city/large town |
|                                                 | Small town                 |
|                                                 | Village/ rural area        |
|                                                 | None of these              |

|                                                      |                                                                 |
|------------------------------------------------------|-----------------------------------------------------------------|
| Which best describes your current employment status? | In paid work                                                    |
|                                                      | Unemployed                                                      |
|                                                      | Retired from paid work                                          |
|                                                      | Unable to work because of disability/ill health                 |
|                                                      | Looking after family, home or dependants                        |
|                                                      | In full-time education including government training programmes |
|                                                      | Other                                                           |

|                                |                   |
|--------------------------------|-------------------|
| What is your household income? | Less than £9,999  |
|                                | £10,000 – £19,999 |
|                                | £20,000 – £29,999 |
|                                | £30,000 – £39,999 |
|                                | £40,000 – £49,999 |
|                                | £50,000 – £74,999 |
|                                | £75,000 or more   |

|                                                       |                                                    |
|-------------------------------------------------------|----------------------------------------------------|
| Which best describes your highest level of education? | No formal qualifications                           |
|                                                       | GCSE/O level/A level or equivalent                 |
|                                                       | Professional qualification below degree level      |
|                                                       | Degree level qualification or higher or equivalent |
|                                                       | Other                                              |

|                                        |                                     |
|----------------------------------------|-------------------------------------|
| How would you describe your ethnicity? | White: British                      |
|                                        | White: Irish                        |
|                                        | Any other White background          |
|                                        | Mixed: White and Black Caribbean    |
|                                        | Mixed: White and Black African      |
|                                        | Mixed: White and Asian              |
|                                        | Any other Mixed background          |
|                                        | Asian or Asian British: Indian      |
|                                        | Asian or Asian British: Pakistani   |
|                                        | Asian or Asian British: Bangladeshi |
|                                        | Any other Asian background          |
|                                        | Black or Black British: Caribbean   |
|                                        | Black or Black British: African     |
|                                        | Any other Black background          |
|                                        | Chinese                             |
|                                        | Other ethnic group                  |
|                                        | Prefer not to say                   |

|                                                           |                                                                                                      |
|-----------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Please select the category that best describes your work. | Modern Professional e.g. teacher, nurse, social worker, artist, police officer (sergeant or above).  |
|                                                           | Clerical & intermediate e.g. secretary, call centre agent, nursing auxiliary, nursery nurse.         |
|                                                           | Senior managers or administrators e.g. finance manager, chief executive.                             |
|                                                           | Technical & craft e.g. motor mechanic, plumber, printer, tool maker, gardener, train driver, fitter. |
|                                                           | Semi-routine manual & service e.g. postal / farm worker, security guard, catering / sales assistant. |
|                                                           | Routine manual & service e.g. HGV driver, porter, cleaner, sewing machinist, bar staff, labourer.    |
|                                                           | Middle or junior managers e.g. office / retail / bank / restaurant / warehouse manager, publican.    |
|                                                           | Traditional professional e.g. accountant, solicitor, medical practitioner, scientist, civil servant. |
|                                                           | Never worked                                                                                         |