How voters evaluate public services: the case of the National Health Service in England

BURDON, MARK,GIDEON

How voters evaluate public services: the case of the National Health Service in England

Mark Gideon Burdon

A key condition for the public to be able to hold democratic governments accountable is that voters must be able to evaluate government performance. This thesis improves our understanding of the operation of this first step in the retrospective accountability process. It asks whether voter perceptions are responsive to real differences in performance, and which factors condition this responsiveness. Drawing on social-psychological and valence literatures, I examine how partisanship inhibits responsiveness of perceptions and whether the types of information that voters receive about performance is sufficient to change perceptions.

My empirical evidence comes from analysis of voters' evaluations of the National Health Service in England between 1983 and 2019 as recorded in cross-sectional and panel survey data, combined with official performance statistics. This is a good case study because the issue of healthcare is highly salient, and clear metrics exist to measure the service's performance over time and between different areas of the country.

I find that although voter evaluations of NHS performance do depend on partisanship, this relationship is not so strong as to crowd out responsiveness to changes in real performance. Voters evaluate the NHS more positively when national and local performance is better, and that recent users of services are more satisfied with health services. However, the publication of plans to downgrade hospital services in some parts of England did not result in changes in opinions among people living nearby. This thesis, therefore, argues that voter perceptions are indeed responsive to public service performance, but only under certain conditions - primarily when there is a consistent flow of credible information for voters to process.

How voters evaluate public services: the case of the National Health Service in England

Mark Gideon Burdon

December 2023

School of Government and International Affairs

Durham University

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy in Politics

Table of contents

List of figures

List of tables

Declaration

Material contained in the thesis has not previously been submitted for a degree in this or any other institution. My Master's thesis (submitted to Durham University in 2018) included a similar statistical analysis to the models included in chapter 8's retrenchment study; however, the material in this thesis is theoretically, substantively and methodologically distinct and improved from that previous work.

Statement of Copyright

Acknowledgements

I began laying the foundations for this thesis six years ago as a Master's student; in that time I became a father to a third child, moved house twice, and returned to employment. One constant throughout this whole time has been the consistent support and guidance I received from my supervisors, Professors Nick Vivyan and Gidon Cohen. They have gone above and beyond to help me situate my work in the literature, think carefully about my theory, avoid getting lost down rabbit holes and produce better research. I am incredibly grateful for their support.

Other members of staff and students at the School of Government and International Affairs have offered feedback on presentations, draft chapter and papers, for which I am grateful. My thanks also to Patrick Kuhn for his feedback on my Master's work, and suggestion of the use of entropy balancing.

The Economic and Social Research Council (now part of UK Research and Innovation) funded me as a 1+3 student, making it possible for me to undertake this study. The Master's in Research Methods expanded my academic horizons and gave me the tools I needed to embark on a quantitative Ph.D. thesis, and I would recommend this pathway to others interested in undertaking a doctorate. I am grateful for the support of my former employers at NHS South Tees CCG, in particular Amanda Hume, who supported my initial application. I am also grateful to the virtual community of R users and other open source software developers who have made statistical analysis such as I have performed in this thesis much simpler in recent years.

I would also like to express my gratitude to my wife Kathrine for believing in me and pushing me to complete my thesis, even when it was most difficult; and to my parents for helping look after our children, especially during my Master's study and the late stages of writing up the

thesis. I am grateful to my friends and extended family members who must surely have grown sick of asking how close I was to the end of my Ph.D. and receiving a vague answer.

Dedication

To my wife Kathrine, without whose unfailing support and confidence I could not possibly have completed this thesis.

1 Introduction

Probably the most famous maxim about voting behaviour is James Carville's proclamation that "it's the economy, stupid" (Bennett 2013). The theory of economic voting - the idea that voters react to how the government is managing the economy - has been well-studied in political science. One of Carville's other statements on the 1993 election has been less discussed: "Don't forget health care" (Bennett 2013).

While economic management is clearly an important part of both winning elections and governing countries, modern states have taken responsibility for providing a range of services that have significant impacts on peoples' lives and wellbeing, including old-age pensions, social care, education, and healthcare (Castles, Leibfried, Lewis, *et al.* 2012); this wide-ranging responsibility for important issues is reflected in high voter salience for issues like healthcare (Clemence 2018, 2019, 2020). As such, it is important to understand the incentives that government has to maintain good quality public services. In this thesis, I study the example of government-funded health provision in England.

In this thesis, I will draw on the literature and existing theories about public opinion and voting behaviour to carry out a series of empirical studies into whether voters are able to evaluate public services based on their performance[1], and under what conditions they do so.

[1] As I will discuss in chapter 2, there are a number of separate but similar concepts in surveys and in the literature relating to how voters evaluate services. In this thesis, I use the word "evaluations" as a broad term that encompasses people's responses to questions about their personal satisfaction, as well as those that ask voters their views on objective service performance (for example, if services are getting better or worse). Where I use the term "satisfaction", this relates to a specific question's wording (e.g. the British Social Attitudes' NHSSAT question, which specifically asks how satisfied the respondent is with the NHS), but this is treated as a particular type of "evaluation".

These studies focus on potential factors drawn from a synthesis of the existing theoretical and empirical literatures on voter behaviour and government performance. The theoretical account that underpins the study is one of 'moderated learning': in brief, I use this term to encompass a range of viewpoints that all have in common the idea that voters have some ability to recognise, process and react to information about government performance. The moderation of that learning may come in a number of different ways: voters are very likely to balance their aims for accuracy with the aim to maintain a coherent world view that fits their partisan identity; they may use heuristics to reach conclusions based on limited data; they may have cognitive limitations that mean only some new information can be accounted for; they may place very little weight on information that contradicts their priors, and so on. However, despite these limitations and biases, a moderated learning approach expects that changes in government performance can be at least partly reflected in changes to public evaluations of services, if sufficient information about service performance ('performance signals') are received.

Voters therefore combine their partisan-informed priors (e.g., government supporters are generally satisfied with public service performance, while opposition supporters are generally dissatisfied) with new information, whether from personal experience or indirect information. I also hypothesize that information about National Health Service (NHS) performance is encountered relatively infrequently by the median voter; I therefore expect both partisanship and performance to play a role in determining voters' evaluations of the NHS, but that voters may not be attuned to very small differences in performance.

In order to test this theory's primary hypotheses, I will make use of high-quality national survey data from three gold-standard survey series: the annual British Social Attitudes cross-sectional survey, the British Election Study Internet Panel, and the annual Understanding Society panel survey. Each of these contain relevant variables on the political and demographic make-up of the respondents, as well as respondents' answers to useful questions about their views on health services. The different types of survey, question wording and coverage mean that each is useful in helping to study different aspects of the overall research question.

Information about public opinion derived from the survey data will be combined with infor-

mation about individuals' partisanship and recent contact with the NHS. I also combine it with information about the external information environment, including official NHS performance statistics and information on planned hospital reconfigurations, to create a series of statistical models and data visualisations that estimate parameters relevant to the testing of my hypotheses about how voters evaluate NHS performance. These can be broadly categorised into descriptive models, which draw on repeated cross-section survey data to find associations between variables using linear regression, and quasi-experimental models, which draw on panel data and exploit differences within individuals over time to estimate causal treatment effects.

The empirical chapters start by examining whether one of the main potential barriers to moderated learning, political partisanship, is so influential on voters' evaluations as to "crowd out" the possibility of voters learning from differences in public service performance. If partisanship is the sole driver of opinion and causes voters to be heavily entrenched in their viewpoints, this severely limits the possibilities for moderated learning about the NHS. Afterwards, the thesis moves on to looking at a set of different mechanisms by which information about variation in public service performance could be communicated to voters, namely information about service performance and provision, potentially communicated via the media or by word of mouth; and personal contact with the health service, which may provide an opportunity for voters to update their views through first-hand experience. The final empirical chapter looks at a particular case in 2016-17 where information about proposed hospital downgrades and reconfigurations in some areas of the country could have impacted people's evaluations.

Overall, the thesis finds that, consistent with a moderated learning account, voters are indeed able to recognise differences in public service performance and change their evaluations accordingly. In years when national waiting time statistics are better, voters of all partisan groups are more satisfied with the NHS; similarly, people living in areas with better waiting times are also slightly more satisfied with the NHS. Voters with recent experience of inpatient services are on average more satisfied with them than others, which suggests that people also take their personal experience of service into account when evaluating the NHS. However, there are factors moderating this learning process: voters' views are also strongly influenced by their partisanship, and planned hospital and A&E closures did not provoke differences in

attitudes or voting behaviour locally in affected parts of England in 2016-17. This suggests that differences in performance need to be clear and consistent in order to elicit changes in people's evaluations of services.

1.1 Scope

In this thesis, I limit my analysis to England, for two main reasons. The first is that Scotland, Wales and Northern Ireland have separate health services with their own operating models and performance targets, despite having the similar NHS branding and funding arrangements (particularly in Scotland and Wales). The second is that I previously worked in NHS commissioning in England, and therefore am most familiar with that system. Some of this chapter's statistics relate to the UK health system overall, but elsewhere in the thesis all statistics relate to England alone.

I have also taken the decision to only study public opinion *before* the COVID-19 pandemic that began in 2020. While there are no doubt many papers and theses that could be written about public opinion in that period, I argue that the conditions of the pandemic, particularly during lockdown, meant that the NHS played a very different role in society, and so the main phenomena being measured - NHS performance and public evaluations of how the health service is being managed - would be too different in their meaning to be comparable to the pre-2020 data. In addition, some regular surveys, including the British Social Attitudes survey, were disrupted by the pandemic, meaning less data was available.

Finally, I have chosen to focus primarily on how voters evaluate public services, rather than on how they assign responsibility for performance or how this affects vote choice. Further details on this choice in the context of the broader voting behaviour literatures are given in chapter 4's theory discussion.

1.2 Chapter summaries

1.2.1 Chapter 2: A review of the literature on voters' evaluations of public services

In this chapter I give a more detailed treatment of why voters' evaluations of public services matter, in light of the implications of retrospective voting theory for government accountability. I discuss the various types of metric that are in use to measure different facets of public opinion as it relates to health services in particular. Two broad approaches to voting behaviour in the literature (the social-psychological approach and the valence or rational choice approach) are summarised, and the expectations about public service performance's relationship to public opinion that follow from those theories is outlined. The social-psychological approach argues that partisanship or party identity is the primary force driving political behaviours and beliefs, and therefore voters' views on public services will be largely determined by their political self-identification. Valence or rational choice approaches, meanwhile, place more emphasis on voters' evaluations of government competence and ability to respond to poor performance at the ballot box.

1.2.2 Chapter 3: The case of the National Health Service in England

In the third chapter of the thesis, I argue that the English NHS is a good case study to test whether voters evaluate public service provision. I present evidence that health care is a highly salient political issue in the UK, which means people will form and maintain opinions about it; that successive governments have created performance metrics about the NHS and aimed to drive public awareness of those metrics in order to incentivise improvements; and that there is significant over-time and geographic variation in performance, which should lead to noticeable variation in evaluations of the NHS among the public. To support the latter argument, I present data on median waiting times between the late 1980s and 2019, explaining how the metric is defined and creating an 'indexed' waiting time measure that aims to allow comparison across the whole time period, mitigating the impact of changes to metric definitions in the late 2000s.

The measures show that waiting times were around four times longer in the late 1980s than in 2008-2010, supporting my argument that there has been sufficient over-time variation in NHS performance for voters to notice and respond to in a measurable way.

1.2.3 Chapter 4: A theory of moderated learning about public service performance

Chapter 4 sets out the thesis' theoretical model of how voters might evaluate public services. Five primary hypotheses are posited throughout the chapter, based on expectations derived from the theory. The chapter begins by briefly summarising the key components of the social-psychological and rational choice approaches, and what they imply for public evaluations of health services. Elements of these approaches are then combined into a broad set of theoretical expectations about how voters evaluate public services. Assumptions about voters are enumerated: their goals and motivations, the constraints on their behaviour, a model of how they process information, and some factors that affect how likely they are to incorporate new information into their opinions. The sources of performance signals are then described and their implications for the theory laid out. These are: personal experience, changes in government, and indirect information. Finally, the chapter explains how changes in the information environment may alter voters' evaluations of health services, and how those evaluations can affect vote choice and political behaviour.

1.2.4 Chapter 5: The impact of partisanship on satisfaction with publicly provided health services

In chapter 5, I carry out descriptive analysis to explore the extent to which partisanship is associated with differences in voters' satisfaction with the NHS. Using individual-level survey data from 1983 - 2018, I will use questions about satisfaction with how the NHS is being managed to measure differences between different groups of partisans over time. This analysis will demonstrate that in general, supporters of the government are more likely to be satisfied

with how the NHS is being managed - in line with expectations about the role of political partisanship. However, it will also show that the influence of partisanship varies over time and between party groups; notably, gaps in satisfaction between supporters of different parties have shrunk over time. In addition, I explore how while changes in government sometimes elicit rapid changes in levels of satisfaction among partisans, this is not always the case, which shows again that the role of partisanship in influencing public opinion about public services is context-dependent and changes over time. I conclude that, as predicted by the theory, while partisanship remains an influential factor, other explanations for differences in evaluations of public services are required.

1.2.5 Chapter 6: NHS performance and public evaluations of the NHS

In this chapter I examine whether voters evaluate health services based on their performance. By combining public opinion data from individual-level surveys with median hospital waiting times in regression models, I estimate how respondents' satisfaction with the health service varies as waiting times differ, both from year to year and geographically. The over-time analysis finds that respondents are less satisfied with the NHS in years when median waiting times are longer, and that this association is of a similar magnitude to that of partisanship. In the geographical analysis, I find that there is a very small association between longer average waiting times for hospital treatment within a local area and worse evaluations of the NHS among local respondents. Taken together, these findings demonstrate that both geographical and temporal variation in performance are associated with differences in NHS evaluations, indicating that people are able to draw on indirect information about health service performance to make judgements about government management of the service.

1.2.6 Chapter 7: Does personal experience change voters' evaluations of health services?

Having shown that indirect performance information is associated with people's views of services, I then go on to test the effect of personal experience of health services. This tests the idea

that personal contact reduces the influence of partisanship on voters' evaluations of services by yielding credible signals about performance. I draw on both cross-sectional and panel data in which respondents are asked whether they have been hospital inpatients in the previous year, and to evaluate health services. In both cases, I find that people who have recently been in hospital give better evaluations than those who have not. This finding is robust to controls for age group and partisanship and persists in within-person models, building confidence that it reflects a causal effect. However, the finding does not apply to non-partisans. Possible explanations for why recent hospital experience may almost uniformly result in higher evaluations of services are discussed.

1.2.7 Chapter 8: Do voters respond to planned health service cuts?

In the final empirical chapter of the thesis, I use panel data to estimate whether voters punished incumbent politicians for changes in local health services. I use as a case study the 2016-17 period, in which a set of plans were published to downgrade or close hospital services in some areas of England, approximately half a year before a snap General Election. By measuring changes in opinions and voting intention before and after the plans' publication, both in affected local areas and unaffected areas, this chapter provides a test for whether voters take changes in local health service provision into account when evaluating governments, in particular when deciding who to vote for. Using a difference-in-differences approach with pre-processing, I find no significant difference in opinions or voting intention between people living in affected areas and those living elsewhere following the publication of the hospital downgrade plans. This indicates that while performance can in some circumstances affect public opinion, local service changes do not necessarily cause voters to update their evaluations of government performance or engage in punishment voting. Several potential reasons for this empirical finding are then discussed.

1.2.8 Chapter 9: Conclusions and questions: How voters evaluate public services

The thesis ends by recapping the progress of the thesis, providing a reminder of the main strands of literature and the basic elements of the 'moderated learning' theory put forward in chapters 2 and 3. Findings from each of the empirical chapters is summarised. I then discuss the overall findings from the thesis as a whole: that voters draw on a mixture of partisan-informed priors and new information about performance, gleaned either directly through service contact or indirectly, in order to evaluate public services. Common themes in the thesis' findings across different chapters are discussed. I reproduce a figure from chapter 5 to illustrate how the general pattern of how public opinion and performance evolved between the late 1980s and 2019 supports the main tenets of the thesis. I end the chapter and finish the thesis by describing two questions that further research - underpinned by newly collected data - could help to answer.

2 A review of the literature on voters' evaluations of public services

The purpose of this thesis is to shed more light on the question of how people evaluate public services in democratic societies. In this chapter, I first discuss what is meant by voters' evaluations of public services, and why it is important. I will then present some of the key findings from the political science literature into how voters evaluate public services.

In this thesis, voter evaluations of public services are the main phenomenon under study. I treat 'evaluations' as an overarching label that includes several potential dimensions: satisfaction with how a service is performing, belief that the service is improving, and approval of the service's current management. Different data sources ask different questions and therefore are closer to one of these dimensions than the other - for example, the British Social Attitudes survey series asks specifically about satisfaction with how the National Health Service (NHS) runs. However, all fall within my umbrella concept of evaluations. Although there may be slight differences between these dimensions (e.g. a voter may not think a service is improving, but approves of its maintained high performance), from an empirical perspective I will be comparing variation within these dimensions rather than across them - e.g. how much satisfaction has changed over time, or how evaluations of service performance differ between demographic groups.

I conceive of these evaluations as being an internal value assigned to how well a service is performing at a particular point in time. Voters may cast their minds back to the past and compare how well they believe a service is performing now compared to then; or draw comparisons be-

tween local performance in different areas; or consider whether performance might be different under a different government. We can measure voters' evaluations directly through survey questions, and discern changes over time within individuals, groups or society in general by repeating the survey.

As is discussed further in the following chapter, in the healthcare sphere there are many measures of 'patient satisfaction' that aim to measure evaluations of particular services and individual experiences of care; individual public service providers in other sectors, such as transport, also have their own customer satisfaction metrics. In this thesis, however, the focus is on how voters evaluate the national government's management of public services.

In this chapter, I will first explain why it is important to understand how accountability works in democracies. I will then provide a brief overview of the theoretical link between accountability and retrospective voting, outline the steps in the retrospective voting process, and then summarise the differences and relationships between retrospective voting, performance voting and other closely linked concepts in the literature.

I will then summarise some key objections and challenges to retrospective voting that arise from the social-psychological literature. Following this, I will review some key findings from empirical studies testing retrospective voting theory's applicability to different contexts, and highlight several key areas of ongoing study.

In the final sections of the chapter, I focus on valence theory. I first explain how this approach has adapted retrospective voting to accommodate some of the challenges posed by the social-psychological literature and by the empirical findings reviewed previously. I will then go into more detail to examine the valence literature's contributions to our understanding of how voters evaluate UK governments and public service provision. The chapter will then end with a general summary.

2.1 Retrospective voting and accountability

In this thesis, I will examine whether voters are able to evaluate government performance in providing public services. This evaluation is the first necessary step towards creating governmental accountability for performance, and encouraging politicians to make policy choices that result in good quality public services.

I argue that voters' evaluations of public services matter because of democratic accountability. If voters are unable to hold governments accountable for provision of quality public services, this has negative implications both for instrumental and non-instrumental reasons.

From an instrumental perspective, governments provide important services upon which society relies. If voters are unable to properly evaluate and hold governments to account for the quality and availability of this provision, it is liable to be eroded or removed over time, because voters would not punish politicians who failed to maintain good public service performance. Democratic accountability, then, facilitates the provision of positive outcomes for the public.

From a non-instrumental perspective, governments derive their legitimacy from the idea that they bring together the collective will of the people, and produce a benefit to the common good (Freeman 2000). If the electorate is unable to make political choices that result in the outcomes they desire, then it is unclear how governments can represent the will of the people. Without accountability for policy choices, the claim of governments to hold democratic legitimacy is threatened. If we accept that democracy is a good in and of itself, then accountability - an important feature of how the public will is translated into governmental action - is also a good in and of itself.

How can voters induce accountability? Manin, Przeworski & Stokes (1999) explain that accountability can be produced through the process of retrospective voting:

> 'Governments are "accountable" if voters can discern whether governments are acting in their interest and sanction them appropriately, so that those incumbents who act in the best interest of citizens win re-election and those who do not lose them.' (Manin, Przeworski & Stokes 1999 p.40)

Therefore, according to this simple 'reward or punish' model of retrospective performance voting, at each election, voters can choose to reward governments for implementing policies that have yielded good outcomes, or punish them for failures. This occurs through a four-step process:

- Firstly, voters recognise conditions or outcomes in their lives and the world around them, and choices that elected politicians make.
- Secondly, voters make a link between the outcomes or conditions they have noticed and the policy choices of political actors, and therefore assign credit or blame to the government.
- Thirdly, they take account of these judgements when making vote choices, punishing those who they blame for poor outcomes and rewarding positive outcomes.
- Finally, politicians take note of the changes in vote choice and alter their actions in order to deliver positive outcomes and thereby increase their chances of political survival (Anderson 2007; Bernardi 2018; Healy & Malhotra 2013).

The idea of retrospective voting sits within a broader framework of performance voting, and is closely related to economic voting and valence models. Figure 1 in Stiers (2022) provides a useful guide to the theoretical distinctions between these concepts: performance voting is the broadest concept, and refers to the idea that voters evaluate political actors' performance and take this into account when voting. Retrospective voting is the specific type of performance voting described above, in which voters use past performance to guide their vote - as opposed to prospective voting, which is based on future expectations. Economic voting is again part of the broader family of performance voting, but on the issue of economic management, and can be retrospective or not (Stiers 2022).

The difference between spatial and valence voting also cuts across these definitions. In spatial models, voters' views exist on a continuum, and voters aim to maximise their utility by seeking policies that are ideologically as close as possible to their preferences on an ideological spectrum, within certain constraints (Downs 1957; Stokes 1963). Valence models, meanwhile, describe competition between candidates over who can best deliver good outcomes on issues

where there is broad consensus as to what a good outcome looks like (Clarke, Sanders, Stewart, *et al.* 2011). Issues often have both spatial and valence elements; Clarke, Sanders, Stewart, *et al.* (2009) gives the example that on the economy, there is a valence dimension over who will deliver high growth, low inflation and stability, and a spatial dimension as to how the proceeds of growth are distributed in society. Issues can also change over time between being primarily valence issues and primarily spatial (Green & Jennings 2017b). Applying this to a public service context, we could hypothesise that there might be a valence dimension where parties compete over who will deliver quality services, and spatial competition over market involvement in that provision.

2.2 Challenges to retrospective voting

The theories described in this chapter so far have all stemmed from what could be broadly described as 'rational choice' approaches. Whereas the rational choice approaches have their roots in economics, an alternative perspective on performance politics is offered by the social-psychological literature, which has its roots in sociology. These approaches place emphasis on the role social group identity, and in particular political partisanship, plays in determining political behaviour.

In Campbell's seminal work The American Voter, partisanship (or party identity) is defined as a "sense of attachment" or a "psychological tie" to a particular party (Campbell 1960 p.121). This is conceptualised as being similar to the group attachment an individual feels to religious or class identity, and that this emerges because of 'political socialisation' at a young age, through the influences of family and social milieu. People can identify to a stronger or lesser degree with a party, which is conceived of as a continuum; those who identify more strongly with a party are more likely to agree with their party's issue positions and vote for them (Campbell 1954, 1960).

How does this relate to evaluations of public service performance? Crucially, this account posits partisanship as an "unmoved mover" that acts as a "perceptual screen", with the voter

tending to "see what is favorable to [their] partisan orientation" (Campbell 1960 p.133). More recent empirical evidence indicates that partisanship is not only associated with differences in opinion (such as whether Jimmy Carter was knowledgeable), but also in differences in perceptions of objective fact: e.g., whether the deficit had increased, or crime rate had increased (Bartels 2002). The argument made by Evans & Chzhen (2016a) is that when people are expressing positive views about government performance, they are doing so because of their partisan beliefs. Rather than evaluating conditions and deciding how to vote based on those evaluations, voters start from a position of 'belonging' to a certain party, and arrange their opinions to be consistent with that sense of belonging.

This arrangement may not necessarily be conscious. There are a number of mechanisms at various stages in the process of information exposure and processing that could result in this screening effect. Firstly, because of established partisan differences in media consumption, voters of different party identities are exposed to different information (both factual and persuasive) about government performance (Hart, Albarracín, Eagly, *et al.* 2009; Knobloch-Westerwick & Meng 2009; Lazarsfeld, Berelson & Gaudet 1968). This difference in media consumption may arise due to variation in upbringing, taste, and social group, or through actively seeking out media that reinforce one's own world view rather than challenge it. Either way, 'selective exposure' to pertinent information helps voters maintain attitudes that align with their political and social identity, as well as the information they already believe.

When partisans are exposed to information, they then process it in a way that is influenced not only by a desire to uncover truth, but also to fit with their pre-existing partisan preferences (Taber & Lodge 2006). Taber, Cann & Kucsova (2009) p.139 breaks this 'motivated reasoning' down into three tendencies: an 'attitude congruency bias', where people are more likely to uncritically accept evidence supporting their priors, a 'disconfirmation bias', where people seek to discredit evidence contradicting their priors, and 'attitude polarisation', wherein the first two processes lead to opposing partisans diverging in their opinions based on the same information.

Note that 'disconfirmation' does not necessarily mean simply choosing not to believe a new

piece of information. Facts can be interpreted differently to accord with a person's priors; for example, through selective attribution of credit or blame. A government supporter is less likely to attribute blame for poor performance to the government, an opposition supporter more likely (Marsh & Tilley 2010). In the context of public services, there are other actors – managers, staff, etc – who could be used to deflect blame for negative service changes (Pierson 1996). The key is that people tend to expend more time and energy scrutinizing and critiquing new ideas that conflict with their existing beliefs than those that don't (Schaffner & Roche 2017). However, strong evidence can induce even partisans to change their opinion when they reach an 'affective tipping point' (Redlawsk, Civettini & Emmerson 2010).

In summary, social-psychological theories of party identity argue that partisanship is the most influential factor overall in determining voting behaviour, and that evaluations of performance are a result of partisan identity. If this account is accurate, this poses a challenge to the idea that voters could evaluate public service performance , and ultimately to the idea of retrospective voting.

2.3 What we know about voters' evaluations

2.3.1 Contingent economic voting

Although the original theoretical frameworks underpinning retrospective voting have needed to be revised in light of the limitations described above, the fact remains that retrospective voting has been observed — especially in relation to economic outcomes — in a range on contexts, in both developed and transitional economies (Lewis-Beck & Stegmaier 2007, 2008; Wilkin, Haller & Norpoth 1997). A recent analysis argues that economic voting in Presidential elections has been evident to some extent in the entirety of U.S. history (Guntermann, Lenz & Myers 2021). However, there are cases where economic voting is not observed or is weaker than other explanatory factors, which suggests that economic voting is not universal, but contingent on the particular context (Anderson 2007). As a result, recent research is aimed at exploring the contours of these contingencies, for example: under what circumstances does

retrospective voting emerge? To what extent do institutional differences lead to challenges to each part of the four-step accountability chain?

Looking at the first step, perception of change, there is debate over whether voters' evaluations of the economy are based on national economic conditions, a theory known as sociotropic voting (Kinder & Kiewiet 1981), or their own personal financial circumstances, which is called pocketbook or egocentric voting (Tilley, Neundorf & Hobolt 2018). In general, studies have found more support for sociotropic voting than egotropic voting (Lewis-Beck & Paldam 2000), with some exceptions, such as the 1997 UK General Election (Sanders 1999). Recent work has found that retrospective voting also occurs at levels between the individual and national, such as the community, region or sub-federal state (Books & Prysby 1999). Pattie and Johnston have a series of articles citing evidence from the British Election Studies of the 1992 and 1997 General Elections to demonstrate that voters take account of conditions in their region or local area when evaluating government performance on the economy (Johnston & Pattie 1997, 2001a; Johnston, Pattie, Dorling, *et al.* 2000; Pattie & Johnston 1998).

Another area of debate is over heterogeneity in voters' evaluations within the same country: Gomez and Wilson (2001, 2006) have carried out a series of analyses to argue that, contrary to previous assumptions, low-sophistication voters are more likely to evaluate the economy sociotropically, with only politically sophisticated voters engaging in egocentric evaluations.

Studies also explore which aspects of the economy matter to voters' evaluations. Meta-analysis of the literature has found that studies of economic voting often find that voters react to differences in unemployment, inflation, and economic growth (Lewis-Beck & Stegmaier 2019), with the first of these being the most important (Lewis-Beck & Paldam 2000).

Moving on to the second step, attribution of credit or blame, there is strong support for the idea that voters' ability to assign responsibility for policy outcomes is key to retrospective voting. Some institutional arrangements impede this ability: dispersion of political power among different levels of government makes economic voting less likely to be observed (Anderson 2006), and coalition governments also make economic voting less likely to emerge — especially those with many parties involved (Fisher & Hobolt 2010). Some studies have used 'clarity of

responsibility' indices to show that other factors including ideological coherence and government longevity, together with the majoritarian-coalitional divide, result in greater and lesser economic voting across and within countries (Nadeau, Niemi & Yoshinaka 2002).

The final step, in which voters translate credit or blame into differences in voting behaviour, can also be affected by context. One such factor here is the electoral system: in highly proportional systems with a high effective number of parties such as the Netherlands, as well as making it more difficult to attribute credit or blame for policies, the improbability of removing particular parties entirely from government seems to result in voters using issue-priority considerations to decide their vote, rather than retrospective voting (Dorussen & Taylor 2001). In the issue-priority model, parties compete to increase the salience of "their" issues - the issues where they believe they perform better with the public, and the public votes for the party who has the best reputation on the issues they see as a priority at that time (Dorussen & Taylor 2001; Whiteley, Stewart, Sanders, *et al.* 2005). There is therefore considerable empirical evidence for the phenomenon of retrospective voting on economic issues, and a degree of consensus on some of the main factors affecting whether it emerges in a particular setting.

2.3.2 Valence models and the limits of rationality

Some of the assumptions associated with the original rational choice approaches to politics have been contradicted or challenged, whether by the social-psychological critiques above or by empirical research testing these approaches. As a result, some assumptions have been either relaxed or replaced in modern retrospective voting literature. In this thesis, a particularly important current strand of the literature is that of valence models - because of their focus on the UK context and ready applicability to public services. Broadly speaking, valence models are a sub-group of models within the retrospective voting framework that emphasize the importance of voters' evaluations of government performance on valence issues, and which accept many of the critiques that the social-psychological approach have made of strict rational choice models. In this section I outline some of the ways in which valence models have adapted retrospective voting to accommodate empirical findings about the limits to voter rationality; the

next section looks in more detail at the findings of the valence literature as it relates to the UK context and to public services.

The term *valence* was originally borrowed by Stokes (1963) from the psychological researcher Kurt Lewin, and used to refer to issues on which parties compete to demonstrate competence in accomplishing a broadly agreed-upon good outcome, rather than proposing alternative policy formulations. For example, the provision of a responsive, well-functioning government is a valence issue, since it is a broadly agreed-upon aim that all parties aim to achieve. Valence issues are set in contrast to *positional* or *spatial* issues such as that proposed by Downs (1957) in which parties aim to locate their policies optimally on a spectrum (e.g. the left-right economic dimension) to attract as many voters as possible.

One the assumptions of strict rational choice models that valence models discard is the concept of the individual making choices in order to maximise their utility. Critics of strict rationality-based approaches were pointing out as early as the 1950s that humans simply do not have the capacity to gather and process all the information necessary to make optimal choices in complex scenarios - instead, they aim to satisfice, selecting the option that is 'good enough' (Bendor 2003; Stevens 2019). Evidence from psychology shows that conscious information processing takes place in 'working memory', which is dependent on attention and executive action and has a limited capacity, making decision-making a mixture of consciously processing a few salient facts while subconsciously drawing on memories (Chai, Abd Hamid & Abdullah 2018; Lau & Redlawsk 2006).

In response to these critiques, valence theory accepts that voters are not fully rational in their information gathering. Clarke, Sanders, Stewart, *et al.* (2009) describe voters are 'cognitive misers' who prefer to use heuristics or cognitive short-cuts to make decisions, rather than expend time and energy gathering new information unnecessarily; and that valence issues' simplicity in contrast to spatial issues may allow less informed or politically sophisticated voters to make judgements more easily (Clarke, Sanders, Stewart, *et al.* 2009). Voters use partisanship (which per Fiorina (1981), is seen as a long-run accumulation or 'storehouse' of previous evaluations) and leader images (the perceived competence of the party leader) as heuristics when deciding

how to vote. Partisanship is therefore dynamic, changeable, and interacts with other factors that themselves affect vote choice (Whiteley, Clarke, Sanders, *et al.* 2016).

In addition to these heuristics, people also draw on their direct experiences of the outcomes of government policy, their evaluations of government performance, and their emotional reactions to these outcomes (Clarke, Sanders, Stewart, *et al.* 2009 p.18). However, these heuristics are not perfect, and voters remain subject to cognitive biases such as the grievance asymmetry, in which they tend to react more to negative stimuli than positive ones (Nannestad & Paldam 1997), and recency bias - which is that voters *intend* to evaluate leaders on the overall trajectory of the economy during their tenure, but in practice generally only use conditions in the election year in their evaluation process (Healy & Lenz 2014). The bias introduced by partisanship and voters' cognitive limitations notwithstanding, voters do still often demonstrate an ability to respond to differences in performance (Bailey 2019 ; De Vries, Hobolt & Tilley 2018 ; Hill 2017).

2.3.3 Valence models in the UK and beyond

As mentioned earlier, the valence model of voting argues that while other factors such as spatial voting and social class affect vote choice, it is the valence dimension, voters' perceptions of competence, that have been most influential in elections for decades (Clarke, Sanders, Stewart, *et al.* 2004). Within that general concept of competence, different valence theorists have proposed several different (but sometimes overlapping) ways of thinking about how voters assess performance and make judgements about candidates. This section describes some of these in more detail.

Green and Jennings (2017b) divide 'competence' into three separate concepts: issue ownership (which issues are typically seen as being strengths for particular parties), government performance (short-term fluctuations in public opinion about government competence based on the current policy environment) and generalized competence (an underlying view of the competence of a party across issues generally).

In their chapter on issue ownership, they find that changes in views on government competence tend to be followed by changes in partisanship, rather than the other way round – in other words, information about past government competence is more predictive of future partisanship than is past partisanship (Green & Jennings (2017b), contra Evans & Chzhen (2016b)). However, analysis of the link between government performance and voters' opinions found that voters only responded to changes in performance in some contexts and not others (Green & Jennings 2017b). This is argued to be because different contexts give rise to different levels of salience for different issues:

> '[W]hen a political leader or party suffers a negative competence shock, we should expect competence to be seized upon politically, to be salient, and therefore most relevant to public opinion and vote choice. This takes us back to Stokes (1963), who argued that valence issues were those issues that shape political competition in a given period because performance and competence become the salient and most relevant political evaluation.'(Green & Jennings 2017b p.197)

This explanation draws on the issue-priority model described earlier. As well as Green and Jennings' findings, there is also empirical evidence from Canadian elections showing that issue ownership only affects vote choice among voters who prioritise that particular issue (Bélanger & Meguid 2008).

In terms of day-to-day government performance on issues, Green & Jennings (2017b) measure correlations between government performance and how well voters rated them on handling of issues across several countries. They found that in general, governing parties tend to be rated more highly in years when performance is better, and that improvements in performance are recognised by the public. This demonstrates that there is a general pattern of voters being able to evaluate government performance across a number of policy domains and contexts. Macro-competence or generalised competence is understood to be the underlying trend in how voters rate a party's handling of many issues; this is operationalised by combining voters' issue ratings for a government or party across different performance domains. These two proposed concepts, and how they relate to the theory underpinning this thesis, are discussed further in

chapter 4.

In addition to exploring different aspects of valence as described above, there are also strands of the valence literature that examine how the features of retrospective voting described earlier (sociotropic vs. egotropic, local vs.. national, and differences institutional contexts) apply in the UK context.

Although primarily focused on the Westminster context with some American applications, there are cases of valence theory being tested in devolved Parliaments and in other countries. Individual-level survey data suggests that valence perceptions were crucial in allowing the Scottish National Party to gain power with support from pro-devolution voters sceptical of independence, since the party scored well on performance evaluations, economic competence and party image compared to Labour (Johns, Mitchell, Denver, *et al.* 2009). In the 2013 German election, the Free Democratic Party performed poorly because while in spatial terms its policy was well positioned, it had a poor reputation for valence (Zur 2017). Analysis of voting in Portugal's 2011 elections directly tested the valence, spatial and patrimony variations of economic voting theory (Lobo 2013). Here, "patrimonial" voting refers to the idea in Lewis-Beck & Nadeau (2011) that people who own more property (land, housing, shares) vote differently, in particular to support pro-free-market candidates. Lobo (2013) finds evidence for all three, but that valence voting is the most explanatory. Similarly, valence factors have also been found to have been. important in 2008 elections in Taiwan (Ho, Clarke, Chen, *et al.* 2013).

The valence model's focus on government competence means that some studies have been carried out into public service performance, including healthcare. However, this has usually taken the form of assessing performance of public services as part of a composite measure or combined analysis across a range of domains, whereas this thesis will look in detail at healthcare performance through multiple analytical lenses. Green & Jennings (2017b) include length of NHS waiting list as one of their government competence metrics and find that governments tend to be rated more highly in years when these metrics are better. They also found that NHS performance was the only issue where differences in performance also affected the opposition party's perceived issue handling (Green & Jennings 2017b).

Johnston and Pattie (2001b) offers a rare example of a study examining voters' responses to public service performance. They argue that while American work on retrospective voting such as Fiorina (1981) analysed a range of valence issues, little attention has been paid to potential public service performance voting in the UK, with the literature's focus generally narrowing to economic issues. Noting that the 1997 UK General Election had a heavy focus on the issues of education and health, the paper uses voters' evaluations of services and attributions of credit and blame from British Election Study data to examine the extent of retrospective voting on different policy issues (Johnston & Pattie 2001b). While prior partisanship has some effect - people who supported the Conservatives in 1992 were less likely to give the Conservative government negative performance ratings in the 1992 to 1997 period - there is an association between negative performance evaluations and subsequent vote choice in 1997, suggesting support for the reward-punish model. Performance evaluations of public services (including healthcare and education) were as important as economic evaluations in influencing vote choice, and in the case of prior Conservative voters, were even more important (Johnston & Pattie 2001b). The authors called for more research into retrospective evaluations of public services, a gap in the literature that this thesis aims to help address.

Finally and for completeness, it is important to note that outside the broad family of performance voting theories, there are some studies in the welfare state literature that do not necessarily take on the assumptions associated with retrospective voting, but which nonetheless measure voters' evaluations of public services and how they react to policy changes, in particular service cuts (retrenchment). A study from Sweden showed that incumbents in councils who cut health services were punished, especially in areas near the affected hospitals (Lindbom 2014); however, a similar study into Swedish school closures failed to find any punishment effect (Wänström, Karlsson & Wänström (2012) in Lindbom (2014)). Analysis of Turkish elections has found that better service access was associated with higher vote shares for incumbents (Adiguzel, Cansunar & Corekcioglu 2023), and cross-national comparison has found that social policy results in punishment voting among voters in different countries for whom social policy is highly salient (Giger 2012). In chapter 8, I will carry out a study into planned retrenchment of hospital services in England from a retrospective voting perspective, to deter-

mine whether plans to change provision result in changes to voters' evaluations of government performance.

2.4 Summary

In summary, for governments to be accountable to the electorate, it is necessary that voters can recognise and respond to differences in government performance. While some early theories posited that voters were fully rational in assessing and responding to information, this has been challenged both by social-psychological accounts of voter behaviour and by the empirical evidence. The social-psychological account argues that voters cannot objectively assess government performance because their opinions are heavily influenced by their desire to maintain views that are consistent with their partisan and social identities. Their views on government performance are therefore more driven by their existing political dispositions, rather than objective reality. Empirical research has also shown that voters are subject to other cognitive biases and limitations that make it unrealistic to expect voters to process large amounts of information about government performance fairly.

On the other hand, despite these limitations, associations between performance and voters' evaluations has been observed empirically in many different countries and time periods, especially in relation to economic outcomes. Valence theorists have argued that voters can use heuristics to gain enough information about governments to make informed evaluations about government performance, without needing to rely on unrealistic assumptions about rationality and perfect information. While the valence literature has included healthcare performance in its broad analyses of government performance, few studies have performed in-depth analysis of how voters evaluate public service performance, especially in a UK healthcare context. In the next chapter, I will describe why healthcare provision in England is an informative case for this type of research.

3 The case of the NHS in England

In order to explore how voters evaluate public services, I have chosen to study the case of the National Health Service (NHS) in England. In this chapter, I outline why there is a strong case that voters are more likely to perceive differences in performance in the NHS than in most other public services: firstly, because the NHS is salient; secondly, because there has been a deliberate effort by UK governments to measure the service's performance and make the metrics public; and thirdly, because there has been considerable over-time variation in health service performance within living memory (and there is also variation - on a smaller scale - between geographical areas in England). In doing so, I select a key set of performance metrics that I will use in Chapter 6 to measure the association between performance and satisfaction.

3.1 The NHS is a salient issue

Voters are more likely to notice and respond to differences in performance where the public service is a more salient issue. The UK's NHS is arguably among the most significant public services in the world. In 2021, 83% of healthcare provision in the UK was government-funded, i.e. through the NHS, local authorities or other Government bodies (Prendergast 2023), and its workforce numbers over one million people (Kirk-Wade, Mirza-Davies, Garratt, *et al.* 2023). It has been variously described as a "national religion" and "political football" in light of its political salience (Barer 2016).

One useful measure of salience is the Ipsos Issues Index, in which a representative sample of British adults are asked what they see as important issues facing the country, and their

unprompted responses are coded in to categories (Clemence & Skinner 2023). As alluded to in the introduction to this thesis, the "NHS/Hospitals/Healthcare" category was in the top three most mentioned issues in 2017, 2018 and 2019 (Clemence 2018, 2019, 2020), with no other public service being consistently that highly rated in the index.

Although highly salient, it should be noted that the debate on healthcare is largely a 'valence' issue on which parties compete over who will manage the service best, since there is widespread consensus on the key principles underpinning NHS funding and provision. Polling from IPSOS and the Health Foundation found that across the UK:

- "90% believe the NHS should be free at the point of delivery,
- 89% think the NHS should provide a comprehensive service available to everyone,

and

- 84% think the NHS should be funded primarily through taxation" (Lilburn 2023)

However, satisfaction with the day-to-day management of the health service waxes and wanes over time. In a nationally representative British Social Attitudes survey in 2022, only 29% of people were satisfied with how the NHS was run, while 51% were dissatisfied, the worst result on record (Morris, Shlepper, Dayan, *et al.* 2023). The most common reason for dissatisfaction (when respondents were allowed to pick three options from a list) was "It takes too long to get a GP or hospital appointment" (69%), followed by "Not enough NHS staff" (55%) and "The government doesn't spend enough money on the NHS" (50%) (Morris, Shlepper, Dayan, *et al.* 2023).

Taken together, this evidence suggests that while the underlying principles of the NHS are well supported, the issue of how well the NHS is being managed - service performance and capacity - are a contested and salient issue. As the most salient public service among the British public, I would argue that of all public services, the NHS is the one that is most likely to be subject to retrospective voting.

3.2 Governments measure and publicise performance

Over the years, the Department of Health and the quasi-autonomous or arm's length bodies it oversees (e.g. NHS England) have created and monitored many metrics that aim to measure some aspect of how well the health service has performed. Performance measurement is generally understood to have been introduced for public services in OECD countries as part of the movement towards New Public Management beginning in the late 1970s, which prioritised greater competition and evaluation within the public sector, enabled by measurable performance standards and measurement of outputs (Dahlstrom & Lapuente 2009). In the UK, rather than reverse the trend towards competition and measurement that started under the Conservatives, the New Labour government of 1997-2010 embraced the use of performance measures to enable greater 'patient choice' in a regulated market, alongside an historic rise in funding (Klein 2007). Publication of performance metrics, in conjunction with choice of provider and payments linked to activity, was seen as a way to improve standards by redirecting resources towards high performing hospitals (Department of Health 2006). Politically, the Blair government portrayed service reforms (including adherence to performance targets) as being delivered in exchange for increased funding levels (Klein 2007). The New Public Management-inspired culture of performance measurement and management in the NHS has been criticised for incentivising poor behaviour that superficially meets targets, neglecting unmeasured aspects of performance, and excessive focus on the short term (Mannion & Braithwaite 2012). However, is is likely that efforts to publicise these metrics and enable patient choice have led to greater public awareness of the issue of health service performance.

3.2.1 How survey and administrative data sources can measure performance

Two different types of metric are often used in the NHS to evaluate how well services are performing. The first are patient or public evaluations of services - how well people feel they are being served; the second is through measurement of administrative data.

Patient service evaluations are used at different levels (e.g. individual clinic, GP practice, hos-

pital, national) and can be carried out by the service provider themselves, or by an external organisation. They can be entirely bespoke (e.g. qualitative research carried out by a hospital to help redesign a particular service) or a standard national survey, such as the GP Patient Survey - a national survey that randomly samples patients registered within each GP practice in England, allowing monitoring of satisfaction with access to primary care (Ipsos 2023).

There is a significant body of work in the healthcare literature focused on defining and measuring the public's reported experiences of healthcare, especially at the level of the individual service or clinic - see e.g. Ahmed, Burt & Roland (2014), Batchelor, Owens, Read, *et al.* (1994), Bowers, Swan & Koehler (1994), Carr-Hill (1992) and Wolf, Niederhauser, Marshburn, *et al.* (2014). There are some critiques, though, of treating satisfaction surveys as reflective of an underlying objective 'quality' of the service. For example, one of the defining assumptions about the concept of patient satisfaction is that satisfaction is relative to expectations, but different population groups have different expectations and priorities, meaning there are systematic differences in what is required to satisfy patients (Ahmed, Burt & Roland 2014; Williams 1994). Similarly, patients appear to indicate that they are 'satisfied' with a service, even if they also recount having had negative experiences or there are failings in their care, perhaps because they feel that the service carried out their 'duty' by providing care, regardless of its quality (Jenkinson, Coulter, Bruster, *et al.* 2002; Williams, Coyle & Healy 1998).

As well as patient service evaluations, several national social attitudes surveys ask about how well the government is managing the NHS on a day-to-day basis. This differs from the measures described previously in three main ways. Firstly, it asks for evaluation of choices by party political actors who make policy choices, rather than clinicians who deliver care. By invoking government involvement, this type of question may activate a political partisan framing in the respondent's mind; Chapter 4 will discuss the influence of partisanship in more detail. Secondly, it asks for an evaluation of the national level of provision rather than any individual service provider, and so requires a broader judgement that transcends individual experiences of care. Finally, the context differs: because the question is more political, it is generally asked as part of general social and political surveys done by outside organisations rather than as part of an internal service evaluation process. This again may lead the respondent to answer as a

member of the public (activating a political framing) rather than as a patient.

One source of this kind of survey data is the British Social Attitudes (BSA) survey, which is used annually by The King's Fund to analyse overall trends in satisfaction with the NHS. The most recent King's Fund report based on BSA survey data before the COVID-19 pandemic found that satisfaction increased by 28 percentage points during the 2000s, with a sharp drop after 2010, followed by inconsistent patterns (Wellings, Appleby, Maguire, *et al.* 2020). The study also found that, with the exception of dental services, people who have recently used NHS services are more satisfied with them than people who have not, and that people aged 65+ were more satisfied than others (Wellings, Appleby, Maguire, *et al.* 2020). An earlier use of BSA data in a basic time-series analysis of the 1997-2010 Labour government found that as performance improved, perceptions of the NHS also improved (Maybin 2010). These examples show that the British Social Attitudes survey contains useful data for understanding public opinion on the NHS, although the full potential of the data has not yet been exploited. In chapters 5 - 7 of this thesis, I will make a more detailed examination of the British Social Attitudes data to explore the relationships between NHS evaluations and partisanship, performance metrics and personal contact with services.

It is not yet clear how people's "patient experience" in local services influences their perceptions of the service nationally. A study of European survey data found that patient experience explains around 10% of variation in satisfaction with the healthcare system, rising to 17.5% when accounting for some features of national systems, personality, expectations and selected other covariates (Bleich, Özaltin & Murray 2009). As pointed out by the authors, this may arise from the influence of external differences (such as media coverage and discourse) or because of conceptual differences between personal experience of local services, and the political abstraction of 'the health system'(Bleich, Özaltin & Murray 2009). There is also evidence that people who have recently used healthcare services are generally more satisfied than those who have not (Edwards 2006; Ziebland, Evans & Toynbee 2011). Laycock (2009) observes that there are often considerable differences between how people rate their own personal experience of health services and how they rate the NHS as a whole, and that this seems to be associated with demographic differences.

An alternative method to measure one dimension of patient experience is through performance metrics such as waiting times. These performance metrics are based on administrative data collected by the service providers according to specific guidance and standards, and reported back to bodies that commission and regulate services (e.g. NHS England). These measures have the advantage of not relying on members of the public's own judgements, and therefore they are in one sense less vulnerable to bias compared to survey responses. However,they only measure one facet of experience, and therefore impose the regulator's view of what is important, rather than taking account of what the respondent deems to be good performance.

In today's NHS, performance measures exist across primary care (General Practice surgeries, dentistry, optometry), acute secondary care (hospitals), community services (e.g. community podiatry), mental health services, and specialist services. However, the most long-established performance metrics are for hospital care.

Although primary care is the sector of the health service with the most patient contact, performance measures for GPs have generally focused on quality of care and population-level metrics, rather than activity-based measures. This is logical, because unlike hospital finance, GP contracts are primarily paid on a capitated basis: adjusted payments are made to GP practices based on the number of patients on their list, rather than for specific activity. By contrast, hospitals report numbers of clinical activities (episodes), which may form part of a longer inpatient pathway (Boyd, Cornish, Johnson, *et al.* 2017), and performance measures often require calculation of waiting times.

There is a dearth of research investigating the links between performance and public opinion, particularly research that accounts for potential confounding factors (e.g. demographic change). This thesis adds significant value to this area of study, by examining NHS performance data and public opinion surveys in the context of retrospective voting theory and using regression modelling (and some quasi-experimental designs) to draw conclusions about how voters evaluate services.

3.3 There is meaningful variation in NHS performance

In order for voters to be able to recognise and respond to differences in public sector performance, there needs to be a meaningful amount of variation in that performance for them to recognise. In this chapter I use hospital waiting time data to demonstrate that there has been meaningful variation in NHS performance over recent decades for which we also have public satisfaction data, as well as less pronounced variation between areas of England in the same time period. This is additional evidence that the NHS has the necessary conditions for voters to be able to perceive and respond to differences in performance. The performance metrics are selected and aggregated in a way that makes them suitable for analysis in combination with information on public satisfaction from British Social Attitudes and Understanding Society survey data in chapter 6.

3.3.1 Over-time variation in waiting times

In this section I examine how performance varies over time. As described above, hospital waiting times are a long-running feature of NHS performance management; in order to make comparisons that span the longest possible time period while also acting as a potential proxy for the broader functioning of the service as a whole, I therefore opt to use these as my performance measure. Using a metric with a long time horizon also makes it possible to observe the effect of long-term consistent flows of performance information associated with gradual change over time. Increased waiting times are associated with dissatisfaction, worse clinical outcomes and anxiety among patients (McIntyre & Chow 2020); and compared to clinical outcomes or other possible metrics, waiting times are more visible and measurable for the public, usually due to media coverage, including at election time (Dekavalla 2010). In a particularly high-profile and personalised example, a Labour Party Election broadcast in 1992 focused on a child's waiting time for surgery to treat glue ear, and the debate it caused became known as the battle, or war, of Jennifer's ear (Bennett 1996; Black 1992).

The Referral To Treatment (RTT) performance standard set by NHS England is for 92% of

people on the waiting list for consultant-led inpatient treatment to have been waiting less than 18 weeks , and so one possible metric that could be operationalised to represent 'performance' would be the percentage of patients waiting less than 18 weeks. However, there are two good reasons not to use this. Firstly, this would effectively dichotomise waiting time into "less than 18 weeks" and "more than 18 weeks" and reduce the information available to the analysis. Secondly, this metric is only available from 2008 onwards, which limits the scope for long-term analysis.

Instead, I will use the median waiting time for treatment in weeks as my proxy for performance. This metric does not have a performance standard attached, but has the advantages that it is a continuous variable, and has been measured and reported since 1988. In order to align with the British Social Attitudes survey data, which this thesis uses as its primary source for information on the public's satisfaction with the service, I compute an average of the median waiting time in weeks for the first 6 months of the calendar year in which the survey was taken, and exclude the second half of the year (when the survey has already been taken).

Two different sets of performance measures exist in the time covered by the BSA: from 1988 until the end of 2010, a set of waiting time measures was recorded for inpatient hospital treatment - referred to from this point as the "legacy" metric. From 2007 onwards, as measures and targets developed, "RTT" (Referral to Treatment) waiting times were recorded, which I will refer to as the "modern" metrics. There are some differences between the two metrics' methodology and in the waiting times they show during the period of time when they were both measured. Because of this, I will perform analysis on both sets of metrics separately for the 1988-2010 and 2007-2019 periods, and test whether the findings are consistent across both. This section provides more detail on the calculation of each type of metric, and how they differ.

The legacy metric measured the length of time between a patient being accepted by a consultant for inpatient treatment, and their admission to hospital. Generally, however, this would only take place after two or three other periods of waiting: the wait for a GP appointment, the wait between a GP referral and hospital outpatient assessment, and then potentially also a wait for diagnostic tests. At that point, if the patient was accepted for inpatient treatment, they were

placed on the "inpatient waiting list" and the waiting time "clock" started (National Audit Office 2001).

In the case of the modern metrics, the "clock" starts from the GP making a hospital referral (Department of Health 2023). The modern RTT metric therefore encompasses more of the total 'waiting' time from a patient perspective, and because of the increased emphasis on performance standards, was more rigorously regulated than the previous waiting times metrics.

The difference between the legacy metrics and the modern metrics is apparent from a comparison of the two during the 2007-2010 period in which both were recorded. Figure 3.1 shows the average waiting times for each type of metric after applying the criteria explained above: only observations between January and June are included, which are then combined to create an average for each year. It shows that the highest waiting times were in the late 1980s, when median waiting times were above 20 weeks. There were rapid falls in the legacy metric between 1989 and 1992 and between 2002 and 2008. The modern metric begins in 2008 with a considerable fall in 2009 (to be discussed in more detail below), but otherwise is characterised by a gradual rise during the 2010s, albeit with much smaller between-year differences than was common in the legacy metric.

For a more detailed comparison of the period covered by both the legacy and modern metrics, Figure 3.2 shows the original monthly waiting times data between 2007 and 2010 (inclusive).

Visual comparison of the trends suggests that there may have been a bedding-in period of around a year from the introduction of the modern metrics, in which there is a considerable difference between the two. The difference falls from around seven weeks to two weeks in mid-2008, narrowing slightly as the legacy series ends in 2010. This could be the result of poor data quality as the new metrics were introduced, or the effect of Trusts putting changes in place to improve the waiting times between referral and outpatient assessment (which were not previously reported).

From mid-2008, the seasonality and effects of shocks appear very similar between the two metrics. This suggests that although the modern metrics encompass more of the patient pathway

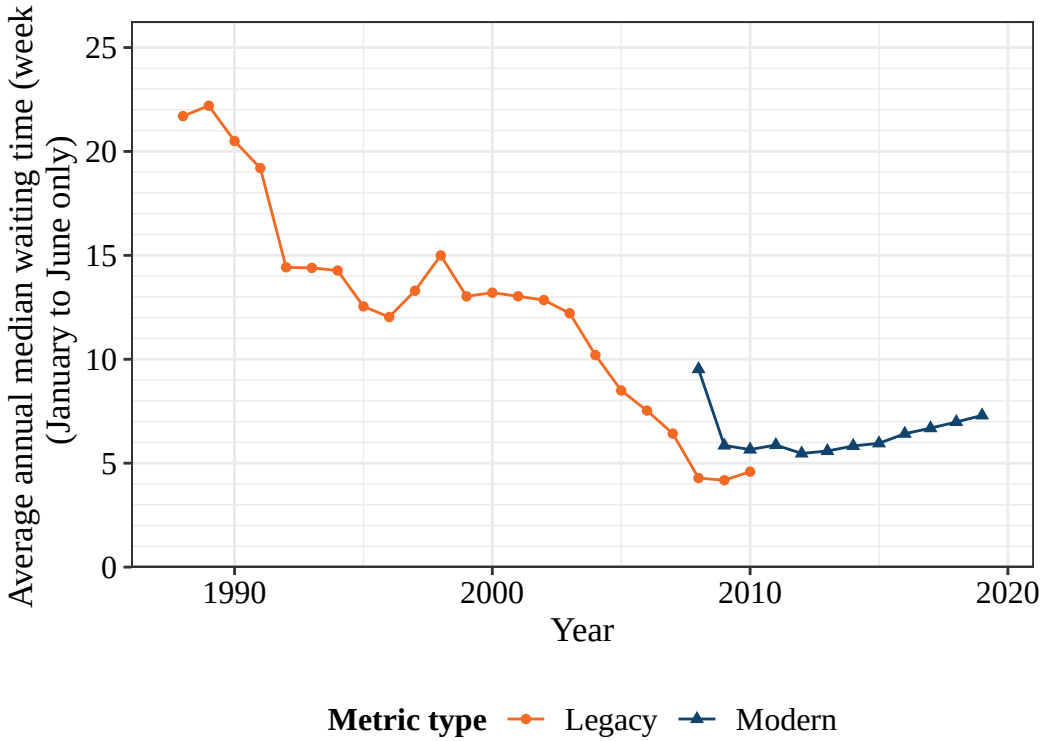

Figure 3.1: A comparison of median waiting times between the modern and legacy metrics, measured in weeks. The legacy metric spans 1988 through 2010 and generally fell over time, whereas the modern metric begins in 2008 and has generally been stable since 2010.

from referral to treatment, the two metrics are similar enough that they are measuring the same underlying process.

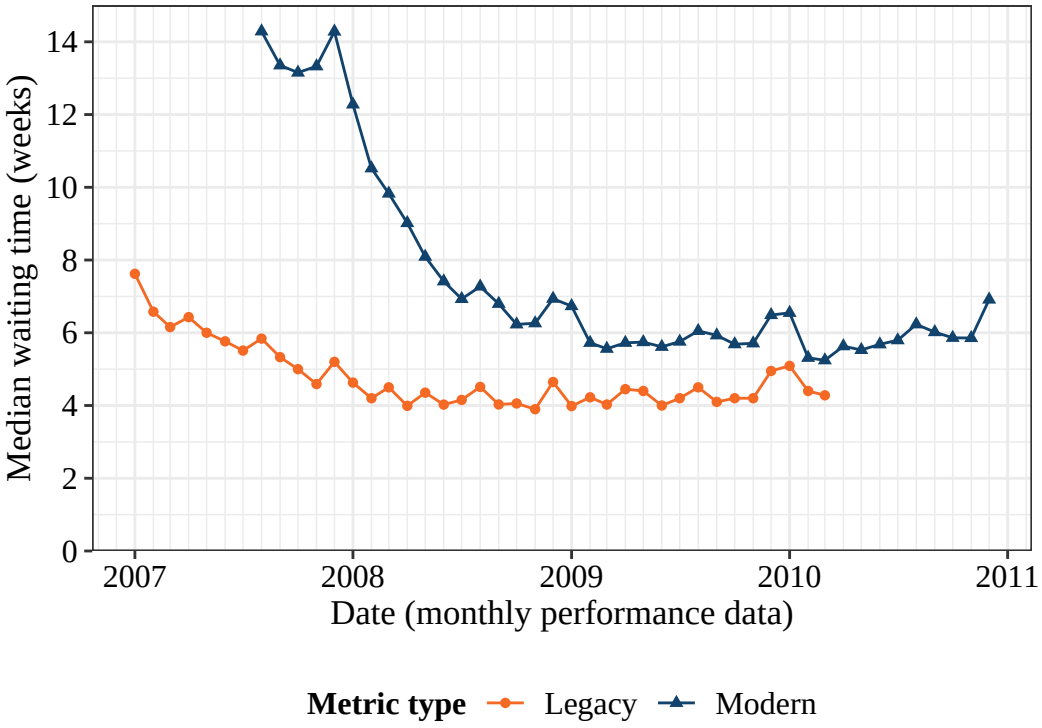

Figure 3.2: A comparison of monthly waiting times, using the legacy and modern waiting time metrics between 2007 and 2011, when both metrics were in operation.

Because the metrics are somewhat comparable across that 18-month period, to give a broader view of the impact of *relative* performance over time across both periods, I will also combine the waiting times from both metrics into an index by dividing each metric by its value in 2009 (the latest complete year for both metrics, and the one with the closest similarity in the data). 1 will therefore represent the 2009 value, 2 would represent double that value, 0.5 half of it, etc. For 2010 onwards, the modern indicator will be used. To reiterate, I will perform all the analysis separately on the legacy and modern data, as well as on the indexed metric.

The advantage of creating the indexed metric is that it allows for a single estimate of the parameters under study, one which uses information from across both time periods. It accounts for the gap between the two metrics as at 2009, while creating a single unifying scale of mea-

surement across both periods. However, it has some weaknesses: firstly, it assumes that the gap between legacy and modern metrics seen in 2009 would have remained stable after 2010. Secondly, it means that results are not reported in the normal "weeks" scale but as multiples of the 2009 value (4.1 or 5.8 weeks, depending on metric), which is less intuitive.

As such, care should be taken in drawing conclusions from the indexed variable that are not supported by both the legacy and modern data. This proves particularly challenging when considering the highest levels of waiting times, since median waits above fifteen weeks were only present in the Thatcher/Major years.

Figure 3.3 shows the indexed annual waiting time metric for the whole of the period under study. It suggests that while performance steadily worsened from 2012 onwards, this has only amounted to a 50% increase from 2009 waiting times, which is relatively much lower than the waiting times that were common in the 1990s and most of the 2000s, which were were two, three, or four times higher than in 2009. This suggests that in the era of the modern Referral to Treatment metric, there is a lot less change in waiting times between years.

Regardless of whether we use only the legacy metric, or add in more recent data using the indexed metric, it is clear that there is considerable variation in waiting times over the longer term. Taking only the legacy metric, median waiting times were over 20 weeks in 1988-1990, but twenty years later they had fallen to a quarter of that in length. The indexed waiting time metric also shows waiting times in the late 1980s being five times as long as those in the late 2000s. We can therefore expect there to be significant changes in public satisfaction with the NHS during this time period.

3.3.2 How performance varies by geography

We can also observe variation between Clinical Commissioning Group (CCG) areas in England to measure geographical differences in performance. This is the most local NHS geography with set boundaries for which performance metrics were reported during the period under study; in mid-2018, there were 191 CCGs, each with a population of around 290,000 (Office

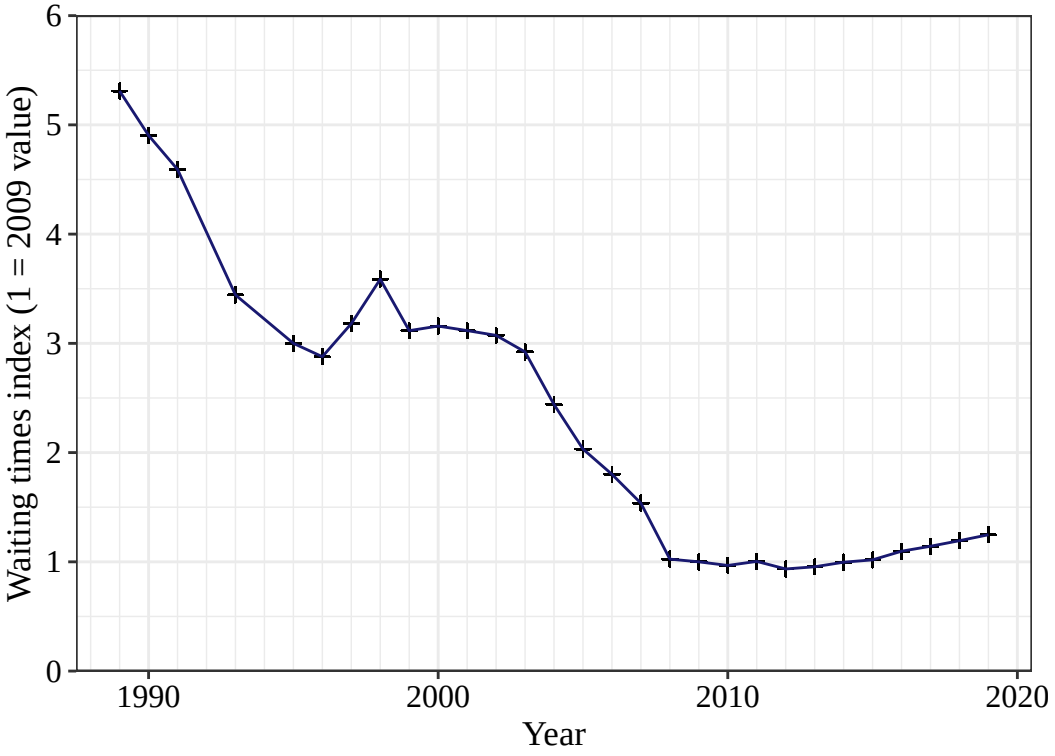

Figure 3.3: Changes in the indexed annual waiting time metric, 1989 to 2019. Combining the legacy and modern metrics shows how waiting times fell considerably in the 1990s and 2000s, with relatively little change since.

for National Statistics 2021). The metric represents the median waiting time from referral to treatment for people registered with GP practices in each CCG, regardless of which hospital provider is giving the care. It is therefore a good proxy for the general performance of hospital services experienced by people living in a given area.

To visualise the variation in this metric, Figure 3.4 shows the distribution of median waiting times across CCGs in each month of 2017, 2018 and 2019. 2016 is excluded as the earliest survey data in the study is from November 2016. Each vertical boxplot represents a month of data.[1]

As Figure 3.4 demonstrates, although there is a clear seasonal pattern, there is also considerable variation between CCGs. This indicates that, while in the longer term performance has changed dramatically from year to year, there is also a meaningful amount of variation between areas in England within the same short time period. As an example, in January of 2018 the shortest CCG-level median waiting time was 4.7 weeks for NHS City and Hackney CCG, while the longest was more than twice as long at 10.4 weeks for NHS North Lincolnshire CCG.

Compare the overlap between months in Figure 3.4 with Figure 3.5, where each CCG has its own boxplot representing the distribution of monthly median waiting times for that CCG within 2018. The variation in waiting times between CCG areas is clearly greater than the seasonal variation within a CCG in this short window of relatively consistent waiting time performance.

This shows that while there are very large differences in national waiting times in the longer term, there is also enough variation in performance between different parts of England that we might expect to also see voters recognising these differences - which would result in different levels of satisfaction with the NHS in different parts of the country.

[1]The box shows the central two quartiles and median. As the *ggplot2* documentation explains, "[t]he upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge." (Wickham 2016)

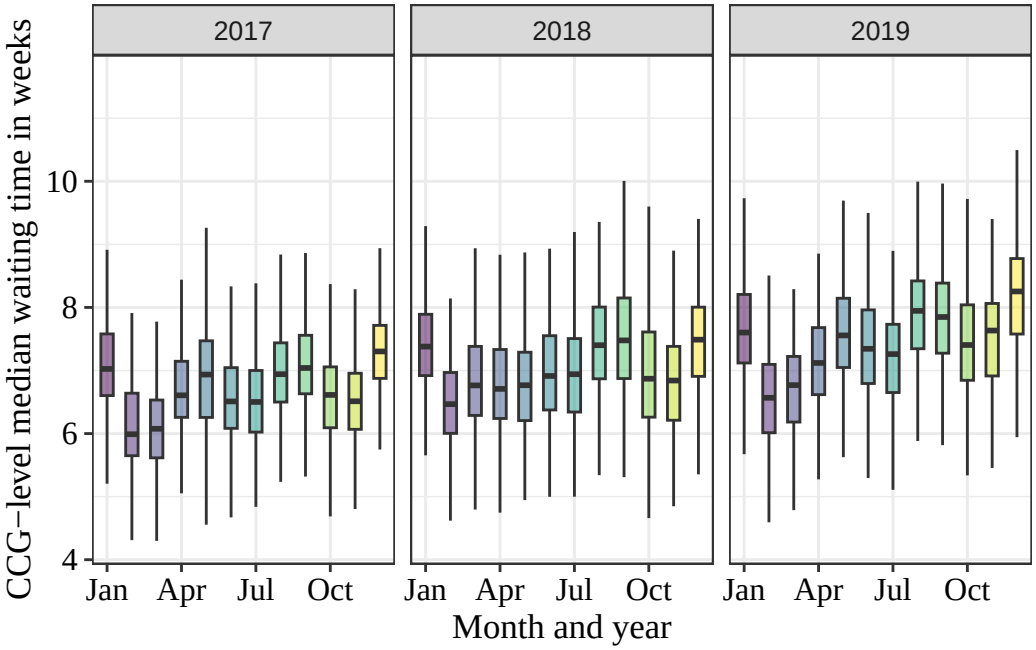

Figure 3.4: How median CCG-level waiting times differ in each month and year, 2017-2019. Each boxplot shows the variation between CCGs, while the difference between boxplots shows the seasonality inherent in the data.

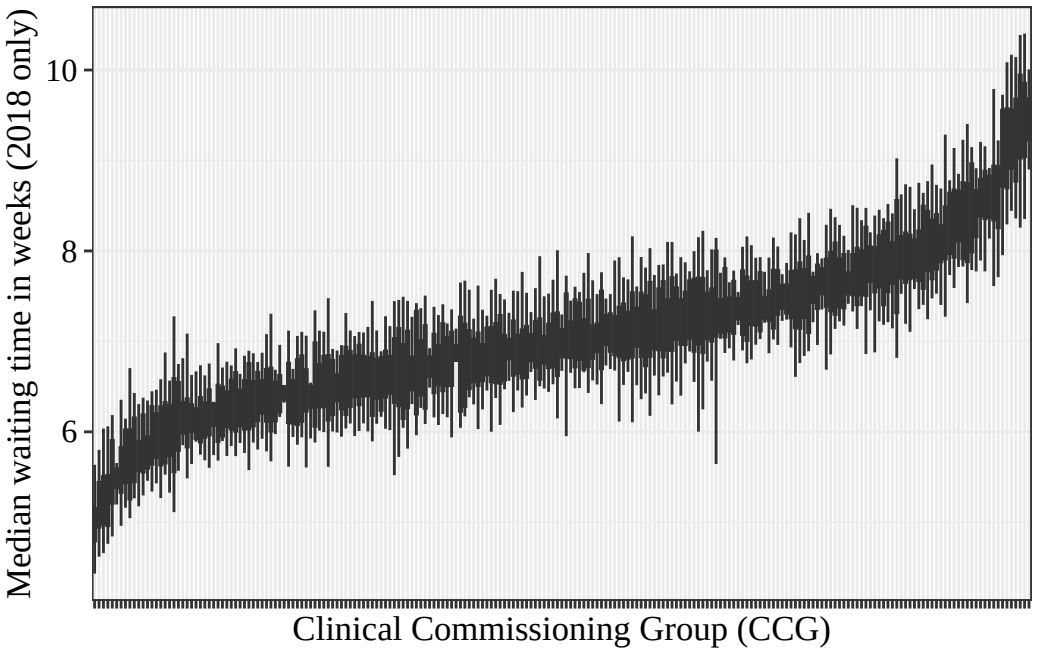

Figure 3.5: Boxplots of monthly median waiting times in 2018. Each boxplot shows how monthly waiting times vary within a Clinical Commissioning Group in 2018, differences between boxplots shows geographical variaion.

3.4 Summary

In this thesis I will use the English NHS as a case study, to determine whether this popular, high-profile public service was sufficiently salient between the early 1980s and late 2010s that voters were able to notice changes in performance, update their evaluations, attribute credit or blame, and change their voting behaviour in response. If not, this poses a challenge to the idea that voters can induce accountability in governments for their provision of public services, and adds to the literature by providing a detailed case study. I will also examine the potential ways in which voters could obtain the information needed to update their views as performance changes, which therefore increases our understanding of how those particular mechanisms, such as direct service contact, affect public opinion about services. In the following chapter I will outline a theoretical model that shows how the findings in the social-psychological and rational choice literatures can be applied to the specific case of the NHS, and what they imply for the process of opinion choice among partisan voters.

4 A theory of moderated learning about public service performance

4.1 Introduction

In the literature review in chapter 2, we saw how social-psychological and rational choice-based theories of voting behaviour are based on different assumptions about how people form opinions, and that these theories predict different electoral and attitudinal responses to external information. In this chapter I draw on the existing literature and apply it to voter evaluations of public services, specifically healthcare. For several reasons discussed later in this chapter, public healthcare provision is distinct from other policy domains typically used as case studies for testing theories of voting behaviour (e.g. macroeconomic conditions), and so here I aim to describe a theoretical framework that is adapted from elements of both approaches, to improve our collective understanding of how voters evaluate public services. While health provision is distinct, most of this framework could also apply to other public services that people interact with periodically, such as social care, welfare, and transport. This means that the theoretical and empirical aspects of this thesis could give insights about public services in general.

The chapter begins by briefly recapping the core elements of the social-psychological and rational choice approaches, and their implications for the study of satisfaction with healthcare services; I will not set out hypotheses based on these theories, but they will inform the theoretical framework and its primary hypotheses.

I will then lay out what the moderated learning theory assumes in terms of voter behaviour:

their goals and motivations, the constraints on their behaviour, a model of how they process information, and some factors that affect their likelihood to accept new information about how well services are being managed ('performance signals'). I will then outline the assumptions in the model about the sources of performance signals: personal experience, changes in government, and indirect information. As key expectations emerge, these will be signposted as numbered 'primary' hypotheses. The chapter ends with explanations of how changes in the information environment affect voters' evaluations of health services, and how those evaluations may feed into vote choice and political behaviour, and primary hypotheses relating to this are spelled out.

4.2 Implications of the main voting behaviour theories for public healthcare provision

4.2.1 Social-psychological summary

As described in the literature review in chapter 2, social-psychological approaches to voting behaviour argue that group identity, particularly one's self-identification with a particular political party, are the principal drivers of vote choice and even affect beliefs about objective political truths. Party identity is usually formed at a young age, is stable over time, and represents not so much an ideological attachment but an affective one: a feeling of belonging to a certain group or type of people. Partisan voters may elect another party's candidate for many reasons, but they usually want to 'come home' to their natural affiliation.

Partisanship affects what people believe about government policy and performance (Bartels 2002; Bullock, Gerber, Hill, *et al.* 2013) .The social-psychological theory of voting predicts that voters with an established party identity will tend to have an optimistic view of government performance when their preferred party is in power, and be more critical when their opponents are in power: that is, partisanship drives performance evaluations, not vice versa (Evans & Chzhen 2016b; Green, Palmquist & Schickler 2002).

These findings - and the basic underlying idea of partisanship as social identity - have been further elaborated in modern political psychology, which also examines how and why this perceptual screen is effective. There are two main competing explanations for how partisanship affects information processing: the first is that people use partisanship as a heuristic, avoiding costly information gathering by using partisanship as a cognitive short cut. The second is that people aim to keep their beliefs aligned with their social identity, which is deeply entangled with their partisan identity, and so protect themselves by discounting dissonant information (Bavel & Pereira 2018). While there is evidence that partisanship can act as a cue in low-information elections (Schaffner & Streb 2002), there is ample experimental evidence that people actively resist information that conflicts with their views, even to the extent of making data analysis more difficult when the data is political (Kahan 2013, 2015; Nir 2011; Taber, Cann & Kucsova 2009; Strickland, Taber & Lodge 2011). Partisans therefore tend to embed themselves in information environments that insulate them from negative signals about their chosen party, and use motivated reasoning to dismiss or downplay any negative signals they do receive.

4.2.1.1 Implications for perceptions of health services

According to the social-psychological perspective as summarised above, we would expect that people who hold partisan views will respond differently to news or information about healthcare policy.

Those who identify with governing parties (government partisans) are less likely to be exposed to negative information about health provision, and when they are, they are more likely to ignore, rationalize, or discount it – leading them to be, on average, more satisfied than others with how the health service is being managed. Conversely, those who identify with opposition parties (opposition partisans) are more likely to consume media that presents negative information about government health policy and accept it unchallenged – leading them to be less satisfied on average. Partisanship, rather than performance, will be the defining factor influencing satisfaction with the health service.

Elements of this account are incorporated into my theoretical framework in this chapter. In a later chapter, I will examine the implications of the social-psychological perspective in greater detail, and use survey data to test whether this social-psychological theory of voting is applicable to the context of public services, and more particularly the English NHS.

4.2.2 Rational choice summary

Whereas the social-psychological approach places partisanship at the centre of voting behaviour, models based on the rational choice theory describe voters as rational maximisers of utility, who use political choices try to obtain the best possible outcomes for themselves.

One potential strategy for making that choice is retrospective voting, where voters evaluate incumbents' past performance and use elections to remove politicians who under-perform according to their preferences or incentivise good behaviour. This relies on a chain of events wherein the voters perceive performance by picking up on 'performance signals', choose whether to assign responsibility for performance to the government, and alter their voting choices accordingly (Anderson 2007). While there is a considerable literature testing theories of rational retrospective voting in different contexts, the main features of the theory have long been established (Downs 1957; Fiorina 1981; Key 2013).

Rather than being an existing social affinity, Fiorina (1981) theorised that partisanship was the accumulation of repeated positive reinforcements from a party, a "running tally" of which party had been consistently more successful at delivering utility for the individual.

In response to criticisms that rational choice places an unrealistic intellectual burden on the electorate, retrospective voting has been reinterpreted as a cognitive short-cut or heuristic, that allows low-information voters to decide their vote based on recent government performance (Healy & Malhotra 2013; Rahn 1993; Schaffner & Streb 2002). Recent work in this vein, referred to as the 'valence politics' literature, accepts that voters have cognitive biases and limitations, but argues that voters' willingness to vote for a party depends on their perceived competence (Green & Jennings 2017a, 2017b).

4.2.2.1 Implications for public healthcare

From a rational choice perspective, voters seek to elect politicians who deliver positive outcomes - in this case, good healthcare performance. If the four-step process of retrospective voting (Anderson 2007) is functioning in the domain of government-provided health services, we would expect the following to occur:

- Firstly, voters receive and evaluate information about performance (waiting times, GP appointment access, service availability, ambulance waits, etc.) and about the decisions government is making - such as announcements about NHS funding.
- Secondly, voters make a link between the performance information and political choices; if the information is positive, they give the government credit for managing the NHS well, but if it is negative they blame the government for managing it poorly.
- Thirdly, voters remember these judgements when deciding who to vote for, rewarding incumbents who they credit for good health service performance and punishing those they blame for poor performance.
- Fourthly, politicians respond to these incentives by making decisions that aim to deliver good NHS performance.

There may, however, be barriers to accurate evaluation of services. In order to make a retrospective judgement, voters need to be able to pick up the performance signals, which assumes availability of performance information, whether through direct contact, news media, word of mouth or another medium. When they do encounter a performance signal, voters may exhibit partisan bias: if a party has previously delivered positive results, voters may use that as a heuristic for current performance. However, repeated consistent signals will 'update' the voter's beliefs.

Looking from a rational choice perspective, while there may be some partisan bias due to selective exposure, the primary driver behind changes in evaluations of services would be objective differences in performance and therefore voters' expected utility from health services. In other

words, we would expect that satisfaction will rise among all voters when performance of health services (broadly conceived) improves, and satisfaction will fall among all voters when performance of health services worsens.

In summary, a rational choice approach to theorizing changes in satisfaction with the health service in England would imply that the public would respond to information suggesting positive or negative changes in performance by updating their opinions of the service.

4.3 Moderated learning

In this section I propose a broad theoretical model for understanding changes in satisfaction with public services, building on and synthesizing the social-psychological and rational choice approaches. The theoretical framework is compatible with a range of competing explanations, to the extent that they allow for voters to learn from new information while balancing it with their priors. I begin by outlining the assumptions in the model about voters' goals and motivations, their constraints, and how they process information. I then describe in more detail the effects of the external environment on this process. Finally, I summarize the implications of the model for voters' opinions and vote choice. Throughout the section I will use illustrative graphics to demonstrate the implications of key elements of the theory.

4.3.1 Scope

In this thesis, I focus on the first step in the four-step retrospective voting process. Voters' evaluations are a vital first condition that are not often examined in detail, especially outside of the economic domain (with a few notable exceptions as cited in chapter 2). While examining vote choice is clearly important because elections determine governance, looking at vote choice on its own could obscure whether retrospective voting failed because voters did not recognise performance signals, did not attribute responsibility to the government, or whether other influences (such as the electoral system, candidate attributes, or leader image) crowded out the influence of performance evaluations. In order to understand how the four-step accountability

chain is working in practice, detailed study of each step in isolation is required. This thesis therefore provides a detailed examination of the first of these necessary steps: voters being able to learn about government performance is the minimum requirement for retrospective voting to take place.

4.3.2 Key assumptions about voters

4.3.2.1 Goals and motivations

Despite the differences between the social-psychological and rational choice theories of voting behaviour, the general premise that voters have both accuracy and directional goals, described from a psychological perspective by Taber, Lodge & Glathar (2001) is compatible with both sets of theories. 'Accuracy goals' are when voters aim to discover and hold accurate beliefs and minimize false beliefs, whereas 'directional goals' are when voters aim to maintain beliefs that are congruent with a predetermined conclusion (Taber, Lodge & Glathar 2001). From a rational choice perspective, partisan bias could be described as a heuristic to reduce cognitive effort by relying on existing information (Dancey & Sheagley 2013; Schaffner & Streb 2002).

I therefore assume that there is some effort on the part of voters to ascertain what is true, how well governments are performing, who is responsible for changes, and so on (accuracy goals). I also assume that there is also a tendency, whether seen as partisan bias or rational use of prior information, to allow partisanship to influence opinion (directional goals). This may be rooted in a desire to align one's views with their social group, or to follow elite cues; whatever the intrinsic motivation, it has the effect of reinforcing partisan thinking and encourages people to be more sceptical of information that conflicts with their existing views.

In the previous chapter, I showed that there is considerable agreement among the public in favour of the publicly provided NHS model, free at the point of delivery. In this individual-level theoretical model, I therefore treat NHS performance as a valence issue, where parties compete on the basis of competence, rather than on the basis of spatial ideological positions. We can think of a voter's opinions as existing along a one-dimensional continuum, with more negative

evaluations further to one side and more positive evaluations further to the other; while this is necessarily a simplification, it is useful for our purposes here. The voter's accuracy objective compels them to try to position their opinion as close as possible to an unobserved "true" value of the NHS' performance. Because this abstract parameter is not directly observable, they must infer its position by assessing the evidence they have available to them: the performance signals they receive through their own experiences or external sources, and their prior internal narrative[1]. Their directional objective acts as a brake on movement, allowing the voter to discount, ignore or diminish evidence that conflicts with their priors. This mechanism will be explored further later in this chapter.

In order to better illustrate the process of opinion change envisioned in this theory, I present a set of idealised diagrams that depict this one-dimensional continuum. In the first diagram (Figure 4.1), more negative opinions are further to the left of the continuum, and more positive opinions are further to the right. The X represents the voter's current evaluation of the service. Further elements will be added to later iterations of the diagram to illustrate the parts of the process under discussion in the section.

4.3.2.2 Constraints on voters' capacity to evaluate

Voters have a limited capacity for gathering information because of the costs involved. From a strict 'rational choice' perspective, it makes sense for voters not to expend energy evaluating candidates for office, since the probability they will cast a decisive vote is minuscule (Downs 1957). Similarly, keeping up with news about health policy, checking waiting times statistics at local hospitals, evaluating GP services and so on require time and energy that could be otherwise spent on other activities. Evidence from Taiwan, where patients can opt to see any doctor of their choice, shows that only half of patients actually make comparisons between doctors when seeking care (Cheng & Song 2004).

[1] While from an external perspective we make a distinction between their 'partisan priors' and new information, from the voter's perspective these are simply different types of fact being drawn on to make sense of government performance.

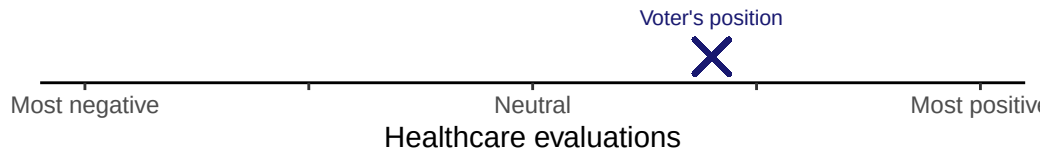

Figure 4.1: A voter's evaluations of the NHS represented as an X on a left-right continuum, with worse evaluations on the left and better evaluations on the right

In addition, because of the links between social and political identity, we can expect selective exposure to different sets of information about public services. This could emerge in several ways. Firstly, people's partisanship is often influenced by their family and social milieu (Campbell 1960). As such, people who are inclined to support the government of the day and therefore be more satisfied with the NHS are more likely to socialise with others of similar views, and hear experiences relayed through that lens (again, negative experiences could be justified away or ignored) and vice versa. Therefore, voters' ability to perceive information about public services is also limited by what signals are received in the first place.

Voters are also limited in their ability to understand information about government performance because they may not have the relevant information to contextualise the new pieces of evidence, making that information processing task more costly. For example, someone with no previous experience of inpatient admissions may not have a frame of reference for what 'good' hospital care is like and will therefore be unable to make a judgement. Past experience shows that the English public tend to prioritise access and convenience over clinical quality; this is not

surprising given the difficulty in measuring 'quality of care' (Hanefeld, Powell-Jackson & Balabanova 2017; Lagarde, Erens & Mays 2015; Tan, Erens, Wright, *et al.* 2015). This suggests that voters prefer to use simpler metrics when making decisions about care; and that even where information is received about public services, people may not allocate the time or cognitive energy necessary to consider the information and integrate it into their opinions. This could be seen as an example of 'satisficing' rather than 'maximising' utility, especially if voters believe that the differences in quality between providers are not meaningful or that such a comparison is not possible.

4.3.3 How voters learn about public service performance

In this theoretical model, voters' information processing is described as a "moderated learning" process. As described in the literature review, we know that there is evidence for the idea that voters react to government performance in general, but that this process of learning from new evidence is limited by cognitive restrictions, partisan biases, motivated reasoning and partisan sorting within the information environment. These limits therefore "moderate" voters' learning process over time.

Theories of Bayesian learning have been previously applied to political science (Grynaviski 2006; Hill 2017). In the Bayesian learning model, posterior belief is a weighted combination of the probability of the voter's prior belief and the 'meaning of new evidence'. Bayesian learning is therefore a type of moderated learning model, because voters' posterior (updated) evaluations of services will be closer to their prior beliefs than the naive meaning of new evidence; in other words, while voters can learn from new information, they continue to take information their priors into account. However, whereas Bayesian learning posits that voters learn at a specific rate according to Bayes rule, my moderated learning model is more general and does not require any particular rate of learning. This flexibility allows it to accommodate not only Bayesian learning perspectives, but a range of alternative interpretations of how voters update their opinions.

As in the retrospective voting model, I posit that voters perceive and react to information about

government performance. However, this does not mean accepting a view of voters as idealized, rational decision makers, but rather taking this ideal as a starting point and adapting it to take account of the reality of cognitive and environmental barriers to learning, including partisan bias.

In this model, voters therefore gather or encounter information that helps them evaluate services, allowing them to satisfy their desire for accuracy; they process this in ways that allow opinion change, but in line with their directional goals, they combine the new information with their existing beliefs. In this way, we can theorize that government performance does matter and can provoke retrospective voting (or, at the least, alter public opinions) under certain circumstances – bounded by the cognitive limitations and biases of the public, as well as an information environment that is conducive to selective exposure to political information.

The literature has shown that some partisans go out of their way to counter-argue against conflicting factual information rather than revise their views (Schaffner & Roche 2017). However, experimental evidence suggests that in general, voters learn as 'cautious Bayesians', updating their views 73% as much as a perfect Bayesian would; and that partisans do accept information contrary to their preferences, albeit at a lower rate (Hill 2017). This finding supports the theory to be tested here, which argues that partisanship can introduce bias into public opinion and make it harder for voters to update their opinions, but that it is not impossible for opinions to change when new information signals contradict a voter's prior.

I assume in this model that as active processors of information, voters make judgements about which performance signals to believe, discard, or partially accept. I model information as having two properties: its value (how positive or negative its implications are), and its credibility (how much weight the voter places on the information, based on their assessment of the probability it is accurate). Both of these are subjective and represent a judgement on the part of the voter. By assigning lower credibility to information that is further away from their prior opinion, voters are able to balance their directional goals and accuracy goals.

4.3.3.1 Priors

In this model, priors represent the existing views of the voter, as shaped by their partisanship, personal experiences and other information they deem relevant. Priors can consist of the accumulated impact of long-forgotten experiences and feelings, allowing both to influence decision-making. I assume that for most voters, partisanship is a major influence for the reasons described above (e.g. constraints on information gathering).

We can think of a voter's prior public service evaluation as having a probability density: a voter may, for example, rate a public service's performance as 7 out of 10, but if evidence emerged to suggest that they should rate it 6 out of 10, this would be more believable than information suggesting it should be rated 1 out of 10.

Returning to the toy example depicted in an earlier diagram, in Figure 4.2, a voter's prior opinion – how they would rate their satisfaction with the NHS given the opportunity - is again represented by a cross. I also add a curve representing the probability density of the voter's prior. This represents the area of the continuum where the voter believes an accurate assessment of the NHS is most likely to fall. Areas where the curve is higher are more compatible with the voter's prior opinion, and therefore the voter is more likely to assign a higher weight to information suggesting these values.

4.3.3.2 New information

In this model, new relevant information about government performance has the potential to alter the voter's opinion. Rather than take each new piece of information as authoritative, the information is processed and accounted for in the context of its credibility (how much weight it will be assigned when combining information) and its compatibility with existing opinion (how close it is to the voter's prior). These pieces of information are 'performance signals' that give clues about the true underlying performance of the public service. Voters understand that there will be some random variation in outcomes from a service with a given underlying quality - that a generally good service may produce an unrepresentative bad outcome from

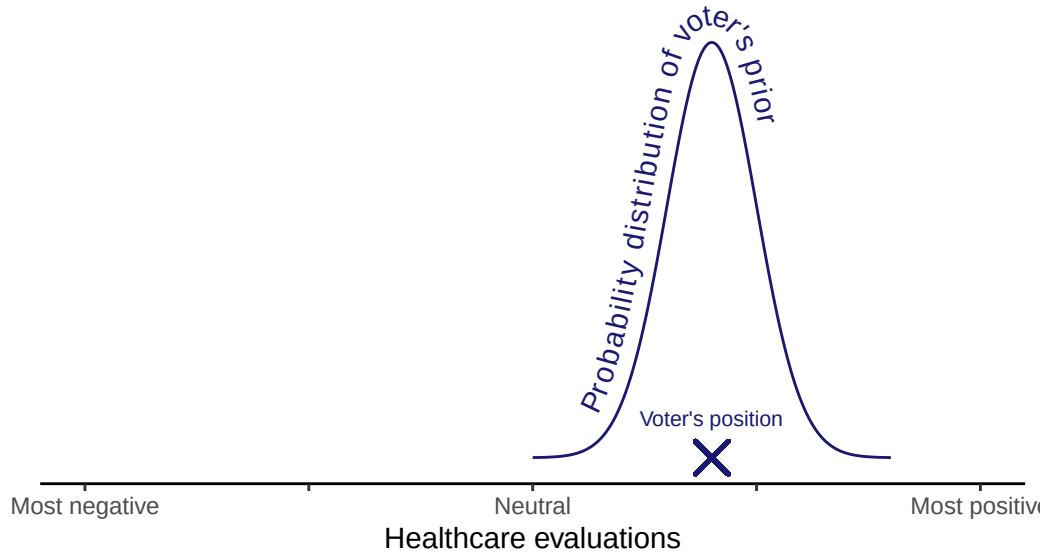

Figure 4.2: A voter's evaluations of the NHS represented as an X on a left-right continuum, with worse evaluations on the left and better evaluations on the right. A normal distribution curve, centred on the X representing the voter's evaluations, represents the probability density the voter applies to different possible evaluations on the continuum.

time to time, for example. Voters can therefore think of these performance signals as a sample drawn from a distribution around the true underlying performance of the service. A purely rational voter would infer the value of that hidden parameter by combining their priors with the new information. Any new information would move the distribution's mean and inform the shape of the posterior, regardless of the value position of that information. They therefore weight and combine information from their prior and different information signals to produce their evaluations, rather than discarding previous information any time a new piece of data arrives.

Having evaluated the information, the voter then can attribute responsibility (credit or blame) for the evaluation. I also assume that voters understand that there is to some extent shared responsibility for public service performance between different actors. For example: the government is responsible for overall funding and strategy, local managers for how their resource allocation is used within national guidance, and clinicians and other staff have a significant influence on patient experience. This may potentially insulate politicians from blame (or avoid them receiving credit) as a result of differences in health service performance; this is in line with welfare state literature suggesting that voters react less to negative service changes where responsibility is distributed (Pierson 1996).

Figure 4.3 takes the voter's prior from our toy example and adds several new items of information, or signals, with varying values but similar levels of variance. The mean value of the new information is equal to the mean value of the existing prior, so the voter's prior opinion does not move, either in a situation of perfect voter rationality or when assuming moderated learning.

However, drawing on the literature, I assume that voters are resistant to updating priors based on new information that is incompatible with their existing opinion. While they can change opinions due to new evidence, this only happens once a considerable weight of evidence has emerged. This is because the voter may assign low credibility to information that is far away from their existing opinion. This information will therefore receive a lower weight when the voter revises their evaluations by combining their priors with new information.

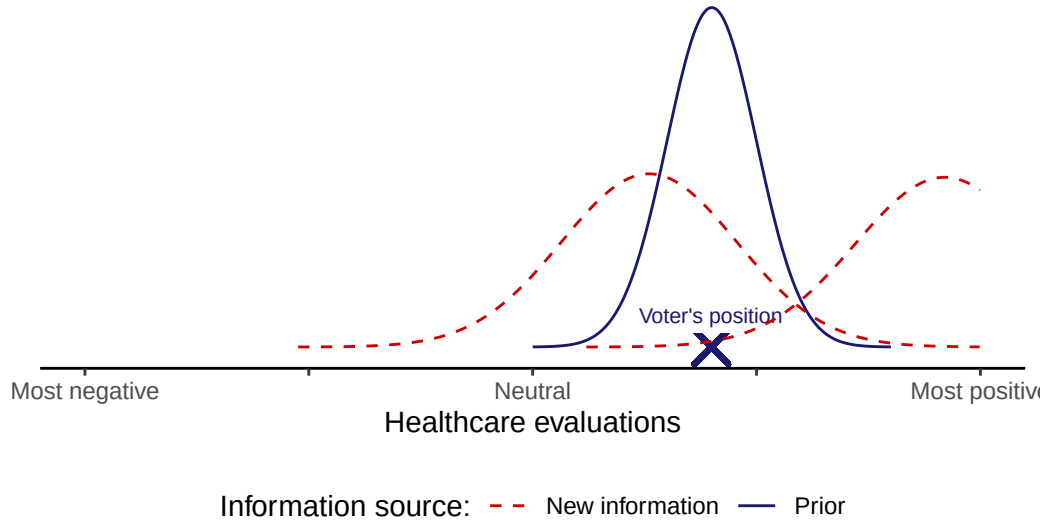

Figure 4.3: An example where the voter's prior opinion does not change when conflicting information emerges, because taken together the mean value of the new information is equal to the mean value of the voter's prior

4.3.3.3 Factors affecting perceived strength of performance signals

As posited above, voters are slow to update their priors because they expect an innate level of noise or variation in performance signals, and therefore place little weight on any given signal - especially those far away from their prior. Because of this, a single positive or negative experience of the health service is unlikely to shift a person's opinion significantly in most cases, and therefore they will wait for sustained indicators of change to alter their opinions. A single incongruent piece of information may be easily dismissed with a justification such as "the NHS is doing well, but the clinic I attended was unusually busy today" or "the NHS is mismanaged and underfunded nationally, but my GP surgery is the exception" (for example, see Ziebland, Evans & Toynbee (2011)).

However, opinion change is possible. This theory assumes that individuals update their evaluations of services when a consistent flow of credible evidence overwhelms the voter's prior opinion. In the toy example, because more negative information emerged, the voter has combined their priors (represented by the blue curve) with new information (represented by the dashed red curve). Although the voter assigns less credibility to information that conflicts with their priors, it has been combined with the priors to create a new posterior opinion and probability distribution. The new data has now been integrated into the voter's "posterior" evaluation of the service. This is represented by a second curve in Figure 4.4. This posterior will serve as the voter's prior when the next piece of performance information is received.

As newer information is given more weight and older information is forgotten or given less weight, voters can entirely change their opinions, given enough evidence. A change in opinion may happen quickly, if the conflicting information is very credible. The mechanism is the same regardless - contradictory evidence overwhelms the voter's prior - but it takes place over a much shorter period of time.

In summary, because people may discount information that contradicts their existing beliefs, a consistent flow of contradictory evidence is required to cause significant shifts in opinion. This can emerge either through an accumulation of new information over time, or through a

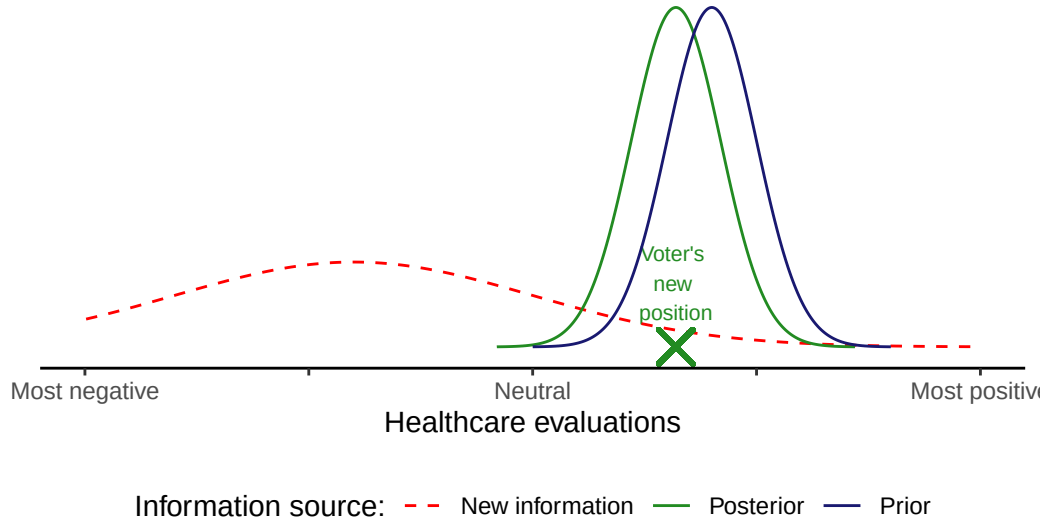

Figure 4.4: Diagram showing how the probability distribution of a voter's opinion has updated to accommodate new information

smaller number of high-credibility information signals.

4.3.4 Sources of performance signals

This section discusses several different types and sources of information ("performance signals") that may lead to changes in opinion about public healthcare, and posits several 'lead' hypotheses about the effect of each of these sources of information on people's evaluations of the health service that will be tested in this thesis. Additional hypotheses that flow logically from the primary hypothesis will be enumerated in the relevant empirical chapters.

4.3.4.1 Government partisanship

This theory assumes that people use their political partisanship to inform their views on public service performance. This means that all else being equal, Labour partisans will be more

satisfied with the NHS under a Labour government than a Conservative one, and vice versa for Conservative partisans. This is because partisans are reluctant to attribute success to their opponents or failure to their own party. In principle, non-partisans' levels of satisfaction should be somewhere in between the two regardless of which party is in power (again, all else being equal).

When a party leaves government, the assumption among its partisan supporters that "the government is doing well" dissipates, because it was based on the logic that 'their' trusted party was in power. Conversely, opposition partisans whose prior assumption was that the government was mismanaging public services because their favoured party was not in power would have that assumption reversed and therefore become satisfied with government performance.

The above assumptions are based on the logic of the moderated learning theory outlined above. However, when considering the effect of government partisanship on satisfaction from a valence perspective, work done by Green & Jennings (2017b) on macro-competence and issue ownership becomes relevant.

In "The Politics of Competence" (2017), as mentioned in the literature review, competence is split into three concepts. Macro-competence or generalised competence describes an overarching view on the part of the public as to how well they consider the government is doing. The concept is primarily used in the book to discuss how 'honeymoon periods' of high popularity are almost universally followed by decline due to the 'costs of governing', but the correlation between voters' views on different issues poses a potential challenge to retrospective voting on public services. If voters' views on other issues (or of generalised competence) influence their satisfaction with the NHS, and the government's objective performance on those other issues is different from their performance on health, this suggests a reduced ability on the part of voters to recognise and judge NHS performance. Green & Jennings (2017b) use data from the 2015 UK General Election BES Survey to show that NHS performance had a non-significant negative effect on likelihood to vote Conservative after controlling for performance of other issues.

I would argue that while this is a potential issue for issue-based retrospective voting, as we

saw in chapter 3, satisfaction with the NHS was at its highest in 2009, the final year of the Blair/Brown Labour government. While this suggests that good health service performance does not determine electoral outcomes, it also suggests that voters are able to distinguish between generalised government competence and performance on the NHS in particular (perhaps because of the issue's high salience relative to other public services).

Issue ownership is argued to be long-term perceptions of which issues are best for which parties. According to this measure of issue ownership, health was Labour's top issue throughout 1983 - 2015 period, with the exception of 1997 - 2001 in which it was their fourth highest issue (see Table 3.5 in Green & Jennings (2017b)). This means that generally voters thought health was a better issue for the Labour party than other issues (e.g. the economy, defence, education); by contrast, during the same period Health was never in the top four issues for the Conservative party (see Table 3.11 in Green & Jennings (2017b)). Changes in issue ownership for health could potentially result from, or contribute to, changes in voters' perceptions of health service performance. However, as there appear to be no changes in issue ownership between the two main parties during the period under study, this does not need to feature specifically in my empirical analysis.

Therefore, with the possible exception of election years, when voters' opinions may still be adapting to the change of government, I expect the following hypotheses to hold:

> PRIMARY HYPOTHESIS I: GOVERNMENT PARTISANS WILL EVALUATE THE NHS MORE HIGHLY THAN OPPOSITION PARTISANS.
>
> PRIMARY HYPOTHESIS II: NON-PARTISANS' EVALUATIONS OF THE NHS WILL BE LOWER THAN THOSE OF GOVERNMENT PARTISANS, BUT HIGHER THAN THOSE OF OPPOSITION PARTISANS.

4.3.4.2 Indirect information about performance and provision

As well as their partisan priors, this theory expects voters to receive 'indirect' information about health service performance and provision (that is, information gained other than through

personal experience of services) in their day-to-day lives. This could include through news consumption, conversations with friends, commentary on social media, official statistics, or even direct observation of facts that suggest pressure on services or other negative aspects of performance (e.g., ambulances queueing, services closing).

In the literature review in chapter 3, we saw that there are separate strands of political science literature dealing with 'performance' and 'welfare state retrenchment'. The evidence suggests that public opinion may respond to changes in the objective quality of what is being provided, as well as to the withdrawal of services. We would expect voters to be less satisfied, on average, in situations where the performance of health services is worse, and more satisfied on average in situations where NHS performance is better, as voters update their views to align more closely with the information they receive about service performance.

> PRIMARY HYPOTHESIS III: VOTERS' EVALUATIONS OF THE NHS WILL BE BETTER WHEN SERVICE PERFORMANCE METRICS ARE BETTER

4.3.4.3 Personal experience

Voters' own experiences of services could be a highly credible source of signals about NHS performance. My assumption is that direct personal experience of health services is the most effective information source for overcoming partisan priors and updating evaluations of performance, because it is not being relayed through a third party such as a newspaper, and therefore voters will perceive it as more objective.

However, there is considerable variation in exposure to performance signals between individuals, since different groups have very different levels of service utilisation. One clear example of this is the difference in health service utilisation by age. For example, Figure 4.5 shows the number of hospital outpatient attendances per person in England by age group in 2019-20. Population data is from ONS mid-2019 estimates (Organisation for National Statistics 2020), while outpatient statistics are from Hospital Outpatient Activity 2019-20 (NHS Digital 2020).

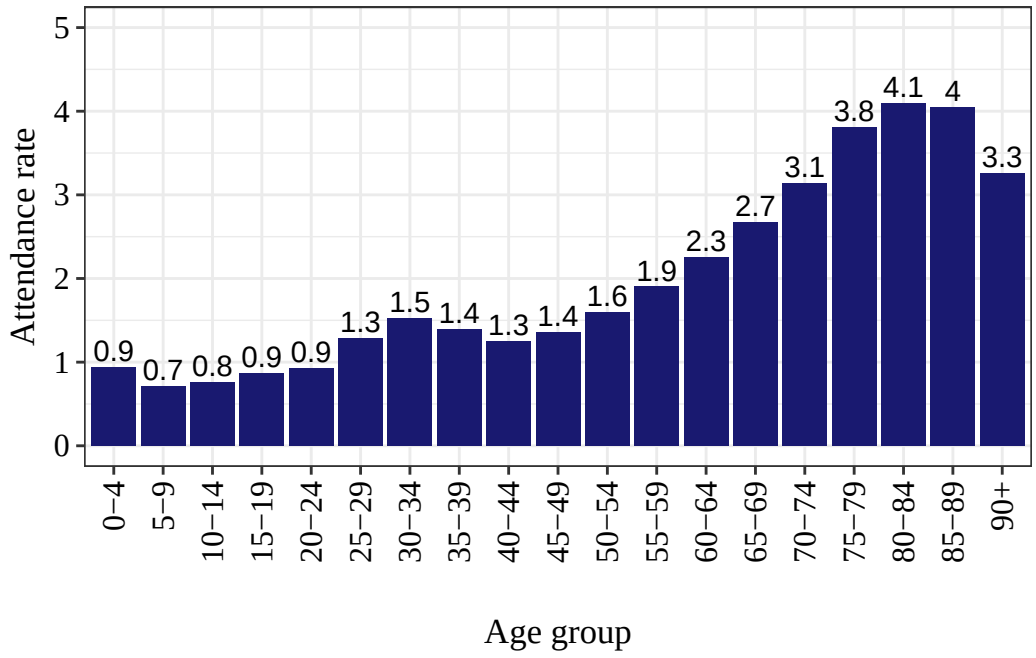

Figure 4.5: Annual outpatient attendance rate per person, by age group, 2019-20. Data relates to England. Population figures are from mid-year 2019/20 ONS estimates, outpatient attendances are from NHS Hospital Episode Statistics.

The group with the highest utilisation is 80-84 year olds, with a rate of 4.1 outpatient attendances per person per year, more than quadruple that of the 20-24 year old group. We should also expect service utilisation to vary by other factors, including gender and chronic health status (Economic Cooperation & Development 2019). This suggests that some voters will have little to no recent contact with NHS services, whereas others will have had very recent contact.

Another feature of personal contact with the NHS that is theoretically relevant is the possibility that information gained through personal experience may be processed differently to abstract or third-party information gained through the media or statistics. If personal experience is processed using a more emotional, personal framing rather than a political or ideological framing, it may not influence views about how services are being managed by the government. Therefore, while personal experience may be seen as unbiased and highly credible, it is possible that this type of performance signal is paradoxically less influential. This possibility is supported by polling showing that patients are more satisfied with the NHS than the public (Edwards 2006) and interviews with patients suggesting this is because they think they are "lucky" to have received good care from an otherwise poor system (Ziebland, Evans & Toynbee 2011). In chapter 7, individual-level data from two surveys will be used to test the effect of recent NHS contact on service evaluations, which will either confirm or challenge these findings.

According to this theory, additional information can replace partisan-informed priors about service performance and cause voters to update their evaluations of services. Depending on the difference between these partisan priors and the performance levels implied by the additional information, this may cause evaluations among partisans to improve or worsen. This implies that NHS evaluations will be different among people with recent experience of NHS services.

PRIMARY HYPOTHESIS IV: RECENT SERVICE USE WILL AFFECT VOTERS' EVALUATIONS OF THE NHS

4.3.5 Expectations about how vote choice responds

The final phase of retrospective voting is for voters to either vote for the incumbent if they are satisfied with their performance, or to vote to remove them if they are dissatisfied. As the final phase, it is contingent on the preceding phases, which this theory predicts are themselves dependent on circumstances. This means that changes in vote choice are only possible where voters have been able to pick up on enough performance signals to change their evaluations of performance; if most voters' opinions about healthcare performance are usually a function of their political views, then the opportunities for changes in voting behaviour - and therefore accountability for politicians - are limited.

This thesis focuses on the 'evaluations' step of the retrospective voting process, and does not attempt to answer the broader question about how voters' evaluations of the NHS affects their vote choice. However, there are reasons to believe that the public might vote retrospectively on health performance: the NHS is highly salient. It is consistently cited as one of the most important issues facing the country (YouGov plc 2020). As a major employer, and provider of life-or-death services to the population, I would expect that there is at least a section of the public who are attuned to performance signals about the NHS and who would factor its performance into their voting intention.

On the other hand, whereas in satisfaction surveys individuals can express their views on the NHS in isolation, when voting it will be one factor of many affecting the decision on how to vote (Clarke, Sanders, Stewart, *et al.* 2011; Dassonneville 2016). Some voters may change their opinions of the NHS, but not change their vote choice; while the theory predicts a change in voting behaviour, changes in voting intention are likely to be less prevalent.

Moving away from theoretical reasons why retrospective voting may or may not take place, there are some difficulties in analysing this phenomenon empirically. Testing hypotheses about changes in voting intention is more challenging than hypotheses about satisfaction: although changes in voting behaviour may emerge through long-term performance shifts, it would be difficult to identify differences in performance levels as the cause, since many other events

that could influence voting intention at a regional level would be occurring at the same time. This makes constructing a convincing counterfactual more difficult, though panel data could provide an avenue for this to be done.

In this thesis, however, I only study vote choice in the context of voter reactions to a specific program of service retrenchment. Because the closure (or planned closure) of a hospital or department is a discrete event, this can be identified as a specific 'treatment' and its effect on vote choice measured. In chapter 8 of this thesis, I will attempt to identify the impact of a specific planned programme of hospital retrenchment on both voters' evaluations of the NHS and their vote intention. Because changes to hospital services could be argued to be positive — such as to address clinical quality issues (The King's Fund 2017) — and they were proposed by local NHS managers, retrospective voting on this issue depends on voters not only being aware of the plans, but also deciding that they are negative, and blaming the government.

> HYPOTHESIS V: HEALTH SERVICE RETRENCHMENT WILL RESULT IN RETROSPECTIVE VOTING AGAINST INCUMBENT POLITICIANS

4.4 Summary

In this chapter I have drawn on insights from the social-psychological and rational choice approaches to develop a novel theory of moderated learning. This theory posits that voters are able to receive and process information about public service performance - also referred to as "performance signals" - make judgements about the signals, and react by altering their views and political behaviours; however, it argues that this learning process is moderated by a number of limitations and biases, notably partisanship.

The influence of partisan priors plays an important role in the theory. No position is taken on whether partisanship acts more like Converse's perceptual screen, which reduces voters' ability to accommodate information that conflicts with their partisan priors, perhaps through motivated reasoning on the part of voters; or whether through a voters combining their priors

with new information in a Bayesian way. In either case, it amounts to voters processing information in a way that balances both their partisan-influenced beliefs with the new information; in any case, it is difficult to distinguish the difference between these examples empirically (Little 2022). I therefore accept the suggestion from Taber, Lodge & Glathar (2001) that voters have directional goals when processing information, but also argue that they have the accuracy goals suggested by rational choice-based approaches: they seek to evaluate governments and respond to information about public service performance.

Another factor that increases the importance of partisan priors is the information environment. I argue that information signals about healthcare services are relatively scarce. As such, for most people most of the time their evaluations of public services will be heavily influenced by partisan priors, rather than recent information. Voters will use partisanship as a heuristic or cognitive short-cut, generally assuming that government services are better when their preferred party is in government than when they are in opposition.

New information can come via different routes. People will treat personal experience of service as being credible, since it is first-hand, which should result in recent service users having different views of services than others. Indirect information about performance may also affect people's views; however, partisan sorting of media sources and motivated reasoning are likely to dampen the effect of conflicting information about performance that is communicated in this way.

Finally, I also discussed the process by which opinion change could effect change in voting behaviour. Vote choice is determined by many different factors, so opinion change may emerge without inducing retrospective voting. In order to test whether retrospective voting could be operational in a public service context, I will examine a planned programme of hospital retrenchment to measure its effect on local public opinion and vote choice.

If this moderated learning framework is an accurate description of how voters evaluate and respond to public service provision, we will see that public evaluations of the NHS are influenced both by voters' partisan identities and by differences in service performance and provision. I have drawn out five primary hypotheses from this theory:

PRIMARY HYPOTHESIS I: GOVERNMENT PARTISANS WILL EVALUATE THE NHS MORE HIGHLY THAN OPPOSITION PARTISANS.

PRIMARY HYPOTHESIS II: NON-PARTISANS' EVALUATIONS OF THE NHS WILL BE LOWER THAN THOSE OF GOVERNMENT PARTISANS, BUT HIGHER THAN THOSE OF OPPOSITION PARTISANS.

PRIMARY HYPOTHESIS III: VOTERS' EVALUATIONS OF THE NHS WILL BE BETTER WHEN SERVICE PERFORMANCE METRICS ARE BETTER

PRIMARY HYPOTHESIS IV: RECENT SERVICE USE WILL AFFECT VOTERS' EVALUATIONS OF THE NHS

PRIMARY HYPOTHESIS V: HEALTH SERVICE RETRENCHMENT WILL RESULT IN RETROSPECTIVE VOTING AGAINST INCUMBENT POLITICIANS

In the remainder of this thesis, I will conduct a series of empirical investigations to test these primary hypotheses (with Roman numerals), as well as a set of secondary hypotheses that follow on from these main questions. Secondary hypotheses (with Arabic numerals) will be described in the relevant chapters.

5 The relationship between partisanship and public service performance evaluations

5.1 Introduction

As discussed in the review of the literature in chapter 2, partisanship or party identity is a key element of social-psychological theories of voter behaviour. It is described as a "sense of attachment" to a particular party, a long-term self-identification similar to one's own conception of belonging to a particular social class or religion (Campbell 1960). Although voters may occasionally vote differently, they self-identify as belonging to a particular party. This concept has long been measured in political surveys and polls using questions like "Generally speaking, do you think of yourself as Conservative, Labour, Liberal Democrat or something else?"

In chapter 4, I set out a theoretical argument that partisanship, while not the only factor affecting people's opinions, is very important in explaining differences in people's views about public services, including their evaluations of how well the current government is managing those services. Only highly credible information about performance (such as that gleaned from personal experience, or consistent indirect information) can alter people's partisan-informed priors, because of voters' desire to maintain a coherent, consistent set of beliefs that aligns with their partisan identity.

In this chapter I will carry out a study with a view to answering the research question: "to what

extent does partisanship explain differences in satisfaction with the performance of England's National Health Service (NHS)?". Using individual-level survey data from 1983 to 2019, I will use questions about satisfaction with how the NHS is being managed to measure differences between different groups of partisans over time. This analysis will demonstrate that:

- while government supporters tend to be more satisfied with health service performance in general, partisan differences have shrunk considerably in the new millennium;
- and satisfaction with the health service increased greatly across all political party identities between 2000 and 2010. Political party identity therefore became a much less reliable indicator of NHS satisfaction than it had been in the 1980s and 1990s.

I conclude that while partisanship remains an influential factor, this chapter indicates the need to explore alternative theoretical mechanisms that could result in variation in voters' evaluations of the NHS. This therefore lends support to the theory set out in chapter 4, which posits that performance, as well as partisanship, can inform public evaluations of services.

5.2 Expectations about the influence of partisanship

As discussed in chapter 4, when people are exposed to information, they are more likely to unquestioningly accept it if it accords with their existing views—but are more likely to dismiss or contest evidence that conflicts with their prior beliefs (Taber, Cann & Kucsova 2009). From a social-psychological point of view, this results from a phenomenon called motivated reasoning, in which partisans aim not only to find out the truth, but the 'right truth' that is consistent with their view of the world (Taber & Lodge 2006). However, from a Bayesian learning perspective, the same observed outcome results from voters combining the information from past experience they hold in their priors with the new information. Little (2022) argues that when we observe differential updating of views between partisan groups, we cannot distinguish between the 'motivated reasoning' and 'Bayesian learning' theoretical explanations. Either underlying expectation is compatible with the moderated learning perspective adopted in this thesis, since this merely requires that voters are able, to some extent, to learn from new information.

Whatever the underlying mechanism, expectations of objective government performance do vary significantly among partisan groups (Bartels 2002). Voters sympathetic to the government will hold more positive views about government performance because they readily accept information that reinforces their positive viewpoint and discard information that suggests government performance is declining or yielding negative outcomes. The opposite will hold for partisans of opposition parties, who will be more likely to assimilate negative information about government performance than positive information. This information may originate from personal experience, anecdote, news report or official information.

Social-psychological theory views partisan priors as being a kind of bias in information processing that either increases or decreases voters' evaluations of services. As such, it follows that voters who have weak priors because they do not see themselves as part of any particular partisan group should have less of this kind of bias. To test this, I hypothesize that non-partisans will be less satisfied with public services than government partisans (who are effectively cheerleading 'their team'), but more satisfied than opposition partisans (who are reluctant to recognise success on the part of their opponents). However, it should be noted that the Bayesian account would not make these assumptions about non-partisans, who may be less likely to update their priors if they are less politically engaged.

Applying this theory to the case of the English NHS, it would predict significant variation in satisfaction with the health service between partisan groups, rapid changes in satisfaction among partisans when governments change, and an otherwise slow evolution of public opinion in the aggregate. In this chapter I therefore set out to test the first two primary hypotheses enumerated in chapter 4:

> PRIMARY HYPOTHESIS I: GOVERNMENT PARTISANS WILL EVALUATE THE NHS MORE HIGHLY THAN OPPOSITION PARTISANS.
>
> PRIMARY HYPOTHESIS II: NON-PARTISANS' EVALUATIONS OF THE NHS WILL BE LOWER THAN THOSE OF GOVERNMENT PARTISANS, BUT HIGHER THAN THOSE OF OPPOSITION PARTISANS.

We can go further and set out some secondary hypotheses that flow logically from the primary hypotheses and the theory. Firstly, in chapter 4 I argued that consistent changes in performance could cause partisans to update their opinions of services. While in this chapter I will not directly measure performance, the corollary of this argument is that when there are noticeable changes in overall performance (such as the over-time changes in median waiting time demonstrated in chapter 3), the influence of partisanship on people's evaluations of services will be reduced. This therefore forms Secondary Hypothesis 1: "*The influence of partisanship will vary over time*".

Secondly, in chapter 4 I also argued that voters use the partisan composition of government as a heuristic for how well the government is performing on public services. It therefore follows that when government changes, voters will adapt their views, with supporters of the previous government becoming less satisfied and supporters of the new government becoming more satisfied. This is Secondary Hypothesis 2: "*Partisan voters will react to changes in government by rapidly revising their opinions of current service provision, as their party enters or leaves government*".

5.3 Data and Methods

The overall thesis aims to describe how voters evaluate public service performance. This chapter aims to explore the extent to which partisanship is associated with voters' evaluations of health service performance, because if this association is very strong, it could limit or 'crowd out' the potential influence of other factors, such as information about service performance, on voters' opinions. Since partisanship (or party identity) is not possible to randomly assign in an experimental setting, and is closely associated with demographics, political opinions and vote choice, I will conduct a descriptive analysis. This will explore overall and group-level associations in observational survey data, using regression models and control variables to adjust, as best as possible, for potential confounding.

I will therefore estimate a variety of regression models,beginning with bivariate logistic re-

gression and progressing to interacted and multi-level models, using repeated cross-sectional survey data to quantify the effect of personal demographic and political factors on satisfaction with the NHS across available years between 1983 and 2018 from the British Social Attitudes Survey. The demographic and political factors will be operationalized at the individual level; the higher level will consist of national and regional dummy variables, allowing the demographics' effect to vary in different geographies. Throughout the thesis, I will use the R statistical programming language[1] to carry out my analysis.

5.3.1 Data sources

In order to carry out this analysis, I decided that the data needed to cover a long period of time for maximum generalisability, and ideally cover periods of time when each of the main parties were in government and in opposition. This will allow differences between party groups (e.g. if Conservative or Labour supporters are systematically more satisfied with the NHS) to be distinguished from differences specific to the roles of government and opposition (e.g. if government supporters are systematically more satisfied with the NHS).

I also decided that the data should contain key variables pertaining to evaluations of the health service, as well as partisanship and a range of demographic variables that could otherwise confound the analysis. For example, if older people are more satisfied with the NHS, but also more likely to identify as Conservative, then not controlling for age would yield an inflated estimate of the association between Conservative identity and NHS satisfaction.

The data source that I judged best suits this analysis is the British Social Attitudes (BSA) survey. The BSA is a high quality, long-running national survey that uses computer-assisted face-to-face interviewing on an annual basis. Its respondents are drawn from a nationally representative sample created using probability sampling, which allows robust analysis and inferences

[1] Versions 4.3.0 and 4.3.1 of R were used throughout the thesis. In general, tidyverse packages were used for data cleansing and wrangling, with the {fixest} and {lme4} packages performing regression models, {marginaleffects} calculating slopes and predictions, and {did} package for difference-in-differences with differential treatment timing. The {modelsummary} package was used for presenting results, {sf} for mapping, and {WeightIt} and {cobalt} for entropy balancing and love plots, respectively (Arel-Bundock, Gassen, Eastwood, *et al.* 2023; Arel-Bundock 2023a; Bates, Mächler, Bolker, *et al.* 2015; Berge, Krantz & McDermott 2023; Callaway & Sant'Anna 2021a; Greifer 2023a, 2023b; Pebesma, Bivand, Racine, *et al.* 2023; R Core Team n.d.).

about the general population. This sampling procedure consists of random selection of post-code sectors, within which addresses are randomly selected, and the respondent within the address is randomly selected (Curtice, Hudson & Montagu 2020).

In order to track public opinion over time, I combine all publicly available waves of the survey into one file. In election years, the post-election government is coded as that year (since all the elections in this period of time took place in spring, before the data collection).

5.3.1.1 Response and predictor variables

In order to measure public satisfaction with the NHS, I operationalise the 'NHSSat' variable in two different ways. The variable represents the respondent's answer to the question "All in all, how satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays?". The options are "Very satisfied", "Quite satisfied", "Neither satisfied nor dissatisfied", "Quite dissatisfied" and "Very satisfied". In my data, these are coded 1-5, with 1 representing "Very dissatisfied" and 5 "Very satisfied".

The first way I use the NHSSat variable is to create a dummy variable for whether the person was satisfied or not, where "Quite satisfied" or "Very satisfied" is coded as 1, and the other possible responses are coded as 0. I will use this as the outcome variable in logistic regression. I will also conduct linear regressions with the NHSSat variable considered as an interval variable. By using the two approaches in combination, being aware of the strengths and weaknesses of each, the resulting analysis will be able to draw more robust conclusions.

In order to measure the effects of different aspects of partisanship, several different predictors will be used. These all derive from the 'party identity' questions in the survey. The standard process in the BSA is to ask "Generally speaking, do you think of yourself as a supporter of any one political party?". If they answer "yes", they will be asked which party. Otherwise, they are then asked "Do you think of yourself as a little closer to one political party than to the others?", and again, if they respond in the affirmative, they are asked which party. If they again give a

negative response, they are asked which party they would likely support if a general election were held. The resulting answer determines the person's party identity.

To determine the association between supporting the government of the day and being satisfied with the NHS' performance, I will first use a 'government partisan' predictor. This is a dummy variable representing whether the individual is a government partisan: a Conservative identifier in any year between 1983 and 1997 or 2010 and 2018, a Labour identifier in any year between 1998 and 2009, or a Liberal Democrat identifier during the Coalition government (2010-2014) is coded[2] with a 1. All other respondents are coded as 0, except for the few Scottish or Welsh Nationalist respondents, since these are at times also in government at the devolved level, making interpreting ideas of 'government partisanship' in that context more difficult.

The third predictor allows a direct comparison between the two main opposing partisan groups, without the influence of minor party identifiers or non-partisans. Here, a "Conservative" dummy variable is used, again based on the Party ID variable, where Labour identifiers are coded as 0, Conservatives as 1, and all others excluded from the dataset.

Finally, the fourth predictor is used to measure predicted satisfaction levels for people in specific party identity groups. This is a categorical variable containing all the party identities (Conservative, Labour, Liberal Democrat/Alliance, Other, None) as different levels.

5.3.2 Models and Expectations

5.3.2.1 Government partisanship and satisfaction

Primary Hypothesis I is that "*Government partisans will evaluate the NHS more highly than opposition partisans*". I will test this by measuring the extent to which government partisanship is generally associated with NHS satisfaction in the period under study, using a bivariate logistic regression where the predictor variable represents whether the respondent is a government partisan, and the response variable is the binary 'Satisfied' dummy variable. I will summarise

[2] In election years, the post-election government is coded for that year (since all the elections in this period of time took place in spring, before the data collection).

the results using the difference in predicted probability of being satisfied with the NHS between a typical government partisan and a typical opposition partisan or non-partisan.

Following this with a multivariate logistic model (model 2), I will check whether this result is robust when also accounting for demographic differences between government partisans and others. I will do this by adding control variables for gender, age group, marital status, socio-economic group, region, ethnic minority status, and year. Government partisanship should continue to show an independent influence on the response variable, if there is a true association between government partisanship and evaluations of the NHS.

5.3.2.2 Differences between party groups

I will then test Primary Hypothesis II ("*Non-partisans' evaluations of the NHS will be lower than those of government partisans, but higher than those of opposition partisans*"). I will test this using a linear Ordinary Least Squares regression, with partisan group as the main predictor, and satisfaction with the NHS (measured on the 5-point Likert scale) as the response variable. The 'partisan group' variable is derived from the 'government partisan' variable described earlier, but distinguishes between opposition partisans and non-partisans. If the theory holds, government partisans will have the highest average satisfaction, and opposition partisans will have the lowest average satisfaction. If the social-psychological view of non-partisans is correct, non-partisans' satisfaction should be between the two, as they are less responsive to partisan cues. If the Bayesian view of non-partisans is correct, this may not be the case.

5.3.2.3 Stability of partisanship's influence over time

To test Secondary Hypothesis 1's prediction that the influence of partisanship will vary over time, I will measure the influence of the 'government partisanship' variable on levels of satisfaction with the NHS in a linear regression model (M4), where 'year' is interacted with the 'government partisanship' dummy. The theoretical expectation here is that government par-

tisanship should be a consistently significant, strongly positive predictor of satisfaction with the NHS regardless of party in government.

As a second check on the influence of partisanship, I will also estimate a linear model that includes only Conservative and Labour identifiers (M5). This will allow me to isolate the difference in satisfaction between partisans of the two parties who have led government. I will report the regression coefficient for the interaction between the Conservative dummy variable and year, to illustrate how this partisan gap changes over time.

5.3.2.4 Partisan responsiveness to changes in government

Secondary hypothesis 2 predicts that "*partisan voters will react to changes in government by rapidly revising their opinions of current service provision, as their party enters or leaves government*". This is based on the idea that the partisan makeup of governments is used as a heuristic by partisans of how well government is performing. To test this hypothesis, I will estimate a multi-level model with party identity group as the categorical predictor (M6). I will supply the model with a variable for all party identity groups (Conservative, Labour, Liberal Democrat, Other and None). I will then generate predicted levels of NHS satisfaction for each different party identity in each year, on the original 1-5 Likert scale. The prediction will be conditional on all the control variables being held at their modal values: a married 17-32 year old White woman in the South East, with an intermediate non-manual socio-economic background. This will give a clear indication of the independent effect of party group on satisfaction with the management of the health service for each year in the data set.

Using a multi-level structure will allow me to estimate the influence of party in each year, without segmenting the data back into individual years and therefore losing information in the process. Gelman & Hill (2007) calls this 'partial pooling' of information. Each model uses the same set of covariates, but moves the over-time variation in partisanship to the second level (or 'random effects') by allowing the intercept and slope for the effect of partisanship to vary across years. This means that the model will estimate the extent to which identifying

with a particular party is associated with higher or lower evaluations of the NHS in each year, allowing it to change over time if the data suggests there has been a change.

Plotting predictions for each main party group over time will show any changes in satisfaction that coincide with changes in government. If this hypothesis is accurate, I would expect that changes in government would result in significant changes in satisfaction among partisans of parties either entering or leaving government, shortly after the election year.

5.3.3 Covariates and sample sizes

In order to isolate the effect of partisanship from potential demographic differences between partisan groups, I include a set of control variables that have been indicated in the literature as being influential on opinions about the NHS or some NHS services. These are gender (Welch & Thomas 1988), age (Calnan, Almond & Smith 2003), social class (Laycock 2009) and ethnic minority status (Laycock 2009). I also control for marital status and region, as other similar analyses have done (Calnan, Almond & Smith 2003; Welch & Thomas 1988).

In order to preserve statistical power and comparability between surveys over time, I have opted for broad categorical measures where practical, and chosen measures with the broadest coverage across years. Age is divided into categories with approximately equal membership.

Social class is measured by the 'Socio-Economic Group Compressed" variable. This variable underwent some changes after 2000, which are described in more detail in the footnotes[3]. However, given the variable's role only as a control to avoid confounding, rather than a key element of the theoretical model, this does not seem problematic for the study. Categories are further compressed for compatibility before and after 2000, such that "Professionals" are grouped with "Employers/Managers". "Other occupation", "Member of the armed forces" and "Occupation not classifiable" are grouped as "Other".

[3]In 2001, coding of Socio-Economic Group changed to the new National Statistics Socio-Economic Classification (NS-SEC). In order to allow time-series analysis, after 2000 NatCen provides variables based on their 'best estimates' of which Socio-Economic Group each individual would have been coded into, had the switch not been made. Analysis has shown that the post-2000 variable results in classification differences in around 17% of cases (Exley & Thomson 2002).

Ethnicity is generalised to a dummy for "ethnic minority". Marital status is updated to the most recent descriptions and the "Divorced" and "Separated" categories are combined, since these were not separately defined in all years.

Finally, respondents are split into nine English regions. Residents of Scotland and Wales are excluded. This thesis takes England as a case study, since other nations of the UK have devolved responsibilities for health provision, which has led to differences in management and performance metrics. The number of respondents in each demographic category is shown in Table 5.1 and Table 5.2 (table split for ease of reading).

Table 5.1: Summary statistics for demographic variables in the British Social Attitudes survey data

Variable	Category	Number of observations
Age Group	17-32	19,428
	33-42	17,851
	43-54	18,709
	55-67	18,360
	68+	18,435
Socio-Economic Group	Professional/employers/managers	17,161
	Intermediate non-manual	18,515
	Junior non-manual	16,036
	Supervisor/skilled manual	16,869
	Semi-skilled manual	15,554
	Other	884
	Unskilled Manual	4,388
Ethnicity	White	82,617
	Ethnic Minority	8,541

Table 5.2: Summary statistics for demographic variables in the British Social Attitudes survey data (continued)

Variable	Category	Number of observations
Party Identity	Conservative	29,021
	Labour	32,119
	Lib Dem/Alliance	9,867
	None	12,301
	Other	9,475
Marital Status	Not married	17,478
	Widowed	10,498
	Married	45,942
	Living as married	7,381
	Separated or divorced	11,411
Region	Greater London	12,076
	North East	5,896
	North West	12,829
	West Midlands	9,746
	Yorkshire and the Humber	9,944
	East Midlands	8,750
	East of England	8,352
	South West	9,705
	South East	15,485

5.3.3.1 Overall sample sizes

Figure 5.1 below shows the unweighted sample size by year. The overall unweighted sample size ranges between 1153 and 3877 respondents (1148.48 and 3863.18 in the weighted data),

depending on the year.[4]

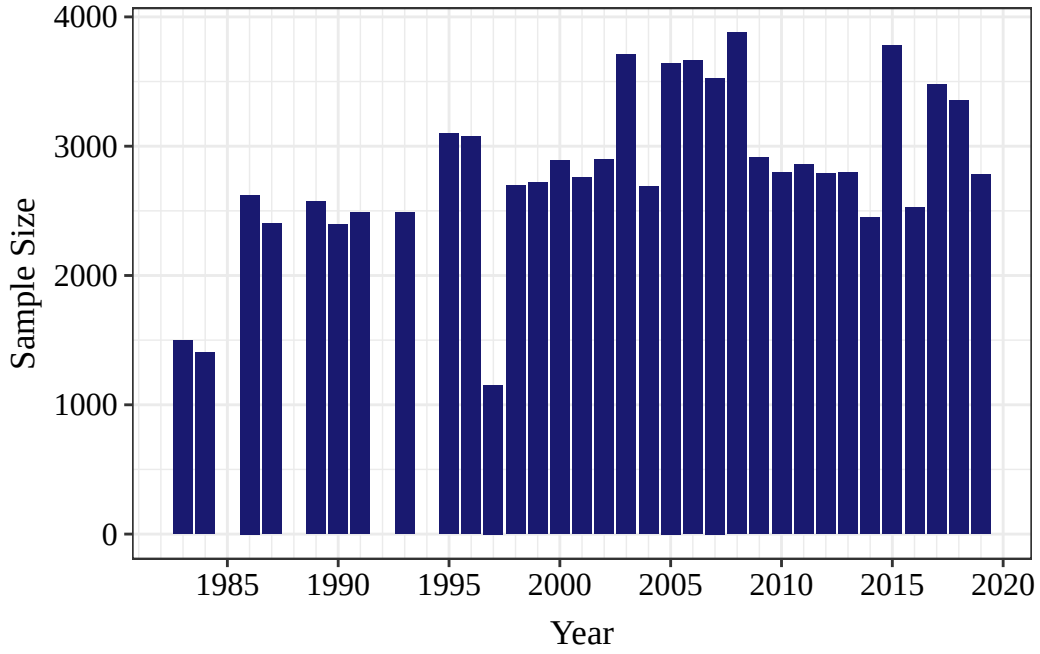

Figure 5.1: Sample size (number of respondents) by year from the British Social Attitudes survey

5.4 Results

5.4.1 Government partisanship and satisfaction

In order to test Primary Hypothesis I ("*Government partisans will evaluate the NHS more highly than opposition partisans*"), I first carry out a bivariate logistic regression (M1). This model aims to predict NHS satisfaction based on whether or not the respondent was a 'government partisan'—i.e., their party identity was aligned with that of the Prime Minister, over all years in the dataset.

[4] In this chapter I use unweighted data, despite the British Social Attitudes Survey's complex sampling structure. While it makes sense to use weights when attempting to estimate parameters that are nationally representative in a particular year, multi-year weights and strata information are not provided, and the more complex models estimated in this chapter already account for demographic differences using statistical controls.

Table 5.3: Logistic regression (M1) results, showing government partisanship is associated with NHS satisfaction

	Bivariate logistic regression (M1)
(Intercept)	−0.163*** [95pc CI: −0.181, −0.145]
Government Partisan	0.352*** [95pc CI: 0.322, 0.382]
Num.Obs.	74 145
Std.Errors	IID

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5.3 shows the estimated coefficients on a log-odds scale for the intercept and the 'government partisan' variable. This indicates that in this time period as a whole, government partisans are more likely to be satisfied than others, and the difference is statistically significant at the $p = .05$ confidence level.

We can express the group difference in terms of probabilities. In this bivariate model, government partisans have an expected 54.7% probability of being satisfied with the NHS, compared to 45.9% for others.

This 8.8% difference in probabilities suggests that partisanship is somewhat associated with NHS evaluations. However, there are very likely to be systematic demographic differences between party groups, which could act as a confounding variable. To test whether this finding is still valid after accounting for demographic differences, I now proceed to estimate a more complex model by controlling for gender, age group, marital status, social class, region, ethnicity, and year. Comparing the magnitude of the coefficients for these controls will also give some indication of how to interpret the magnitude of partisanship's influence on evaluations.

As Table 5.4 shows, the positive effect of government partisanship on satisfaction with health provision remained significant after controlling for demographic variables and year, confirming the expectations of the partisanship-based account of satisfaction as described in H1. Taking a hypothetical respondent with all covariates set at their modal values (a married 17-32 year old White woman in the South East, with an intermediate non-manual socio-economic background), the average satisfaction over all years for a government partisan is 38.3% compared

Table 5.4: Multivariate logistic regression (M2) results, showing government partisanship is associated with NHS satisfaction when controlling for demographic differences

	Multivariate logistic regression (M2)
(Intercept)	−0.473*** [95pc CI: −0.621, −0.325]
Government Partisan	0.377*** [95pc CI: 0.345, 0.409]
Female	0.016 [95pc CI: −0.019, 0.051]
Ethnic Minority	−0.115*** [95pc CI: −0.178, −0.051]
Age category: 33-42	0.005 [95pc CI: −0.046, 0.055]
Age category: 43-54	0.062* [95pc CI: 0.011, 0.114]
Age category: 55-67	0.177*** [95pc CI: 0.124, 0.230]
Age category: 68+	0.645*** [95pc CI: 0.587, 0.703]
Marital status: Widowed	0.031 [95pc CI: −0.038, 0.101]
Marital status: Married	0.100*** [95pc CI: 0.053, 0.147]
Marital status: Living as married	−0.001 [95pc CI: −0.067, 0.065]
Marital status: Separated or divorced	0.001 [95pc CI: −0.060, 0.062]
SEG: Intermediate non-manual	0.037 [95pc CI: −0.013, 0.088]
SEG: Junior non-manual	0.094*** [95pc CI: 0.041, 0.147]
SEG: Supervisor/skilled manual	0.115*** [95pc CI: 0.065, 0.165]
SEG: Semi-skilled manual	0.178*** [95pc CI: 0.126, 0.231]
SEG: Other	0.199* [95pc CI: 0.043, 0.354]
SEG: Unskilled manual	0.263*** [95pc CI: 0.184, 0.342]
North East	0.327*** [95pc CI: 0.252, 0.401]
North West	0.185*** [95pc CI: 0.124, 0.245]
West Midlands	0.199*** [95pc CI: 0.135, 0.264]
Yorkshire and the Humber	0.248*** [95pc CI: 0.183, 0.312]
East Midlands	0.336*** [95pc CI: 0.269, 0.403]
East of England	0.198*** [95pc CI: 0.129, 0.266]
South West	0.346*** [95pc CI: 0.281, 0.410]
South East	0.107*** [95pc CI: 0.049, 0.164]
Num.Obs.	70 087
Std.Errors	by: serial

Coefficients for year dummies are excluded for brevity, but visible in the coefficient plot.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

to 29.9 for others—a 8.4% gap, slightly narrower than the bivariate model's prediction. The difference is statistically significant at the p = 0.05 level; 95% confidence intervals were 7.7% - 9.2%. Standard errors are clustered by individual ID, to mitigate potential repeated sampling of the same individuals.

Coefficients, including control variables, are shown in Table 5.4 and Figure 5.2 (year dummy variables are only shown in the plot). As expected, there are demographic differences in satisfaction; respondents who are older, married or living outside Greater London were more likely to be satisfied with the NHS. Professionals/employers/managers were modelled as less likely to be satisfied than most other socio-economic categories.

Three features of the results in particular raise questions. Firstly, there are a number of demographic dummy variable coefficients that are similar in scale or larger than the coefficient for government partisanship: the variable with the largest estimated effect in the model is the "68+" age category dummy, and several of the region indicators are very similar in magnitude to government partisanship. This indicates that while partisanship is clearly important, it is not dominant and other factors have a significant association with satisfaction with the health service in England.

The second feature of note is the comparatively large effect of year dummy variables, indicating that over-time differences may be more significant than between-groups differences in satisfaction. The 'year' dummy estimates range from around -0.78 in 1997 to 0.71 in 2010, where the reference category is the year 1983. This compares to an estimate of 0.38 for government partisanship. This implies that there were large shifts in overall satisfaction over the time period under study that were much more influential on individuals' satisfaction than demographic or partisan differences. This finding demonstrates that there is considerable variation in NHS evaluations that is time-variant and not explained by government partisanship.

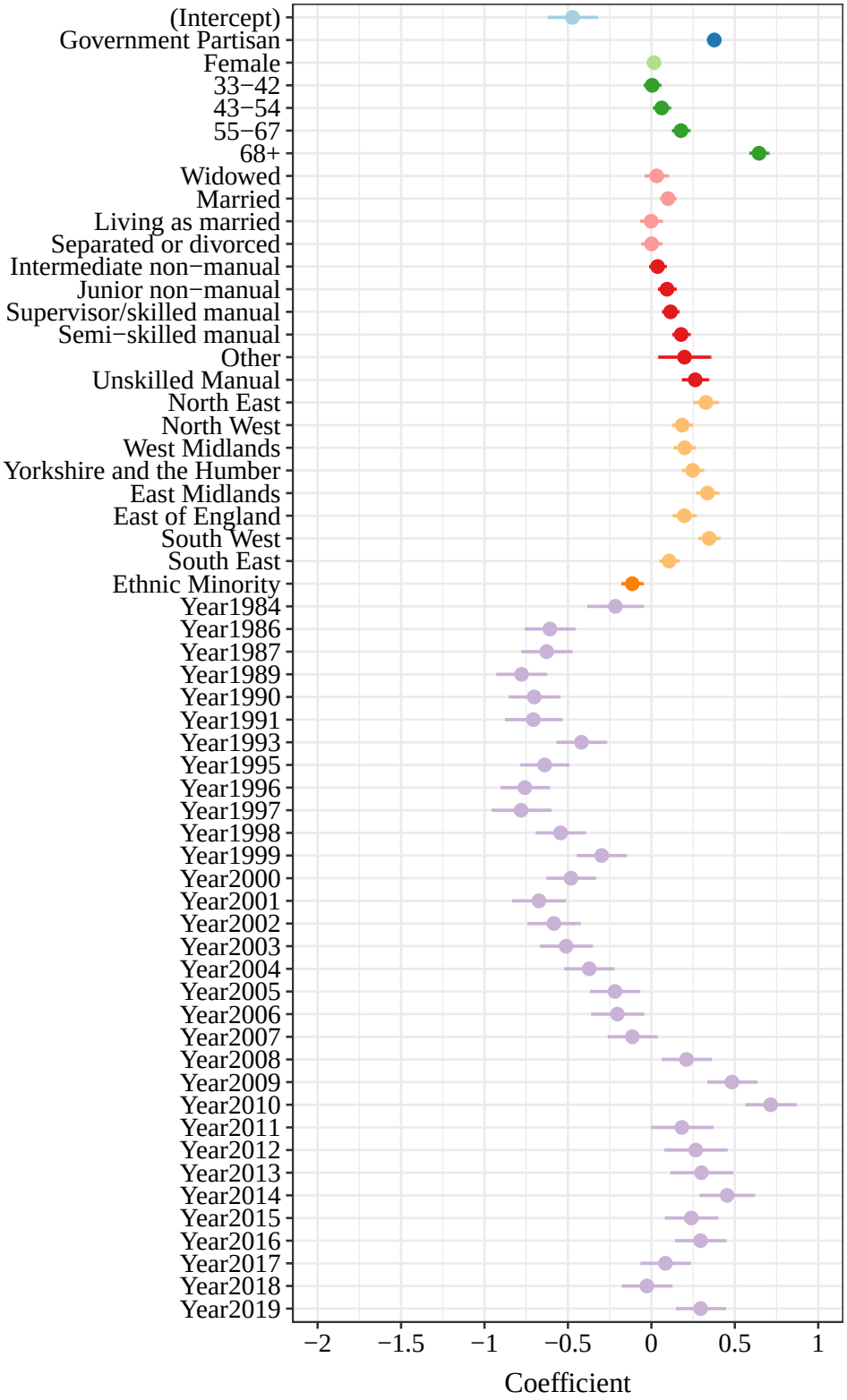

Figure 5.2: Regression coefficients from the multivariate logistic regression (M2), showing government partisanship is associated with NHS satisfaction when controlling for demographic differences. Points representing different categories of the same variable share the same colour on the plot

5.4.2 Differences between party groups

I will now use a linear regression model (M3) to test Primary Hypothesis II ("*Non-partisans' evaluations of the NHS will be lower than those of government partisans, but higher than those of opposition partisans*"). The predictor variable is a categorical measure of party identity with three levels: government partisan, non-partisan and opposition partisan. NHS satisfaction on a five-point scale is the response variable. Demographic controls as listed in the Data and Methods section are included, as is a dummy variable for year.

Results from the regression are shown in Table 5.5 (year dummy variables excluded for concision). In the results table, "Non-Partisan" is the reference category, which means that the coefficients for "Government Partisan" and "Opposition Partisan" represent the estimated difference between each of those groups and the non-partisan group. Both coefficients are statistically significant at the $p = .05$ level, and the estimates indicate that satisfaction for government partisans was estimated to be around 21% of a step on the 5-step Likert scale higher than that of non-partisans (95% CI: 18% to 24% difference), whereas opposition partisans were estimated to be around 5% of a step less satisfied than non-partisans (95% CI: -8% to -2% difference).

This means that Primary Hypothesis II is validated: during the period under study, government partisans evaluated the NHS more highly than non-partisans, and non-partisans evaluated the NHS more highly than opposition partisans.

5.4.3 Stability of partisanship's influence over time

The next analysis tests Secondary Hypothesis 1: "The influence of partisanship will vary over time". I do this by estimating M4, a linear regression model that aims to measure the effect of government partisanship on levels of satisfaction (on a five-point scale) in each year, by interacting year and government partisanship.

The results are shown in Figure 5.3. The points represent the coefficient of the estimated effect of being a pro-government partisan. The shaded areas represent 95% confidence intervals;

Table 5.5: Multivariate OLS regression results from M3, showing that government partisans are more satisfied on average than non-partisans, whereas opposition partisans are less satisfied on average than non-partisans.

	Multivariate OLS regression (M3)
(Intercept)	3.037*** [95pc CI: 2.949, 3.125]
Government Partisan	0.214*** [95pc CI: 0.184, 0.243]
Opposition Partisan	−0.052*** [95pc CI: −0.080, −0.024]
Female	−0.025* [95pc CI: −0.044, −0.005]
Ethnic Minority	−0.060** [95pc CI: −0.096, −0.024]
Age category: 33-42	−0.032* [95pc CI: −0.061, −0.003]
Age category: 43-54	−0.022 [95pc CI: −0.051, 0.007]
Age category: 55-67	0.066*** [95pc CI: 0.036, 0.097]
Age category: 68+	0.372*** [95pc CI: 0.340, 0.405]
Marital status: Widowed	0.015 [95pc CI: −0.024, 0.054]
Marital status: Married	0.027* [95pc CI: 0.001, 0.054]
Marital status: Living as married	−0.029 [95pc CI: −0.067, 0.008]
Marital status: Separated or divorced	−0.032+ [95pc CI: −0.067, 0.002]
SEG: Intermediate non-manual	−0.007 [95pc CI: −0.035, 0.022]
SEG: Junior non-manual	0.048** [95pc CI: 0.018, 0.078]
SEG: Supervisor/skilled manual	0.027+ [95pc CI: −0.002, 0.055]
SEG: Semi-skilled manual	0.063*** [95pc CI: 0.033, 0.093]
SEG: Other	0.114* [95pc CI: 0.024, 0.204]
SEG: Unskilled manual	0.095*** [95pc CI: 0.051, 0.139]
North East	0.185*** [95pc CI: 0.143, 0.228]
North West	0.115*** [95pc CI: 0.081, 0.149]
West Midlands	0.116*** [95pc CI: 0.080, 0.153]
Yorkshire and the Humber	0.151*** [95pc CI: 0.115, 0.187]
East Midlands	0.204*** [95pc CI: 0.166, 0.242]
East of England	0.127*** [95pc CI: 0.088, 0.165]
South West	0.228*** [95pc CI: 0.191, 0.264]
South East	0.090*** [95pc CI: 0.057, 0.122]
Num.Obs.	70 087

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

dotted vertical lines represent years in which that party either entered government, left government, or changed coalition status (e.g. the Conservatives changing from coalition partners to single-party government in 2015).

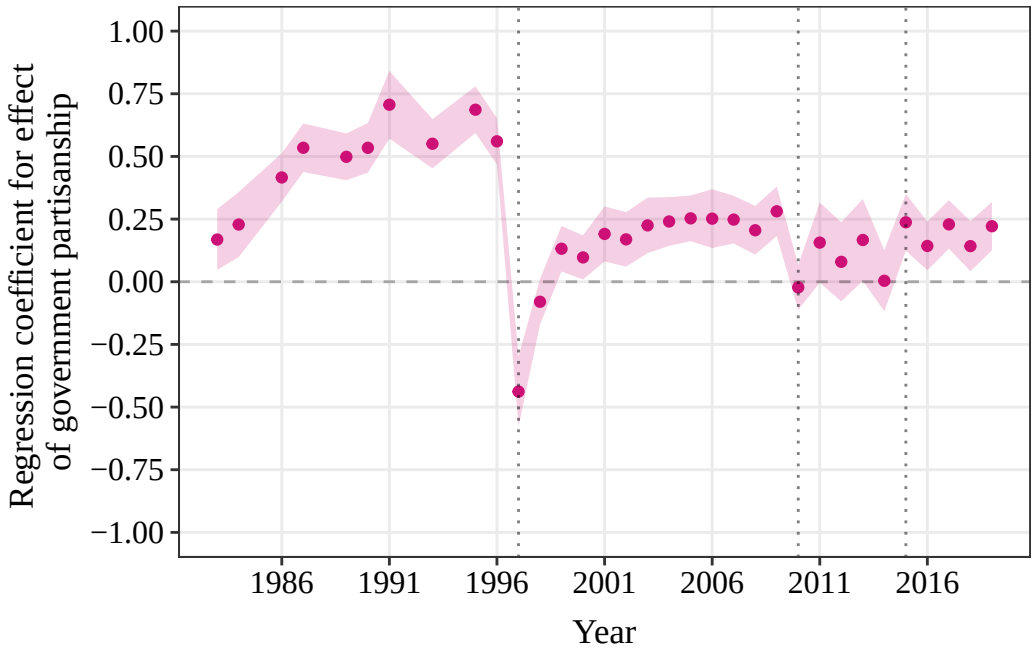

Figure 5.3: Plot of the 'government partisan' coefficient in each year from OLS regression (M4), showing how the estimated association between government partisanship and satisfaction with the NHS changes over time

Figure 5.3 shows that partisanship became increasingly important during the 1983-1987 parliament, with its estimated coefficient rising to above 0.5. This represents an estimated additional half of a point of satisfaction on the 5-point Likert scale on average. This coefficient remains around 0.5 until the 1997 election.

Following the change of government in 1997, the effect of government partisanship fell to around 0.25, which is approximately half its previous size. During the Conservative/Liberal Democrat coalition years[5], the effect appears to fall slightly, although there is greater variation and uncertainty in the data. With the exception of those years, the coefficient for government partisanship remains at or around 0.25 for the remainder of the years under study.

[5] 2010-2014, as the 2015 data was collected after the election

While an argument could be made that normal partisan distinctions broke down somewhat during the 2010-2015 coalition, which could have resulted in partisanship having less influence, the lower estimated effect of government partisanship from 1997 onwards indicates that while government partisanship is clearly associated with increased satisfaction in general, other factors grew in importance after around 1997.

One potential explanation for changes in the effect of partisanship on the model is that non-partisans or minor party identifiers could have changed their views. To rule this out, we can examine the differences between supporters of the main two parties to determine if the above finding still holds when excluding all others.

Figure 5.4 shows the marginal effect of being a Conservative identifier, as compared to a Labour identifier, from the M5 regression model. This model measures the average partisan gap between Conservative and Labour identifiers in NHS satisfaction; all non-Labour, non-Conservative identifiers are excluded from the dataset. Because the model contains an interaction term between year and the "Conservative" dummy, the marginal effects of this variable are analogous to the coefficients for the "Conservative" dummy if separate models were estimated for each year.

It shows that Conservatives were more satisfied with the NHS than Labour identifiers throughout the Thatcher and Major governments, at times being on average more than half a point on the 5-point scale more satisfied. With the exceptions of 1998 and possibly the year 2000, Labour identifiers were more satisfied with the NHS than Conservative identifiers during the Labour government, although the gap was approximately half the size. The Coalition years were statistically ambiguous, although a small gap then opened up from 2015, with partisans of the governing Conservatives again more satisfied with the health service.

This finding confirms the results shown on the previous graph, while also showing that the change in strength of partisanship was not driven by changes in the baseline (e.g. if there were shifts among non-partisans or people who identify with minor parties). It demonstrates that, among partisans of the two main parties, the influence of partisanship on satisfaction has

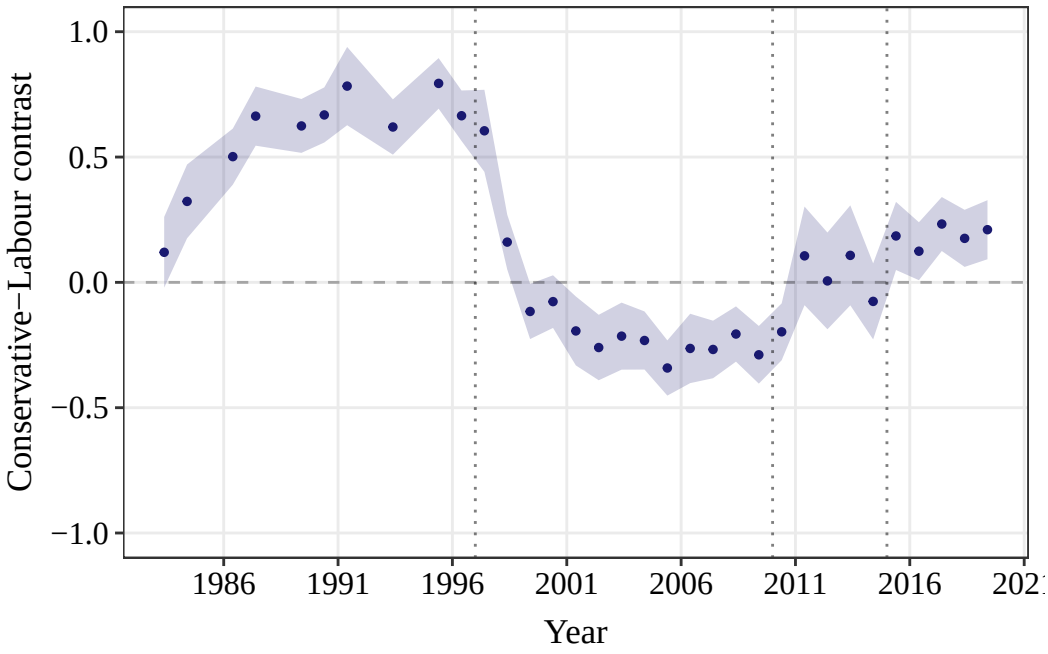

Figure 5.4: Marginal contrasts from two-party Conservative-Labour regression model (M5), showing how the partisan gap in NHS satisfaction between Conservative and Labour has varied over time

varied over time and overall has reduced since the 1980s and 1990s. This supports Secondary Hypothesis 1.

5.4.4 Partisan responsiveness to changes in government

Finally, I test Secondary Hypothesis 2 ("*Partisan voters will react to changes in government by rapidly revising their opinions of current service provision, as their party enters or leaves government*"). To do this, I estimate a multi-level regression model (M6) with party group as the main predictor. Party group has a fixed effect, but is also entered into the model as a grouping variable with its own random intercepts for each party group. Year is added as a random effect slope, to allow the predicted satisfaction levels for each party group to vary over time. This is equivalent to estimating a model for each year, but has the benefits of partial pooling as described in the Data and Methods section.

In order to examine how predicted satisfaction changes in response to changes in government, I create a set of predictions for a hypothetical typical respondent. All covariates are held at their modal values—in other words, predictions for a married 17-32 year old White woman in the South East, with an intermediate non-manual socio-economic background—with only party identity varying over time. These predictions are then plotted as time series for each of the three main parties; the shaded area represents the 95% confidence intervals for these predicted values, calculated using the delta method (Arel-Bundock 2023b). Government and coalition changes are again represented by dotted vertical lines, but only where the party in focus was involved.

In Figure 5.5, there is a clear decline in satisfaction after 1998, following Labour taking office in 1997[6]. However, after the Conservatives take office in 2010, Conservative satisfaction actually falls; after they take majority control in 2015, it falls again. After losing their majority in 2017, Conservative identifiers' satisfaction falls slightly, then recovers.

[6] Although 1998's satisfaction is higher than that in 1997, because of the ambiguity of asking about government performance during an election year, it is likely wise to ignore election-year anomalies in this data series.

While it is arguable whether taking and losing majority control of government in 2015 and 2017 should elicit a change in views among Conservative identifiers (since in any case the government was Conservative-led), the fact that Conservative satisfaction fell after 2010 rather than rose, means that the evidence for Secondary Hypothesis 2 is somewhat mixed in the case of Conservative identifiers.

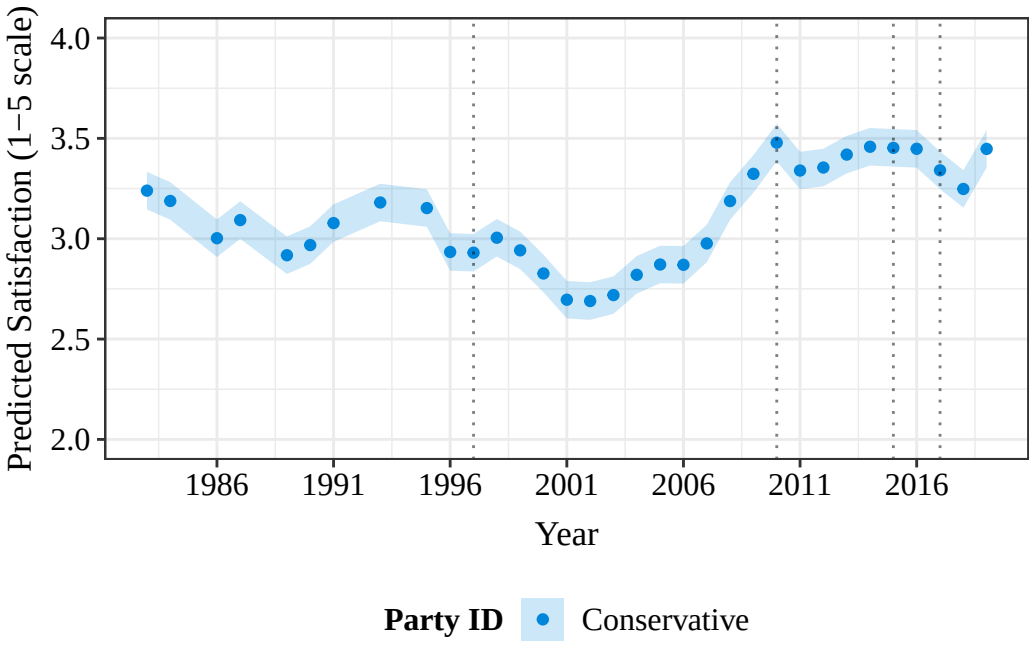

Figure 5.5: Predicted mean satisfaction from M6 (Conservative identifiers)

Among Labour supporters, as shown in Figure 5.6, the change to a Labour government in 1997 was followed by a large spike in satisfaction until 1999. When the party left office in 2010, there was a significant fall in satisfaction. In the case of Labour, therefore, there seems to have been a clear case of partisans realigning their views as governments changed.

Among Liberal Democrats, however, Figure 5.7 shows that entering the Coalition government in 2010 was accompanied not by a rise, but by a fall in satisfaction with the NHS. Satisfaction fell again after the party left government in 2015. This again presents a mixed picture as to the validity of Secondary Hypothesis 2, since we would expect joining government to have a positive effect on Liberal Democrat evaluations of public services, including NHS satisfaction.

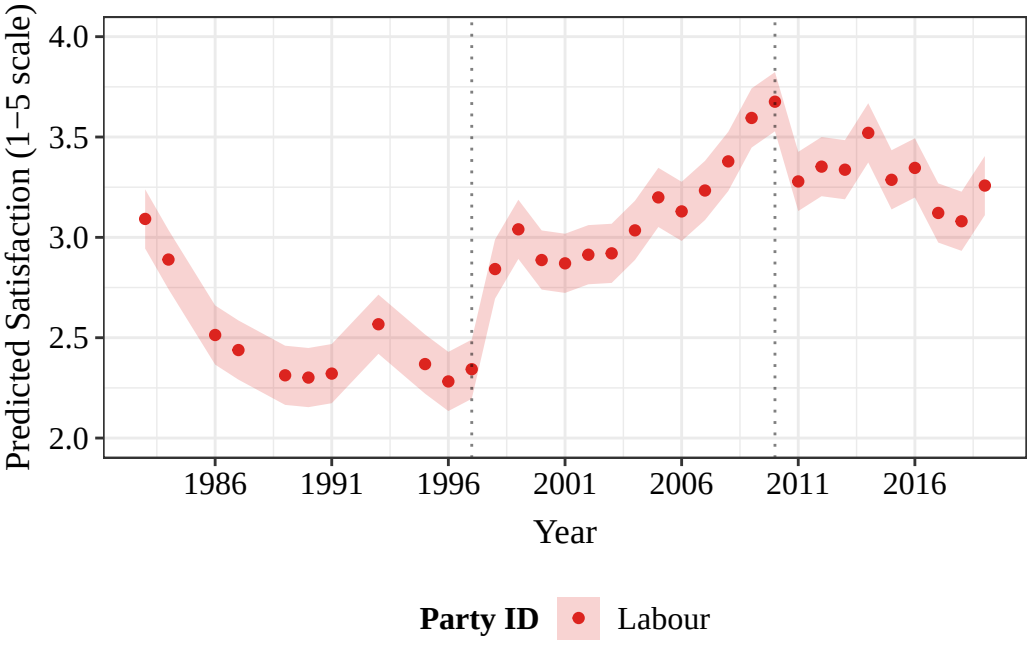

Figure 5.6: Predicted mean satisfaction from M6 (Labour identifiers)

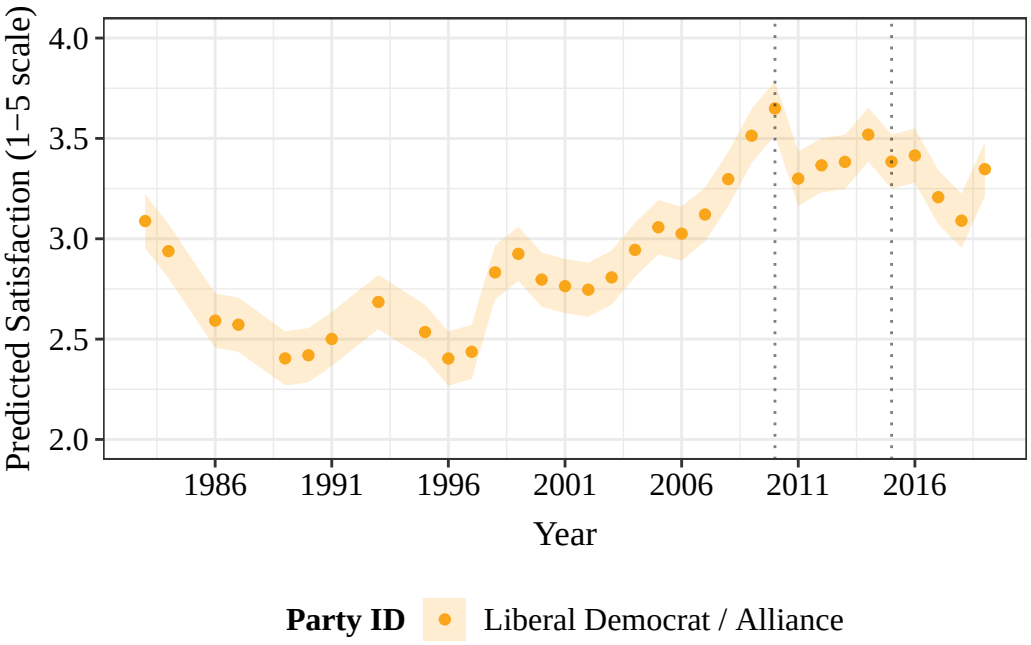

Figure 5.7: Predicted mean satisfaction from M6 (Lib Dem/Alliance identifiers)

There is therefore mixed evidence in relation to Secondary Hypothesis 2. Some changes in government resulted in dramatic changes in satisfaction among some party groups, while in other cases satisfaction moved in the opposite direction to that predicted by the theory. This shows that while partisanship is associated with different levels of satisfaction, there are clearly other factors that are at least as important for understanding public evalutions of the NHS.

On a more general note, there are some patterns in the modelled data that, while not directly relating to the hypotheses above, are of interest in relation to the general question of how and when partisanship is most influential. There is a particularly noticeable shift in satisfaction over time is the overall rise during the 2010s, the magnitude of which is much greater than differences between party groups at the time. Figure 5.8 is a rescaled closer examination of the output of M3 during this period, with only the two main party groups included for clarity. The predicted satisfaction for a typical Labour identifier in 2000 is 2.9 out of 5 compared to 2.8 for a Conservative; by 2010 this has increased to 3.7 for the Labour identifier and 3.5 for the Conservative. The predicted difference between typical party identifiers (i.e. with demographics held at their mode) is therefore 0.1 - 0.2 depending on the year, whereas the difference over time between 2000 and 2010 was 0.7 points for the typical Conservative and 0.8 points for the typical Labour partisan.

Although overall satisfaction reduced slightly from 2011 onwards, all partisan groups were more satisfied in 2018 than their equivalents for most of the 1990s.

I also note that in the multivariate (M2) model outputs, partisanship is less strongly associated with satisfaction than either the survey year or the age group of the respondent. As discussed in the results section, this indicates that over-time variation may be more important than between-groups variation, with the exception of age differences. These age differences may also point to an alternative theoretical explanation, such as the role of personal contact with services.

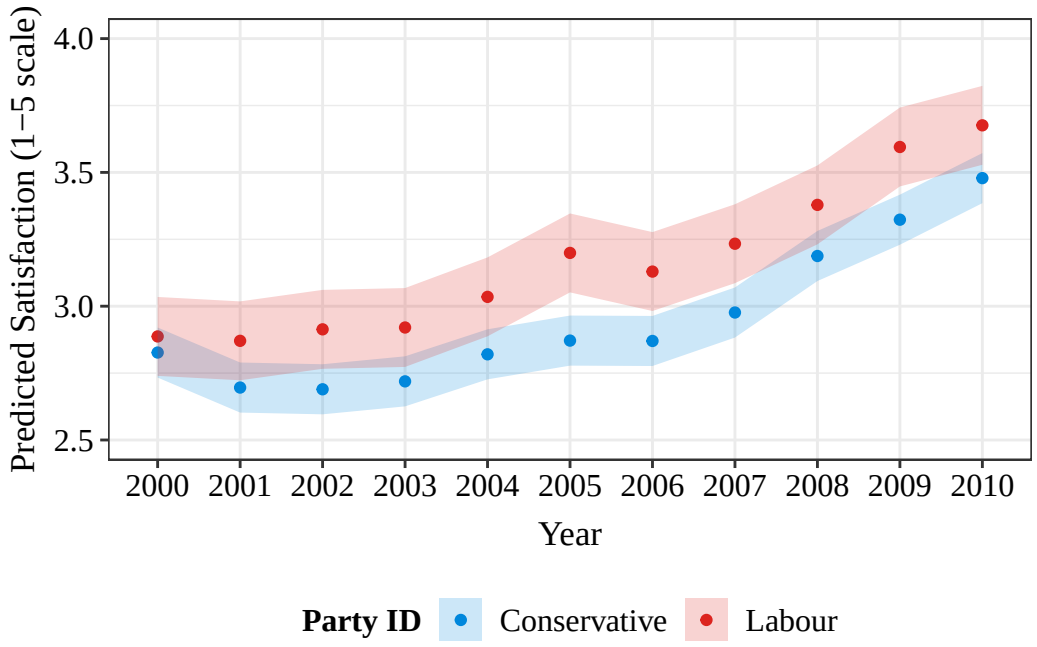

Figure 5.8: Predicted mean satisfaction from M6 (Conservative and Labour identifiers from 2000-2010)

5.5 Discussion

In this chapter, survey data over three decades was used to explore the association between political partisanship and satisfaction with the NHS, with the aim of answering the research question: "to what extent does partisanship explain differences in satisfaction with the performance of England's National Health Service (NHS)?"

I found that respondents who identify with a party in government are more likely to be satisfied with the NHS, compared with others. This confirms that there is a general association between partisanship and evaluations of the NHS. In this time period, the model predicts that government partisans were approximately 8% more likely to be satisfied with the NHS than opposition partisans. These findings suggest that people being more likely to report higher levels of satisfaction when their preferred party is in government. The models therefore provided evidence to support Primary Hypothesis I ("*Government partisans will evaluate the NHS more highly than opposition partisans*").

Similarly, I also found that government partisans had much higher NHS satisfaction levels, on average, than non-partisans; and that non-partisans had higher average NHS satisfaction levels than opposition partisans. This indicates that there is not only a positive association between supporting the current government and NHS evaluations, but there is also a negative association between supporting an opposition party and NHS evaluations. This supports Primary Hypothesis II ("*Non-partisans' evaluations of the NHS will be lower than those of government partisans, but higher than those of opposition partisans*").

However, the results also indicated that there is considerable variation in evaluations that is not explained by partisanship. Firstly, the association between government partisanship and NHS satisfaction seems to have reduced over time. Multi-level modelling estimated that partisan gaps grew notably smaller during the 1997-2010 Labour government, and even smaller still during the 2010-2015 Conservative/Liberal Democrat coalition period. Secondary hypothesis 1, the idea that "*the influence of partisanship will vary over time*", was therefore supported. This points to the potential for performance and other factors to play a stronger role in some contexts and time periods than others.

In some areas the evidence was ambiguous. When examining trends in model-adjusted NHS satisfaction over time by party group, there was no clear evidence that partisans always react to their party gaining or losing power. In some cases, such as Labour identifiers when their party entered or left office, there were rapid changes in satisfaction in the expected directions. In the cases of Conservative and Liberal Democrat partisans, though, this was less clear-cut, with the expected revisions either not materialising, or appearing in the opposite direction (such as Liberal Democrat satisfaction falling following the 2010 election). This therefore only partly supports Secondary Hypothesis 2, which argued that "*Partisan voters will react to changes in government by rapidly revising their opinions of current service provision, as their party enters or leaves government*", saw conflicting evidence.

Overall, the evidence from this chapter suggests that while partisanship is clearly very important and explains a proportion of variation in people's evaluations of public service performance, its importance is context-dependent, and much variation is not explained by par-

tisanship. This is encouraging from a retrospective evaluation perspective, as a very strong association between partisanship and evaluations could have 'crowded out' any possible influence from information about service performance. In the following chapters, I will examine how differences in health service performance and provision—communicated directly through personal experience, or indirectly through word of mouth and media coverage—are associated with people's evaluations of the NHS in England, and by extension, with public services more generally.

6 NHS performance and public evaluations of the NHS

6.1 Introduction

The central research question in this thesis is to what extent the public is able to recognise and make judgements about variation in public service performance. The more capable they are of doing this, the greater their ability to hold governments accountable for the services they fund and deliver - but according to the 'moderated learning' theory put forward in this thesis, voters' ability to recognise these changes may be impeded by partisan priors and scarce information.

In chapter 5, I analysed the relationship between partisanship and public opinion and found that while partisanship is important, there is still considerable unexplained variation in evaluations of the health service. In this chapter I explore the association between performance of health services in England - as measured by official statistics - and the public's views of the service. In existing work on retrospective voting in the UK context, the reaction of voters to other aspects of government performance have been tested, such as economic performance (Johnston & Pattie 2001a). Where health service performance has been analysed, this is generally in combination with other variables; for example, Green & Jennings (2017b) finds that the length of NHS waiting lists was associated with better issue handling ratings for the Labour government between 2004 and 2010. This study adds to the literature on public service performance evaluation by performing a similar analysis across a much longer period of time, and

also by measuring the effect of geographic variation in performance.

The analysis combines public opinion data from individual-level survey responses with official NHS performance statistics, and uses regression modelling to estimate the association between different lengths of waiting times and levels of satisfaction with the health service. In this way, it gives an overall view of the relationship between public service performance and public satisfaction that also accounts for demographic and political variation in the public.

I will use this strategy to carry out two tests: the first will measure changes in average satisfaction between years over three decades and compare this to changes in annual National Health Service (NHS) performance. This over-time analysis will show that people are less satisfied with the health service in years when waiting times are longer, and that this influence is of a similar magnitude to the influence of partisanship. As I will discuss, these findings lend support to the theory that both partisanship and performance information play significant roles in informing voters' satisfaction with the NHS in England.

I will also test whether the overall association between performance and satisfaction holds among respondents with different party identities, building on the over-time analysis performed for the first test. This will confirm that, as expected, people of all party identities were more satisfied with the government's management of the NHS in years when average waiting times were shorter, and less satisfied in years when they were longer. The implications of these findings for the moderated learning theory will be discussed.

I will then create a set of models to estimate how much people's evaluations of the NHS varies depending on geographical differences in NHS performance. This will show that there is a very small association between longer average waiting times for hospital treatment within a local area and worse evaluations of the NHS among people living there. This test of geographical differences in opinion is a hard one, since there is considerably more variation in national performance over time in the long term than there is between areas of England in a given time period, and national media is more able to communicate signals about NHS performance than local media, because it has greater reach with the public. Therefore, if an association is found,

this indicates that voters are able to recognise and respond to even weak signals about service performance.

Taken together, these findings support Primary Hypothesis III ("*voters' evaluations of the NHS will be better when service performance metrics are better*") by demonstrating that both geographical and temporal variation in performance are associated with differences in NHS evaluations.

6.2 NHS performance and public evaluations: expected associations

In this thesis' literature review chapter, I summarised how different schools of thought within political science have made different arguments about the relationship between public opinion and government performance. From the rational choice and valence perspectives, voters evaluate government performance in order to determine which vote choice will maximise the expected utility they will receive (Clarke, Sanders, Stewart, *et al.* 2004; Fiorina 1981). By contrast, some proponents of social-psychological theories argue that voters shape their opinions to match those of their peers within the same social and political group - an example of motivated reasoning (Bartels 2002; Taber, Cann & Kucsova 2009). In the economic voting literature, the rational choice perspective has been tested in relation to government performance in general, especially on economic management (see contrasting reviews by Lewis-Beck & Stegmaier (2007) and Healy & Malhotra (2013)), and research into differences in performance among local authorities has shown that this may result in changes to voting patterns (James & John 2007). However, there is a dearth of research into the relationship between health service performance and satisfaction, outside of descriptive work by think tanks such as (Maybin 2010). In this chapter, I will provide a detailed analysis of that relationship performance in the context of the theory I put forward in chapter 4.

In that chapter, I described how public opinion would respond to performance differences from a "moderated learning" perspective. I proposed that the public has partisan-informed priors

about public service performance but that voters can change their minds in response to credible evidence that disconfirms those priors. In this chapter I propose to examine responsiveness of public opinion to variation in the NHS in England's performance. The 'moderated learning' theory in chapter 4 included a primary hypothesis about the effect of performance:

> PRIMARY HYPOTHESIS III: VOTERS' EVALUATIONS OF THE NHS WILL BE BETTER WHEN SERVICE PERFORMANCE METRICS ARE BETTER

In order to test this hypothesis, I create two tests for my theoretical expectations, one of which is more demanding than the other. Each test addresses a set of secondary hypotheses that helps to validate or challenge the overall primary hypothesis.

The first set of tests relates to that national picture: over the longest possible time horizon, which in this study is three decades long, do people give better evaluations of the NHS in years when performance is better? This test requires people to be able to recognise health service performance (not necessarily in their own local area or from personal experience) and to be able to form a relative judgement about it in comparison with previous years. There are reasons to expect them to do so: over a long time period, people can accumulate scarce signals about performance and alter their views accordingly, and they can use information for this test from any source, local or national. This is important, since the national media sets the tone and frames the political debate, including publicising poor NHS performance. By contrast, local media in the UK has been in serious decline since 2009 (Digital, Culture, Media and Sport Committee 2023). I can therefore set out my first Secondary Hypothesis for this chapter:

> Secondary Hypothesis 1: variation in national performance over time will be positively associated with NHS satisfaction.

As well as setting expectations about overall public opinion, my theory predicts that changes in performance can affect the opinions of partisans, regardless of their party identity. In other words:

> Secondary Hypothesis 2: variation in national performance will be positively associated with better NHS satisfaction, within all party identity groups.

I will check whether this holds true by repeating the over-time analysis, but applying it to each broad partisan group within the electorate. This will indicate whether public responsiveness to performance changes is limited to some types of partisan voter, or whether the influence of performance signals can shift partisan priors across the political spectrum.

The second set of tests, which is the more demanding of the two, focuses on whether people living in areas with better NHS performance evaluate the health service more highly than people living in areas with worse performance. If so, the electorate could apply pressure and even threaten to remove politicians in poorly-performing areas, even when national performance is good. This would therefore introduce local accountability for public service performance, but would require voters to recognise and process information about performance in their local area, as distinct from the national picture.

I argue that this is a fairly challenging requirement. On the one hand, there are clear mechanisms by which people would get specific information on local service performance: people's personal experiences will be of local services, and those of their social networks are also more likely to be clustered around where they live. In addition, performance metrics are measured locally, and along with other heuristics that could indicate poor performance (e.g. ambulances queueing at A&E or a hospital going into special measures), they can be noticed and reported in local media.

On the other hand, as discussed elsewhere in this thesis, personal contact with services is a relatively infrequent event. Indirect information is also limited, because the reach and resources of local media in England are much less than those of the national media.

From the moderated learning perspective, one would expect people' priors to be modified only by significant disparities between local and national performance for three reasons. Firstly, because in this theoretical framework, people (and especially partisans) are resistant to change; secondly, because in an information environment weighted heavily towards national news rather than local news, people's views would be disproportionately affected by national performance signals; and finally, because if differences between local and national performance are small, voters may believe that the difference in signals results from random variation rather

than an innate objective difference in performance. In addition, as the data in this chapter will show, long-term national differences in NHS performance are much larger than geographical differences within the same time period. This can be summarised in two further hypotheses:

> Secondary Hypothesis 3: variation in local performance will be positively associated with better public evaluations with the NHS.

> Secondary Hypothesis 4: NHS evaluations will be more strongly related to variation in national over-time performance than geographical performance within England.

For completeness, I will also test whether the association between variation in local performance and NHS evaluations holds within all partisan groups. This is the geographical equivalent of Secondary Hypothesis 2.

> Secondary Hypothesis 5: variation in local performance will be positively associated with better public evaluations with the NHS, within all partisan groups.

6.3 Data sources and measures

6.3.1 Public opinion

In order to measure any possible over-time effects of changes in national NHS performance, I will return to the British Social Attitudes (BSA) cross-sectional survey series data previously used in the chapter on partisanship. The series starts in 1983 and asks about NHS satisfaction as well as a set of political and demographic variables.

The BSA is appropriate for this task for several reasons. Firstly, in order to have maximum statistical power and generalisability of my results, I am aiming to make the widest possible comparison over time; when combined with the performance data, the BSA will give me three decades of annual responses, much more than was possible with the British Election Study or other newer surveys. Secondly, the BSA is collected annually. As this chapter will show,

there is considerable seasonal variation in NHS performance within a year, so to minimise bias when estimating over-time variation, an annual snapshot of opinion over many years is more helpful than a more regular survey over a shorter period. Finally, year-on-year differences in aggregate public opinion are often small, so having a longer time period increases the overall variation in satisfaction available to my analyses.

For the over-time analysis, therefore, I will use survey data for each year that has both waiting times and NHS satisfaction data before the COVID-19 pandemic: this is every year from 1988 to 2019 except for 1992 and 1994. The number of respondents in each year is shown in Figure 6.1.

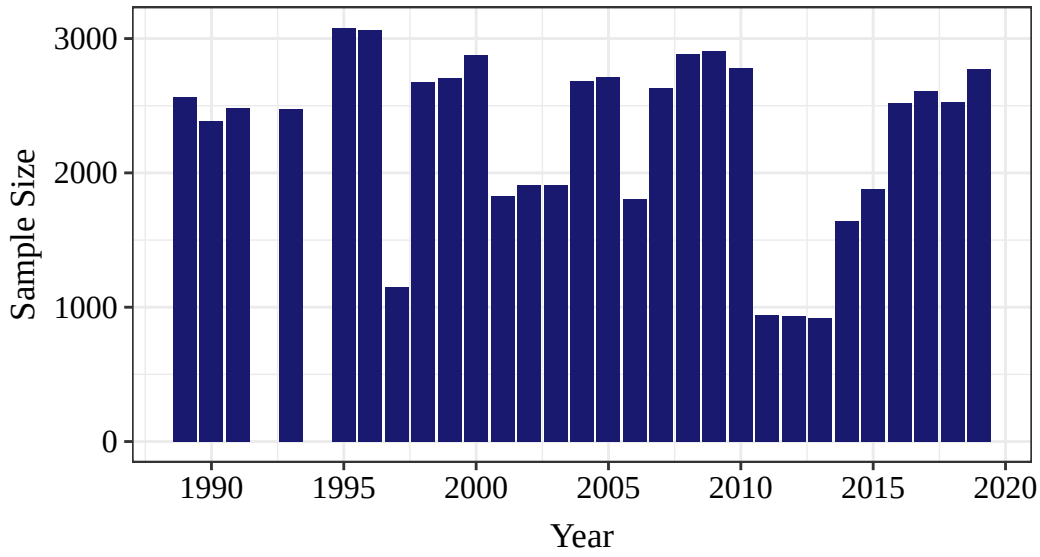

Figure 6.1: Sample size (number of respondents) per year from the BSA survey

To measure differences in average satisfaction in different geographical areas, I will take data from waves 10-18 of the British Election Study Internet Panel, which took place between 2016 and 2019 (Fieldhouse, Green, Evans, *et al.* 2019). I chose these waves because before wave 10, the data does not contain the respondent's Middle Super Output Area - which is a small geographical area, a subsection of a local authority, usually containing 5,000 - 15,000 people (Office for National Statistics 2022), so a detailed geographical comparison would not be possible; and wave 18 is the final pre-pandemic survey wave.

Table 6.1: Sample size (number of respondents) in the British Election Study Internet Panel by wave and year

Survey wave	Calendar year	Respondents
10	2016	20,600
11	2017	21,142
12	2017	24,013
14	2018	21,124
15	2019	21,188
16	2019	27,283
17	2019	24,751
18	2019	6,904

In each wave, the panel is asked a range of questions on their political attitudes. Their responses are combined with demographic variables held by YouGov (who conduct the survey). The sample size for almost all of the waves is over 20,000 respondents, which means that a detailed analysis of geographical variation is possible using this dataset. However, its utility for over-time analysis of the current question is limited because questions about the NHS are only asked in the 2016-2019 period. The number of respondents in each wave, and the year each wave falls in, is shown in Table 6.1.

6.3.2 NHS performance data

NHS performance data was taken from official NHS England and Department of Health statistics. For Referral to Treatment (RTT) waiting times data, this was taken from monthly commissioner-based incomplete pathway waiting time statistics (NHS England 2023); information on historic pre-RTT waiting time metrics were taken from the March 2009/10 Inpatient Waiting Lists file held on the UK National Archives' copy of the Department of Health website (Department of Health 2012). The combination of these two metrics allows for the longest possible time series comparison of hospital waiting time performance; chapter 3 contains more details on these metrics in Section 3.2.1.

6.4 Research design (over-time variation)

In this section, I will explain the research design of the over-time national performance models. I will first describe the variables taken from the British Social Attitudes cross-sectional surveys and NHS waiting times data, then the model specifications. I will then show the results of the main models that examine the overall effect of national performance on satisfaction over time.

6.4.1 Variables

The over-time models aim to measure changes in evaluations of the NHS from year to year. The response variable to operationalise these evaluations is the British Social Attitudes (BSA) survey series' NHS satisfaction variable that was previously used in chapter 5. Details for the predictor are shown in Table 6.2.

Table 6.2: Details of the NHS satisfaction response variable for the over-time performance models.

Variable	Variable name and source	Question wording	Possible values
NHS satisfaction	nhssat from British Social Attitudes (BSA)	All in all, how satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays?"	"Very satisfied", "Quite satisfied", "Neither satisfied nor dissatisfied", "Quite dissatisfied", "Very dissatisfied" and "Don't Know" (treated as missing)

The predictor, which is used as a proxy for long-term NHS performance in general, is a measure of average national waiting times. Figure 6.2 shows the over-time trend in the modern and

legacy versions of this metric, previously produced in Chapter 3.

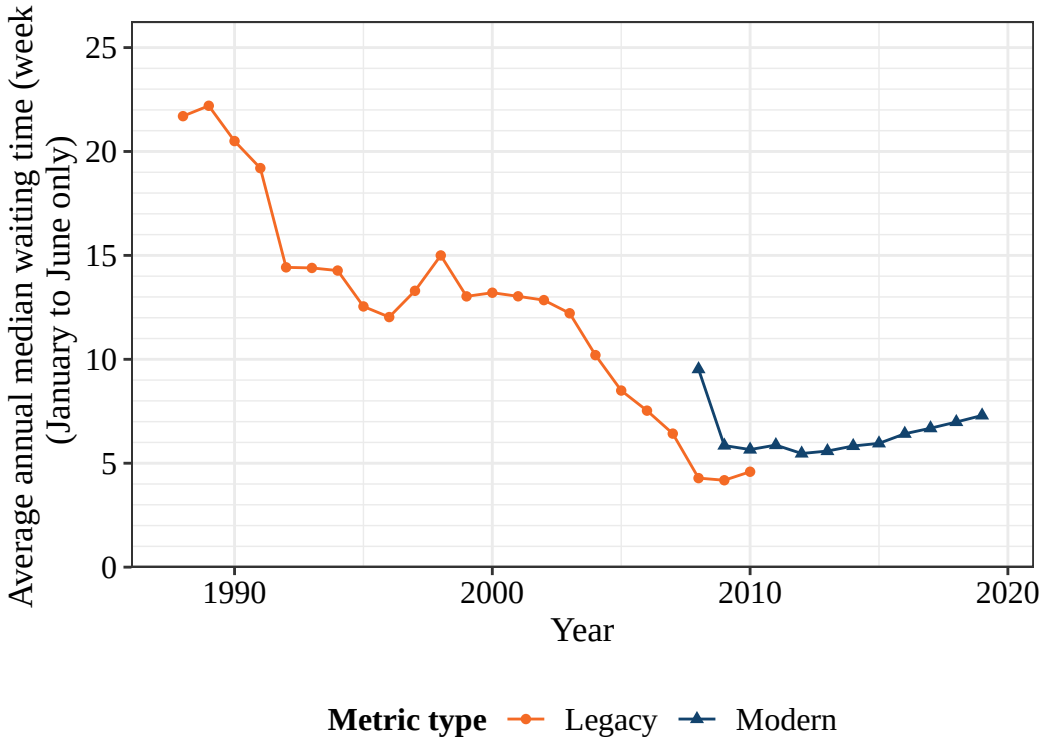

Figure 6.2: A comparison of median waiting times between the modern and legacy metrics, measured in weeks. The legacy metric spans 1988 through 2010 and generally fell over time, whereas the modern metric begins in 2008 and has generally been stable since 2010.

As described in that chapter, this metric takes the median waiting time in weeks for the first 6 months of the calendar year in which the survey was taken and summarises these as an average. This therefore gives an idea of typical waiting times in the first half of each year; the BSA fieldwork dates differ from year to year, but usually start in June or July and finish in November; taking the January-June window therefore gives a somewhat recent view of NHS performance that includes data from the previous winter. See Table 6.3 and Chapter 3 for more details.

Table 6.3: Details of the average national waiting times predictor variable used in the over-time performance models.

Role	Variable	Variable name and source	Question wording/ definition	Possible values
Predictor	Average national waiting times	Average (median) waiting time in weeks, from NHS waiting times data	Mean of the median national waiting time in weeks for the first 6 months of the calendar year in which the survey was taken	Numeric representing number of weeks

As with the previous set of models, I again use demographic controls in some model specifications. These control for differences in gender, age, marital status, socio-economic group, region, partisanship and ethnicity. Details of how these controls are constructed are given in Table 6.4. The control variables reduce the risk that other differences between years (e.g. the sample having a different age makeup in one year than the next) could contaminate the estimated association, and allow a more precise estimate of the parameter. The self-described party identity variable will also be used as an interaction term in some models, as described below in the "Models" section.

Table 6.4: Description of the British Social Attitudes control variables used in the over-time models

Variable	Variable name and source	Question wording/ definition	Possible values
Female	gender from BSA	Transformed into dummy variable	1 = female, 0 = male
Age group	age from BSA, grouped	Grouped into five similar-sized categories	17-32, 33-42, 43-54, 5-67, 68+
Socio-economic group	seg from BSA		Professional/employers/managers, Intermediate non-manual, Junior non-manual, Supervisor/skilled manual, Semi-skilled manual, Unskilled manual, other
Marital status	marstat from BSA		Not married, Widowed, Married, Living as married, Separated or divorced
Region	gor from BSA		North East, North West, Yorkshire and the Humber, West Midlands, East Midlands, East of England, South West, Greater London, South East

Variable	Variable name and source	Question wording/ definition	Possible values
Self-described party identity	partyid2 from BSA,	Re-grouped into fewer categories	Conservative, Labour, Lib Dem/Alliance, None, Other
Ethnic minority status	ethnicity from BSA,	Dichotomised from original variable	1 = ethnic group other than White, 0 = White

6.4.2 Models

This chapter aims to answer the research question: "Does variation in public service performance affect voters' evaluations of government performance?". The over-time models are designed to look specifically at differences in healthcare performance from year to year.

In order to test Secondary Hypothesis 1 (*"variation in national performance over time will be positively associated with variation in NHS satisfaction"*), I estimate a set of regression models that measure the association between the national average median waiting time for hospital treatment (between January and June in each year) and satisfaction with the NHS.

I first estimate a set of simple bivariate linear regression models that estimate the association between median waiting time in weeks and satisfaction with the NHS. Separate models are estimated for the legacy (pre-2010), modern (post-2008) and indexed waiting time metrics. These models will give a first look at the overall relationship between the variables without accounting for other potential influences.

I then estimate a set of "full" multivariate linear regression models that include the demographic control variables described earlier (gender, age group marital status, socio-economic group, region, self-described party identity and ethnic minority status) for the legacy, modern and indexed waiting time metrics . These models will give a more accurate view of the independent association between waiting times and NHS satisfaction by filtering out the in-

fluence of nuisance variables and therefore accounting for differences in sample composition over time.

As the graphs earlier showed, there was considerable variation in waiting times during the 1988-2019 period. I test whether there is a statistically significant association between the average median waiting time in a year and average satisfaction with the NHS. In years with longer waiting times, the satisfaction levels predicted by the model should be lower than in years with shorter waiting times.

I will then test Secondary Hypothesis 2 (*"variation in national performance will be positively associated with better NHS satisfaction, within all party identity groups"*). To do this, I will repeat the full multivariate models, but interact the party identity variable with the "waiting times" predictor. This will allow the effect of waiting times to differ by party identity (Conservative, Labour, Lib Dem/Alliance, None and Other). By comparing group-level predictions from this model, we will be able to see whether a statistically significant relationship between waiting times and NHS evaluations exists within each party group.

6.5 Results (over-time variation)

6.5.1 Is national over-time hospital waiting time performance associated with NHS satisfaction?

This section presents the results of regression models that estimate the association between hospital waiting times and public satisfaction with the NHS.

For a descriptive view of the relationship between satisfaction and waiting times, Figure 6.3 compares the indexed waiting time metric against average satisfaction. In the panel showing average waiting times, the y axis is inverted to make comparison with NHS satisfaction easier, since the relationship is posited to be a negative one. Note that unlike the regression model results, this visualisation does not account for the influence of other factors (e.g. changes in the sample's age profile over time) because it is a raw average, not control-adjusted output from

a regression model. To allow comparison across this time period, the indexed waiting times metric is used.

The two time series are fairly similar. In the 1990s, both waiting times and satisfaction improved, though improvement slowed in 1995 and waiting times lengthened between 1996 and 1998, which was accompanied by a deep fall in mean satisfaction. Between 2000 and and 2004 both waiting times and satisfaction were approximately flat, and until 2010 both improved significantly and rapidly. From that point on, whereas the indexed waiting time metric saw a small but consistent worsening trend, average satisfaction was erratic from year to year, with a general downward trend across the 2010s. In general, then, trends were very similar between performance and satisfaction, albeit with some more volatility and variation in average NHS satisfaction.

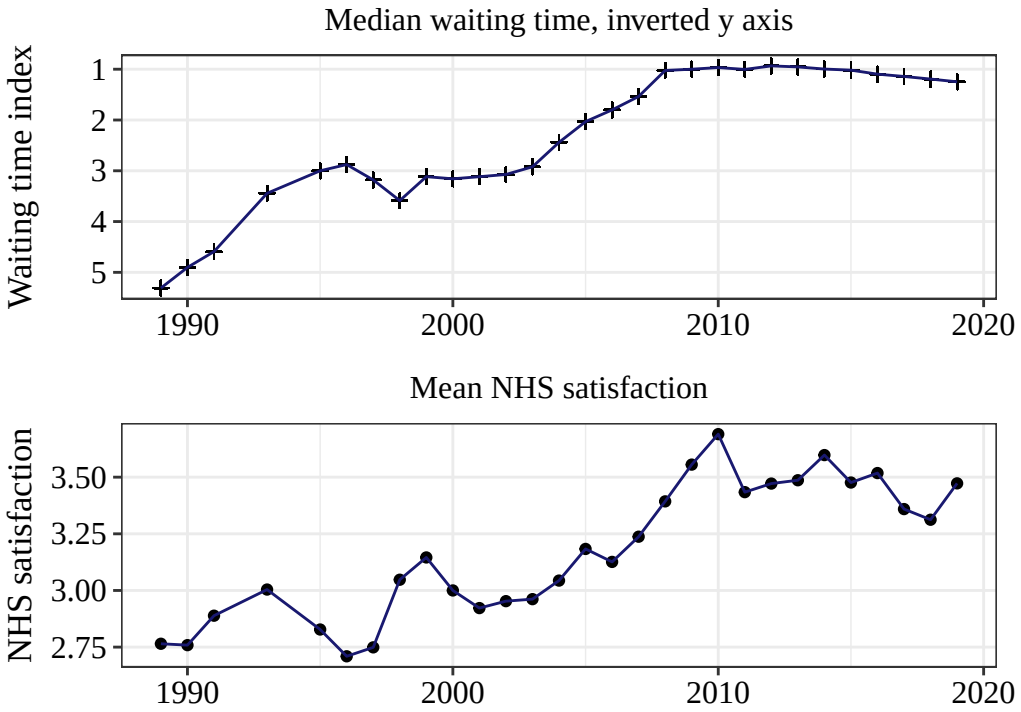

Figure 6.3: Comparison of median waiting times (from the indexed metric) and mean satisfaction from British Social Attitudes survey in each year

Table 6.5 shows the ouput summary from the bivariate linear regression model using the wait-

Table 6.5: Bivariate linear regression results showing that longer median waiting times in a given year are associated with lower satisfaction (BSA data)

	Legacy metric	Modern metric	Indexed metric
(Intercept)	3.566***	3.819***	3.623***
	[3.539, 3.593]	[3.738, 3.901]	[3.604, 3.642]
Median waiting times (weeks)	−0.043***	−0.051***	
	[−0.045, −0.041]	[−0.063, −0.039]	
Waiting times (indexed to 2009)			−0.194***
			[−0.201, −0.187]
Num.Obs.	49 503	25 315	66 245

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ing time metrics (legacy, modern and indexed) and satisfaction data from the British Social Attitudes surveys. Coefficients and 95% confidence intervals are also provided. Table 6.6 is the equivalent for the multivariate models.

In the simple bivariate models, longer waiting times are associated with slightly lower satisfaction. The coefficients for waiting times in the bivariate models suggest that each additional week of average waiting time reduces satisfaction by around 4-5% of a step on the Likert scale, depending on the time period.

Because of the possibility of bias, it is the full models with controls that I will use for the main part of this analysis, rather than the simple bivariate models. The models include controls for gender, age category, marital status, socio-economic group, region, ethnic minority status and party identity. However, as Table 6.6 shows, the inclusion of controls changes the results only slightly.

The multivariate regression results support Secondary Hypothesis 1: *variation in national performance over time will be positively associated with NHS satisfaction.* Average satisfaction was higher in years when waiting times performance was better, and lower in years when waiting time performance was worse.

As with the bivariate models, the coefficients suggest a one-week increase in median waiting times is associated with a reduction of 4-5% of a step on the Likert scale for NHS satisfaction;

Table 6.6: Multivariate linear regression results showing that longer median waiting times in a given year are associated with lower satisfaction (BSA data)

	Legacy metric	Modern metric	Indexed metric
(Intercept)	3.370 [3.311, 3.429]	3.798 [3.693, 3.904]	3.505 [3.458, 3.553]
Median waiting times (weeks)	−0.044 [−0.046, −0.042]	−0.053 [−0.066, −0.041]	
Median waiting times (2009-indexed)			−0.200 [−0.208, −0.192]
Female	−0.026 [−0.051, −0.002]	−0.062 [−0.094, −0.030]	−0.035 [−0.056, −0.014]
Age category: 33-42	−0.012 [−0.048, 0.023]	−0.011 [−0.062, 0.041]	−0.014 [−0.046, 0.017]
Age category: 43-54	0.010 [−0.027, 0.047]	−0.068 [−0.118, −0.017]	−0.013 [−0.045, 0.018]
Age category: 55-67	0.111 [0.073, 0.150]	−0.030 [−0.083, 0.022]	0.069 [0.036, 0.102]
Age category: 68+	0.440 [0.398, 0.482]	0.212 [0.155, 0.268]	0.369 [0.333, 0.405]
Marital status: Widowed	−0.013 [−0.063, 0.036]	0.075 [0.010, 0.140]	0.011 [−0.032, 0.053]
Marital status: Married	0.018 [−0.015, 0.051]	0.098 [0.055, 0.140]	0.035 [0.007, 0.063]
Marital status: Living as married	−0.028 [−0.075, 0.019]	−0.002 [−0.059, 0.055]	−0.021 [−0.060, 0.018]
Marital status: Separated or divorced	−0.019 [−0.062, 0.024]	0.010 [−0.044, 0.063]	−0.023 [−0.059, 0.013]
SEG: Intermediate non-manual	0.006 [−0.030, 0.043]	0.034 [−0.010, 0.078]	0.014 [−0.016, 0.044]
SEG: Junior non-manual	0.058 [0.021, 0.095]	0.067 [0.016, 0.118]	0.058 [0.026, 0.090]
SEG: Supervisor/skilled manual	0.029 [−0.007, 0.064]	0.063 [0.015, 0.111]	0.044 [0.013, 0.075]
SEG: Semi-skilled manual	0.083 [0.045, 0.121]	0.083 [0.034, 0.132]	0.088 [0.056, 0.120]
SEG: Other	0.167 [0.024, 0.309]	−0.058 [−0.481, 0.365]	0.168 [0.031, 0.305]
SEG: Unskilled manual	0.131 [0.077, 0.185]	0.103 [0.020, 0.187]	0.130 [0.083, 0.178]
North East	0.264 [0.210, 0.317]	0.071 [0.000, 0.142]	0.198 [0.153, 0.244]
North West	0.202 [0.158, 0.246]	−0.002 [−0.057, 0.052]	0.124 [0.087, 0.161]
West Midlands	0.164 [0.117, 0.210]	−0.025 [−0.085, 0.034]	0.094 [0.055, 0.133]
Yorkshire and the Humber	0.219 [0.172, 0.265]	0.003 [−0.056, 0.062]	0.139 [0.100, 0.178]
East Midlands	0.230 [0.182, 0.278]	0.040 [−0.020, 0.099]	0.164 [0.124, 0.204]
East of England	0.170 [0.120, 0.220]	−0.054 [−0.112, 0.003]	0.099 [0.059, 0.140]
South West	0.274 [0.227, 0.321]	0.081 [0.023, 0.139]	0.206 [0.167, 0.245]
South East	0.115 [0.074, 0.156]	−0.020 [−0.075, 0.036]	0.062 [0.027, 0.097]
Ethnic minority	0.005 [−0.042, 0.052]	−0.118 [−0.170, −0.067]	−0.040 [−0.078, −0.003]
Partisanship: Labour	−0.143 [−0.170, −0.115]	−0.027 [−0.065, 0.011]	−0.148 [−0.172, −0.124]
Partisanship: Lib Dem/Alliance	−0.162 [−0.199, −0.124]	0.038 [−0.018, 0.093]	−0.142 [−0.175, −0.108]
Partisanship: None	−0.140 [−0.178, −0.102]	−0.143 [−0.190, −0.096]	−0.160 [−0.192, −0.128]
Num.Obs.	46 625	24 224	62 575

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and an increase of one index unit reduces satisfaction by 20% of a step. The difference between the longest and shortest values for this metric is just under 18 weeks in the legacy metric and 4 weeks in the modern one, so we would expect the difference in average satisfaction between worst and best years to be around 78% of a step on the Likert scale in the legacy dataset, and around 20% of a step on the Likert scale in the modern dataset.

The indexed metric suggests a difference of around 20% of a Likert step for a change of one index unit. Note that one index unit is approximately four weeks in the legacy metric, and almost six weeks in the new metric - so as expected, the coefficient for the indexed variable implies similar effect size to the original metrics.

We can use model-based predictions and contrasts to better understand the scale of the effect size. This shows that a standard deviation increase in median waiting time in the indexed metric is predicted to be associated with a reduction in NHS satisfaction of approximately 25% of a step on the 5-point Likert scale (95% CI 24% - 26%). The maximum difference based on the range of waiting times in the whole dataset is predicted to be associated with a reduction in NHS satisfaction of 88% of a step on the Likert scale (95% CI 84% - 90%). Based on this analysis, I would argue that national NHS waiting times are indeed associated with differences in satisfaction, and that the effect size is moderate.

Figure 6.4 shows the predicted satisfaction levels by waiting times for a hypothetical 'typical' respondent in the "full" multivariate models with controls. One line shows the results from the model that is based on the modern waiting time metric only, and the other for the legacy waiting time metric only.

As the graph shows, the models for the legacy and modern metrics both approximately agree on the extent to which changes in waiting times affect predicted satisfaction. This shows that while there are differences in overall satisfaction levels in different time periods, both the older and newer waiting time measures showed a similar level of public responsiveness to performance differences. Note that, as discussed earlier, there is a built-in difference between the two waiting time metrics, likely due to differences in when the 'clock' starts and stops in the different metrics.

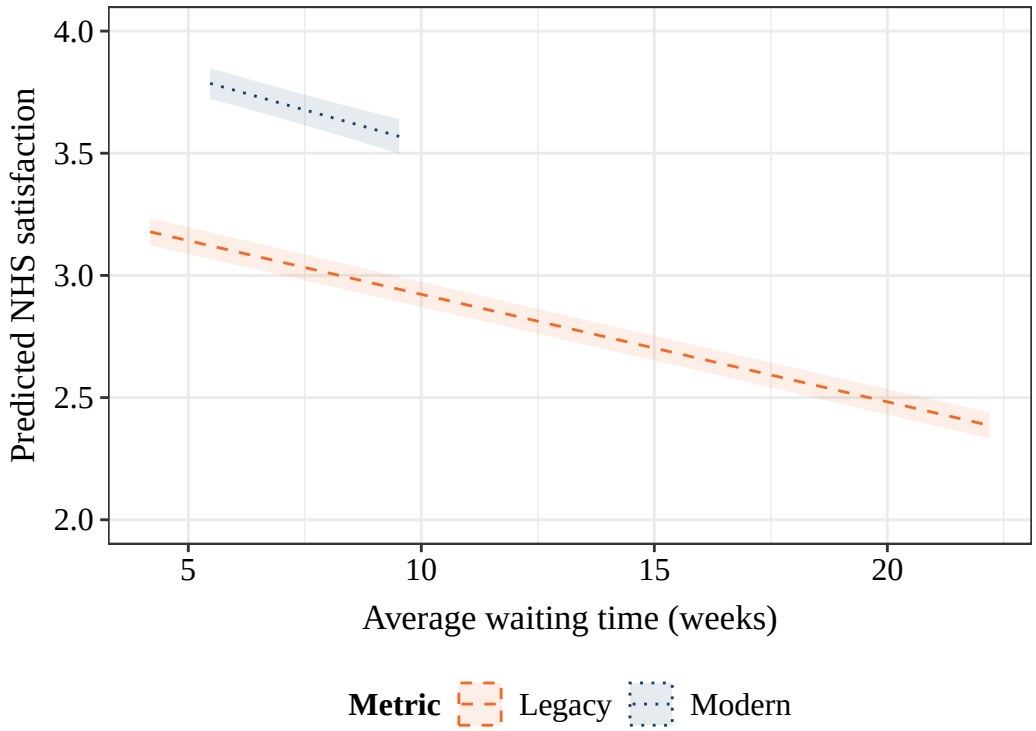

Figure 6.4: Predictions from the multivariate linear model, showing that longer waiting times were associated with lower predicted satisfaction with the NHS in both legacy and modern waiting time metrics (BSA data)

To get an overall view across the whole time period of the study, Figure 6.5 shows the predictions for a typical respondent from the full multivariate model using the 2009-indexed waiting time metric. Because the model uses indexed waiting time rather than weeks, it is on a different scale from the other models and needs to be shown separately.

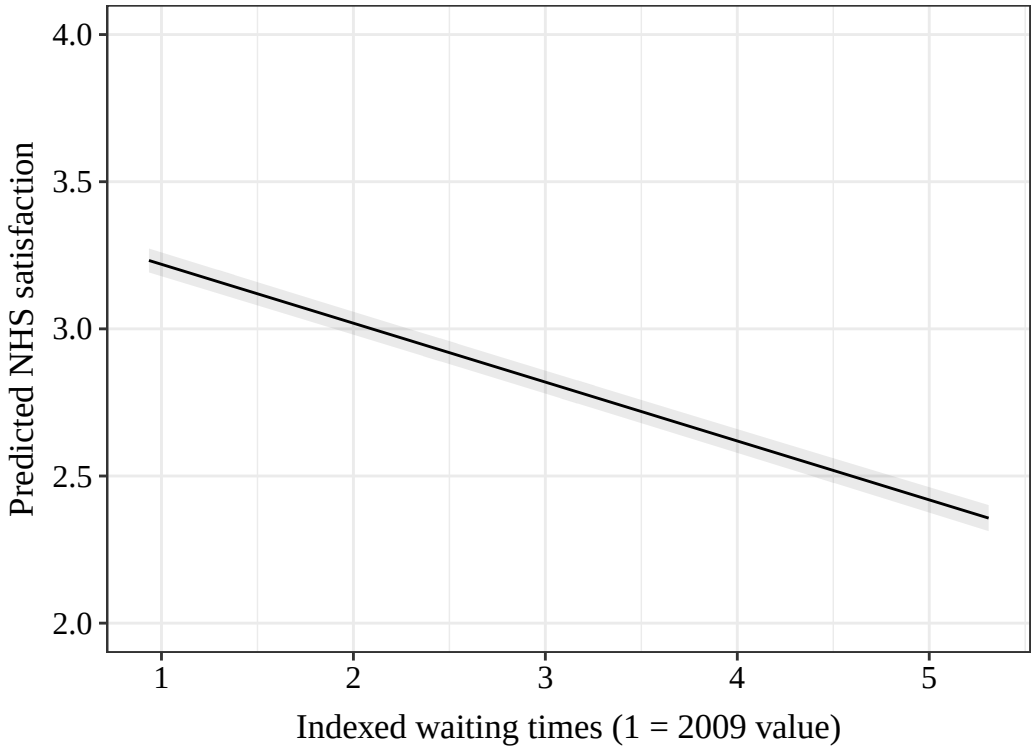

Figure 6.5: Predictions from the multivariate linear model, showing that longer waiting times (using the indexed metric) were associated with lower predicted satisfaction with the NHS (BSA data)

When considering the size of this difference, a possible yardstick is to consider it in comparison to the influence of partisanship. In chapter 5, I wrote that until 1998, the difference between Conservative and Labour identifiers was 50% of a step on the BSA Likert scale for NHS satisfaction. As the same variable is being measured here, I can compare relative associations from this model with chapter 5's partisanship model.

As Figure 6.5 shows, predicted NHS satisfaction for a typical individual changes by 20% of a step for each additional unit of indexed waiting times. Cross-referencing with the graph of average

waiting times by year in Figure 6.3 shows that there was an approximately one-unit change from 2000 to 2005, and another from 2005 to 2010, with the total range between best and worst performance in the period under study being around 4 units. This means that the predicted difference in satisfaction between 2000 and 2010 was slightly less than the typical difference in satisfaction between Conservative and Labour identifiers; and that the range between best year and worst year in performance terms result in a predicted satisfaction gap (~80% of a Likert step) that is bigger than a typical in-year partisan gap. This means that the influence of long-term changes in national waiting list performance on satisfaction with the NHS is similar to that of partisanship.

6.5.2 Does this association apply to all party identity groups?

We can make a basic initial assessment of this question by again graphing average waiting times and comparing this to average satisfaction; but this time, showing group average satisfaction for Conservative partisans, Labour partisans, Liberal Democrat/Alliance partisans, other partisans, and non-partisans separately.

The general trends of partisan satisfaction shown in the bottom panel of Figure 6.6 were previously discussed in chapter 5. As a brief summary, they show that government supporters are usually the most satisfied and opposition supporters are usually the least satisfied; and that there were wide partisan gaps until 1998, after which point these gaps reduced. The rapid improvement in waiting times shown in the upper panel between 2000 and 2010 was accompanied by a common upwards movement in satisfaction, whereas similar movement (albeit from a lower base) in the 1990s was not accompanied by a universal improvement in satisfaction. Since 2010, smaller partisan gaps have persisted and overall satisfaction levels have varied relatively little, in line with near-flat changes in waiting times. Overall, the graph provides visual evidence that long-run changes in waiting time performance are associated with changes in average satisfaction for all types of partisans; however, regression modelling is needed to confirm that this is statistically significant and to mitigate for potential confounding effects.

Figure 6.7 shows the group-average marginal effect of waiting times on satisfaction within each

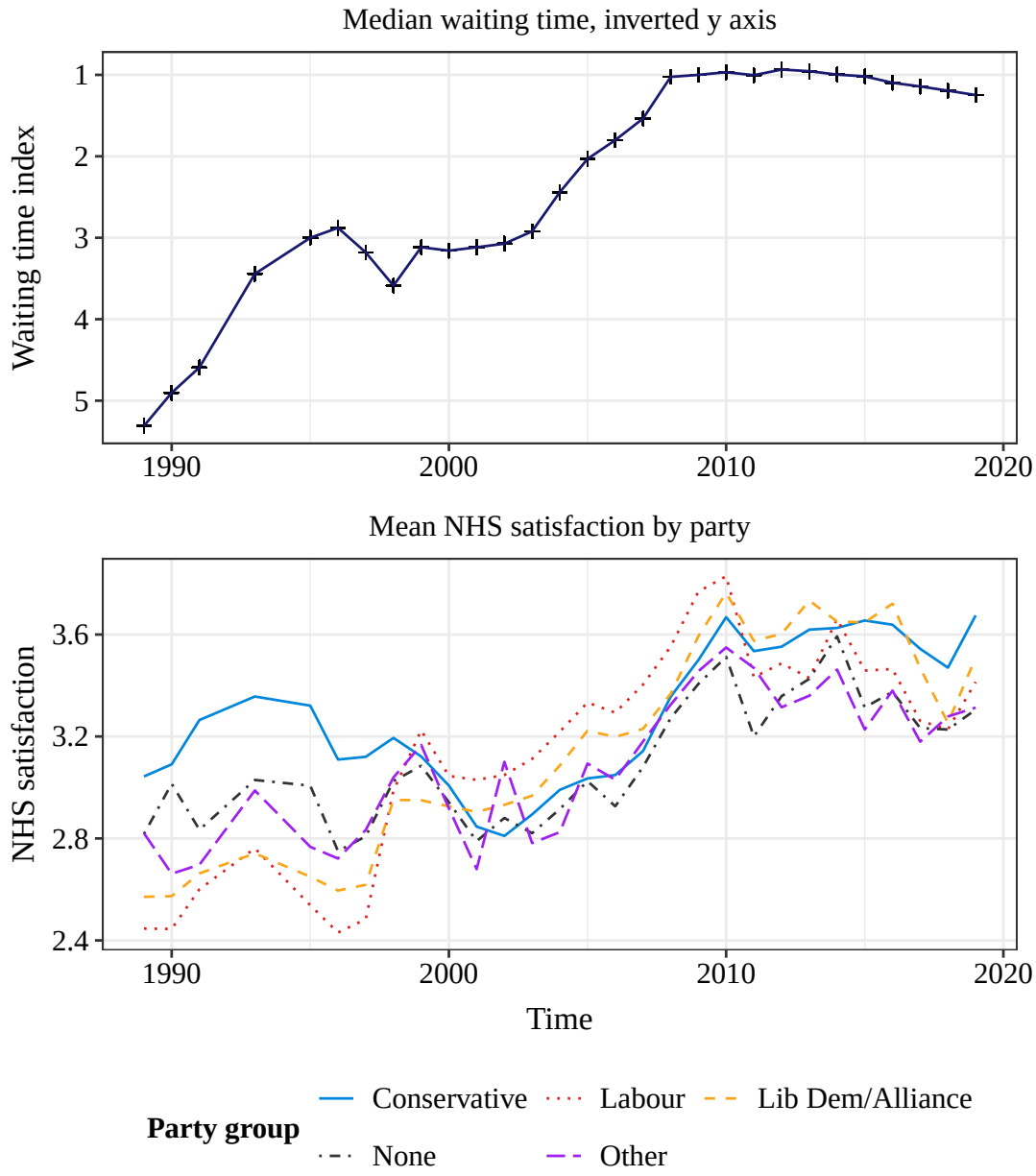

Figure 6.6: Line graph of mean NHS satisfaction from the British Social Attitudes survey data, showing how satisfaction varies by party and median waiting times

party identity group, for each of the legacy, modern and indexed metrics. Estimates come from the "full" multivariate regression models, modified so that party identity is interacted with waiting times. A negative marginal effect indicates that longer waiting times result in reduced average satisfaction. These marginal effect estimates can be interpreted as being analogous to a regression coefficient, if the model only contained respondents from that particular partisan group.

In Figure 6.7, for all partisan groups in all three types of waiting time metric, the estimates are negative. There is some heterogeneity in the effect size for different partisan groups, but there is consistency in the sign and statistical significance of the effect. This means that, while there is some variation in the size of the effect between metrics and groups, NHS satisfaction was lower in years with worse waiting time performance for respondents of each political party identity group. This supports Secondary Hypothesis 2: *variation in national performance will be positively associated with better NHS satisfaction, within all party identity groups.* Responsiveness to changes in performance is therefore a cross-partisan phenomenon, and not driven by any one partisan subset of the public.

The results of this study suggest that long-run changes in NHS performance and their influence on public opinion about the NHS in England are important, because, despite the influence of partisanship and the scarcity of performance signals, people of all partisan groups are able to recognise and respond to long-term changes in national waiting times. This implies that national performance and individuals' partisanship are of similar importance in understanding public opinion about the NHS.

6.6 Research design (geographical variation)

Having explored over-time variation in national performance and its relationship with satisfaction, I will now move on to the set of tests for geographical variation. In this section, I describe the British Election Study and NHS performance variables that I have chosen to use

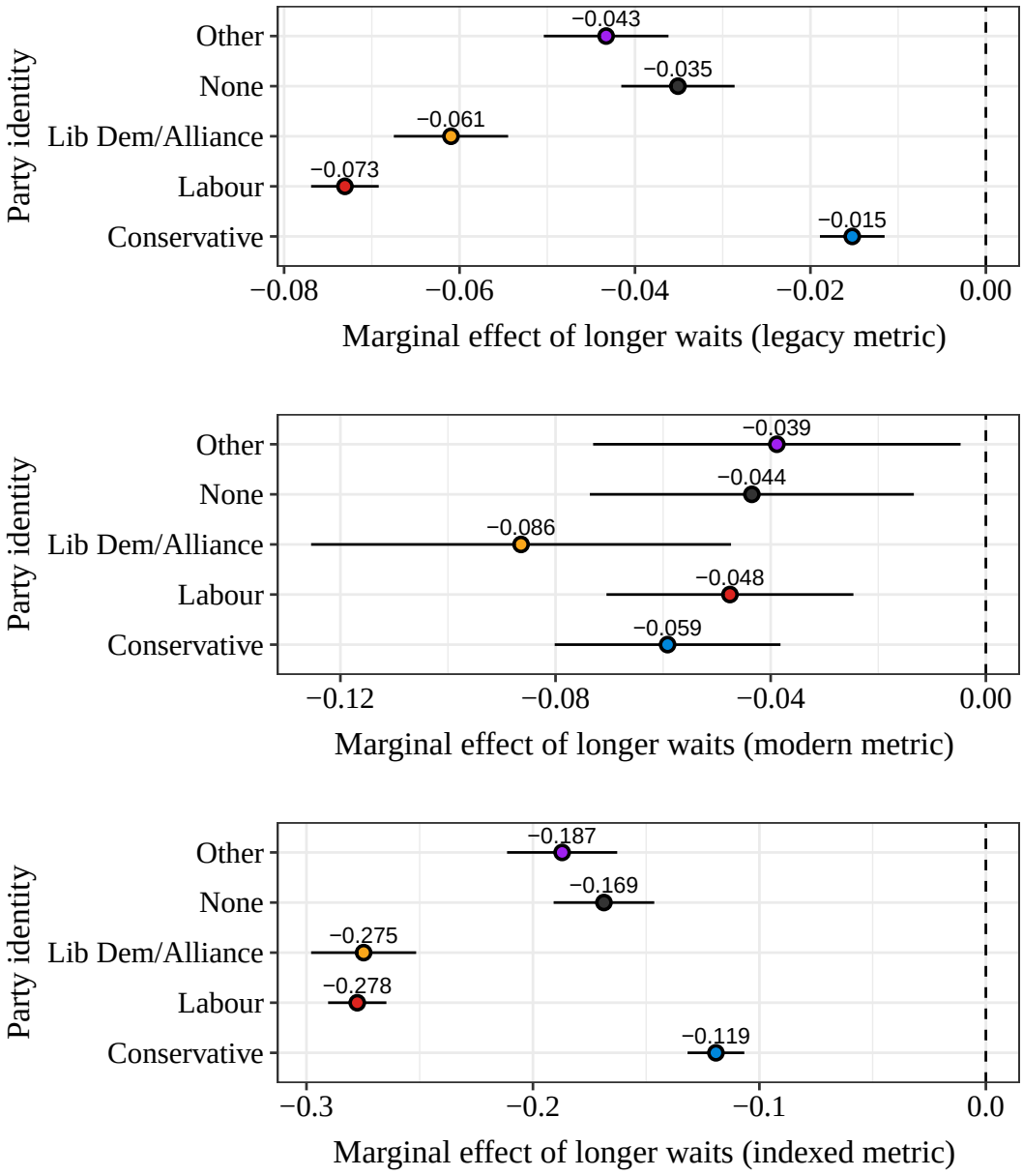

Figure 6.7: Marginal effect of waiting times from 'full' multivariate regression models, by party identity and metric (BSA data)

in modelling the 'geographical variation' models. I will then explain how each model will be specified and which assumptions they will test.

Following this section, I will show the results from these models. I will then turn to the over-time national performance models and describe their research design, since the data sources and model specifications are different.

6.6.1 Variables

The British Election Study asks its internet panel in some waves about whether the NHS is improving or getting worse. The response is used as the response variable in these models to represent voters' evaluations of the health service. I code the responses as numeric on a 1-5 scale, with 1 being "Getting a lot worse" and 5 being "Getting a lot better". Full details of the response options and question wording are given in Table 6.7. This question, while not asking explicitly about satisfaction, asks about a slightly earlier part of the chain of actions needed for accountability: it seeks to determine what judgement, if any, the respondent has about the relative change in recent performance in the NHS. From a retrospective voting perspective, we would expect positive change to be recognised and rewarded; in a moderated learning theoretical framework, we would expect that voters may not update their views in the absence of consistent, clear performance signals.

Table 6.7: Details of the response variable for the geographical variation models

Variable	Variable name and source	Question wording	Possible values
Evaluation of NHS change	changeNHS	"Do you think that each of the following are getting better, getting worse or staying about the same?": "The NHS"	"Getting a lot worse", "Getting a little worse", "Staying about the same", "Getting a little better", "Getting a lot better", "Don't Know". "Don't Know" coded as missing.

As this study measures local performance, the key predictor variable for this section of the study is median waiting times for inpatient treatment, reported at the Clinical Commissioning Group (CCG) level and measured in weeks. An example of variation in this metric is shown in Figure 3.4, and further details of the metric's construction can be found in Section 3.3.2.

Although the British Election Study data does not map respondents to Clinical Commissioning Groups, it does does report their Middle Super Output Area (MSOA). In most cases, an MSOA is wholly within a CCG; MSOA boundaries are drawn within local authority boundaries, and CCG boundaries were drawn to be mostly compatible with local authority geographies. In Figure 6.8, solid blue boundaries are CCG boundaries as of April 2018 (Office for National Statistics 2023a) and dotted orange ones are MSOA boundaries as of April 2011 (Office for National Statistics 2023b).

To assign respondents to a CCG, the central point of each MSOA is found and assigned to whichever CCG it falls into. For a small number of respondents living on the edge of an MSOA, this may incorrectly assign them to a neighbouring CCG. However, since neighbouring CCGs

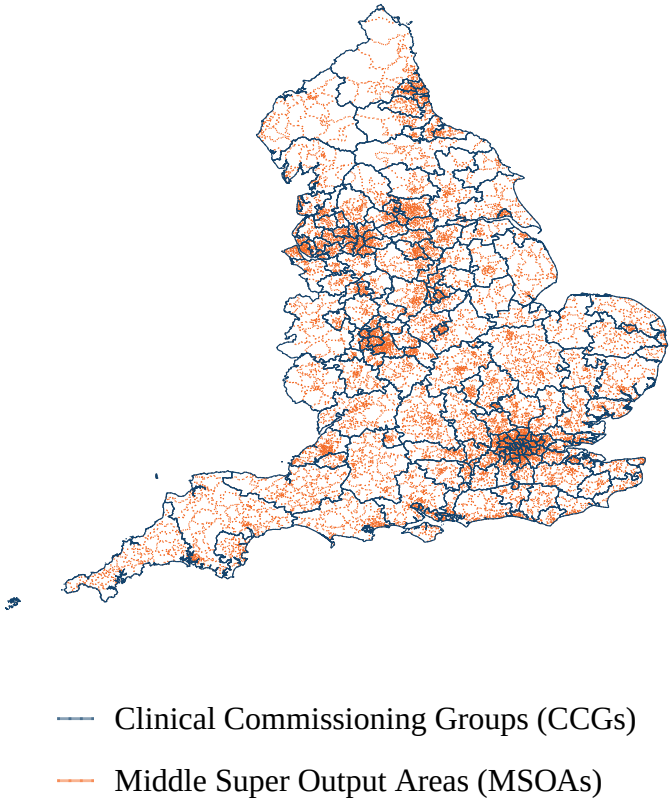

Figure 6.8: Map of Middle Super Output Area (MSOA) and Clinical Commissioning Group (CCG) boundaries

will in almost all cases share providers (e.g. hospitals) to a large extent, this should only introduce a small amount of error. In addition, since some CCGs merged or changed during the time period, I assign MSOAs to CCGs separately for each year based on the CCG boundaries as of April that year (Office for National Statistics 2023a). 1,178 observations (0.7% of the total) were dropped due to their MSOA not mapping to a CCG in the dataset.

Figure 6.9 shows MSOA centroids overlaid on CCG boundaries for 2018 as an example.

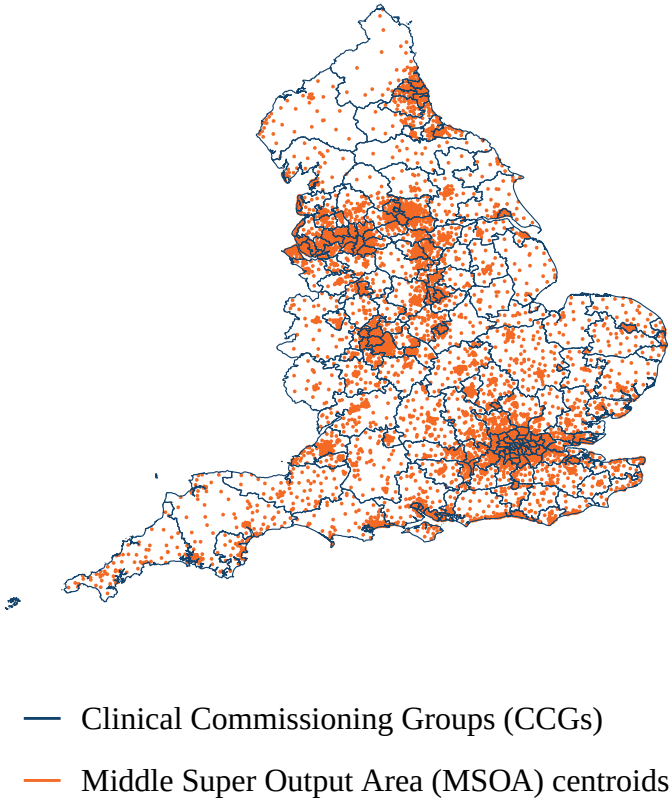

Figure 6.9: Map showing MSOA centroids within 2018 CCG boundaries

Different types of geographical area (e.g. post-industrial urban area, wealthy rural area) have different demographic and political profiles. This would not be problematic if NHS performance was randomly distributed throughout the country, but NHS performance is likely to be

correlated with demographics in multiple ways that are not necessarily obvious. For example, the average age is 6 years higher in rural areas, and older people are more satisfied with the NHS on average, which would lead to average evaluations in those areas being higher, all else being equal. On the other hand, healthcare performance is generally worse in rural areas because of isolation and underinvestment (Appleby & Roberts 2012; Coleman 2023) - so failure to control for age would lead to the model underestimating the strength of the correlation between performance and NHS evaluations.

To avoid this sort of confounding effect arising, the model controls for a set of political and demographic variables that are likely to be relevant to the relationship between performance and evaluations of the NHS. I control at the individual level, to account for the possibility that the panel contains a disproportionately high number of respondents from a particular group. Details of the control variables are shown in Table 6.8.

Table 6.8: Details of the British Election Study control variables for the geographic variation models.

Variable	Variable name and source	Question wording/ definition	Possible values
Female	gender from British Election Study (BES)	Transformed into dummy variable	1 = female, 0 = male
Age group	age from BES, grouped	Grouped into five similar-sized categories	17-37, 38-49, 50-60, 61-68, 69-85
Region	gor from BES		North East, North West, Yorkshire and the Humber, West Midlands, East Midlands, Eastern, South West, Greater London, South East

Variable	Variable name and source	Question wording/ definition	Possible values
Self-described party identity	partyid2 from BES,	Re-grouped into fewer categories	Conservative, Labour, Liberal Democrat, None, Other
Ethnicity	ethnicity from BES,	Re-grouped into fewer categories	White British, Other White background, Ethnic minority

The BES internet panel data includes variables for party self-identification, gender, region and ethnicity; these are the variables that will be used as controls in some of the model specifications (more detail below).

6.6.2 Models

To estimate the effect of local NHS performance on individuals' evaluations of the NHS, regression models will then be estimated that measure the association between the median waiting time for inpatient treatment at CCG level, and the individual's evaluation of whether the NHS is getting better or worse.

The first model is a simple bivariate linear regression model showing the association between median waiting time and NHS evaluations. One potential concern with using this data is the fact that there are repeated measures of the same individuals across different survey waves, and linear regression assumes independence of observations. I deal with this concern by *only including the first observation from any individual respondent.* This means that the data is no longer in a panel format .but a cross-section with unique observations.

The second model allows repeated measures of the same individuals, but adds in controls for differences that could act as 'nuisance variables' by exerting influence on the outcome while being correlated with the key predictor (because they may vary geographically): age category,

party identity, ethnicity, gender and region. The survey wave is modelled as a random effect, in order to increase the precision of the estimates by partially pooling information across survey waves. As long as the results for this model are consistent with model 3 (i.e. repeated observations are not introducing bias into the results), I consider this model to be the most reliable, since the controls are needed to avoid bias, and adding wave as a random effect allows me to estimate the main effects with greater precision than if it was a fixed control variable.

The third and final model main model is identical to model 2 (the multi-level model with wave as a random effect), but as was the case with model 1 (the bivariate model), it only includes the earliest response from any individual. This provides another check on whether repeated measures of the same individuals are adding bias to the results.

In order to check whether the association between local performance and local NHS evaluations holds within each partisan identity group, I will then re-estimate the second model but add in an interaction term between median waiting time and partisan identity. This allows the effect of median waiting time to vary between different party identity groups and will therefore show whether Secondary Hypothesis 5 is supported by this data.

6.7 Results (geographical variation)

6.7.1 Effect of local performance on NHS evaluations

In order to test Secondary Hypothesis 3 (*"variation in local performance will be positively associated with better public evaluations with the NHS"*) and Secondary Hypothesis 4 (*"NHS evaluations will be more strongly related to variation in national over-time performance than geographical performance within England"*), I estimated the effect of local NHS performance on health service evaluations among people living in different Clinical Commissioning Groups in England. Three model specifications were used, as described in the previous section. Coefficients and 95% confidence intervals from these models are shown in Table 6.9.

Table 6.9: Regression outputs from local performance models, showing that longer CCG-level waiting times are associated with slightly lower satisfaction in the BES data

	1: Bivariate	2: Multi-level	3: Multi-level with no repeats
(Intercept)	1.851***	2.645***	2.614***
	[1.792, 1.910]	[2.563, 2.728]	[2.510, 2.717]
Median waiting times (weeks)	0.009*	−0.014***	−0.015***
	[0.000, 0.017]	[−0.020, −0.009]	[−0.024, −0.006]
Num.Obs.	56 920	165 827	56 920
Random effects:	None	Wave	Wave
Individual-level controls:	No	Yes	Yes
Repeat obs:	No	Yes	No

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As Table 6.9 shows, in the bivariate model with no repeat observations, the coefficient for waiting times is very close to zero (0.9% of a step on the Likert scale), with 95% confidence intervals of zero to 1.7% of a step on the Likert scale. It has the opposite sign than expected: this implies that there is a very small positive association between higher waiting times and better NHS evaluations. However, although the model is not estimated using repeat observations, the bivariate model does not account for geographical differences in other relevant variables that may also correlate with NHS evaluations.

When including controls for other relevant variables, the model results are very consistent. Model 2 produces a coefficient of -1.4% of one step on the 5-point Likert scale for the effect of an additional week of median waiting times (95% CI: -0.9% to -2.0%). Model 3 produces a coefficient of -1.5% (95% CI: - 0.6% to -2.4%). We can be confident that the statistical significance of the main effect is not due to repeat observations of the same individuals, because the results for model 3 are consistent with those for model 2. As predicted by Secondary Hypothesis 3, shorter local waiting times were associated with higher satisfaction.

One way to think about the size of the effect is to extract the predicted slope of the relationship between median waiting times and NHS evaluations for a hypothetical respondent. As the second model (full multivariate model with controls and wave random effects) is the one I consider most informative, I will use that one for this analysis.

Figure 6.10 shows model-based predictions for this relationship, for wave 18 (November - December 2019), which was the wave with the greatest range of waiting times across CCGs. All control variables are held at their median (for numeric variables) or mode (for categorical variables), meaning that these are predictions for a hypothetical White British Conservative woman aged between 18 and 37, living in the South East. The only explanatory variable that is altered is CCG-level median waiting times (represented on the X axis). This comparison gives us a prediction of how much a typical person's evaluations would differ if they were in the best-performing NHS area, compared to the worst-performing NHS area.

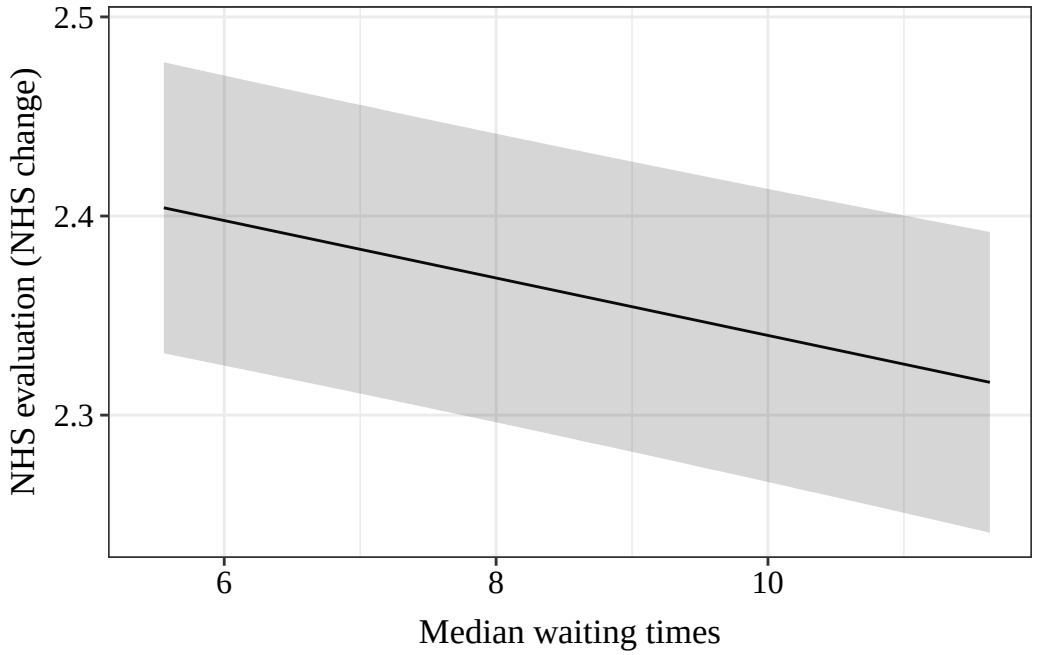

Figure 6.10: Predictions from the geographical regression model using BES data, showing predicted evaluations are lower where median CCG-level waiting times are longer

The average difference between the highest and lowest waiting times was associated with a difference in NHS evaluations of 10% of a step on the 5-point Likert scale (i.e. for whether the NHS is getting better or worse). The difference is statistically significant at the p = 0.05 level (95% CI: 7-14%).

The model also predicts that a standard-deviation difference in median waiting times (0.92 weeks) is associated with an average difference of 1.3% of a step on the Likert scale. This esti-

mated effect is very small compared to the standard deviation of the NHS evaluations variable, which is 0.90 (90% of one step).

These model results therefore support Secondary Hypothesis 3: *"variation in local performance will be positively associated with better public evaluations with the NHS"*. The relationship ran in the expected direction, so that shorter median waiting times were associated with slightly better evaluations. However, the effect size was very small: variation in performance between different local areas was only very slightly reflected in variation in public opinion. This implies that local differences in health service conditions are a measurable but not very important factor in determining people's views about the NHS.

This is not surprising, because the 'test' posed by this model was quite difficult, since it requires people to respond to local and not national conditions; the lack of performance signals about the local picture may have impaired the public's ability to make accurate judgements about local performance differences. Combining these results with the results from the over-time models shows that Secondary Hypothesis 4 (*"NHS evaluations will be more strongly related to variation in national over-time performance than geographical performance within England"*) is supported. A change of one standard deviation in over-time national waiting times was associated with a change of 0.25 Likert steps, whereas a difference of one standard deviation in local waiting times in the geographical models was associated with a change of 0.013 Likert steps.

6.7.2 Does this association apply to all party identity groups?

In order to test Secondary Hypothesis 5 (*'variation in local performance will be positively associated with better public evaluations with the NHS, within all partisan groups'*), I re-estimate Model 2 (multi-level model with wave random effects) with an interaction term between the predictor (CCG median waiting time) and the response variable (NHS evaluation). The association between local NHS performance and voter evaluations is therefore able to vary by partisan group, which means that we can determine whether all partisan groups were receptive to differences in performance signals.

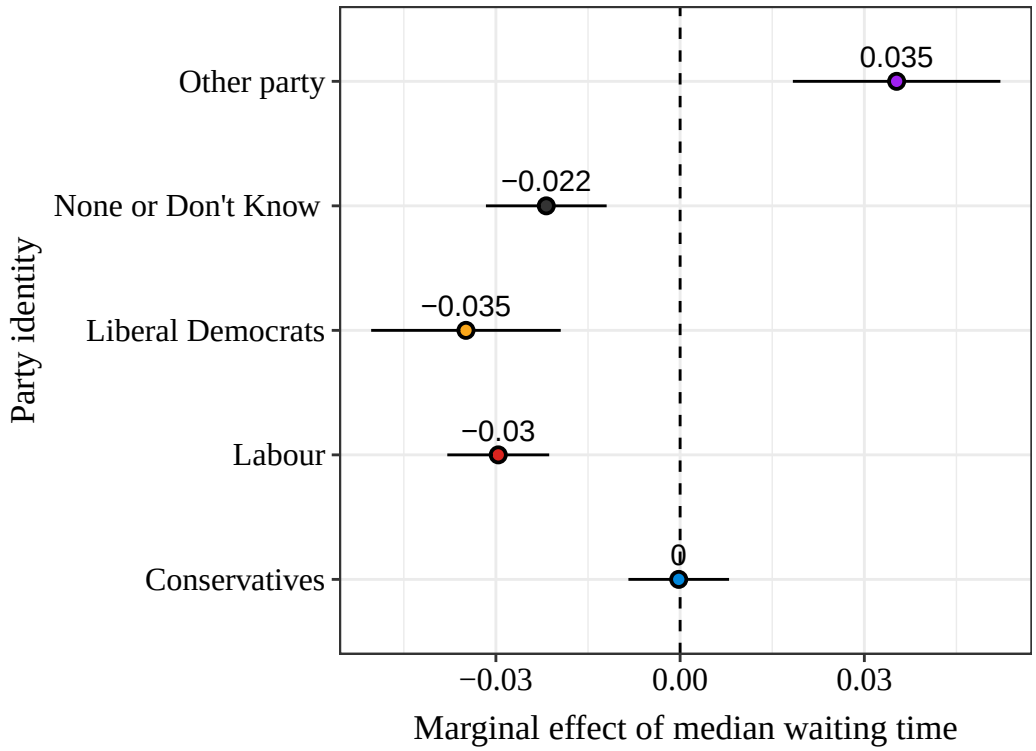

Figure 6.11: Group-average marginal effects of CCG-level waiting times, by party identity group. This shows that for most partisan groups, longer waiting times were associated with worse NHS evaluations in the BES data.

Figure 6.11 shows the group-average marginal effect of a one-week change in CCG-level median waiting times on NHS evaluations in the model. For Labour and Liberal Democrat identifiers, as well as for non-partisans, evaluations were slightly worse (around 3% of a step on the five-step Likert scale for NHS change) when waiting times were one week longer. However, the difference was not statistically significant for supporters of the governing Conservative party, and the coefficient was positive for supporters of other smaller parties.

The estimated effects for this model are very small and low-precision (noisy), but the lack of a statistically significant negative effect among Conservatives and minor party supporters means that we do not find evidence for Secondary Hypothesis 5: variation in local performance was not positively associated with better public evaluations with the NHS,within all partisan groups, although it was for identifiers of the main opposition parties.

6.7.3 Are we missing important information about variation in performance?

At this point we may challenge the model on the basis that it operationalises performance based on median hospital waiting times alone, and question whether an alternative measure of local NHS performance would be more strongly associated with public opinion.

One way to explore whether there is such a measure is to look at how much of the variation in people's opinions about the NHS differs by geography, without specifying a particular performance metric. This is possible by using 'empty' variance components models with random effects at different levels (e.g., for CCG or region) but without supplying any performance or demographic variables associated with them. In effect, this allows me to allocate respondents into their geographical areas without choosing a particular performance metric, as well as into the year of their response, and then quantify how much of the total variation in responses derives from differences *between respondents in the same area*, how much of it derives from differences *between geographical areas*, and how much derives from *differences between years*. If local performance (other than median waiting times) is strongly associated with differences

in average evaluations between local areas, we would expect to see the geographical grouping variables explaining a large proportion of variation.

Intraclass correlation coefficients can be extracted from these models and transformed into percentages: these percentages represent the proportion of the total variation that is explained, in a statistical sense, by each 'level' in the data (individual, local area, and time).

I estimate variance components models to test sources of variation at the Clinical Commissioning Group and Government Office Region levels[1]. Random effects for individual respondent ID, geographical area and survey wave were included in both models. The variance was then decomposed into intraclass correlation coefficients and segmented by random effect.

Figure 6.12 visualises the percentage of total variation explained by each level in the model.

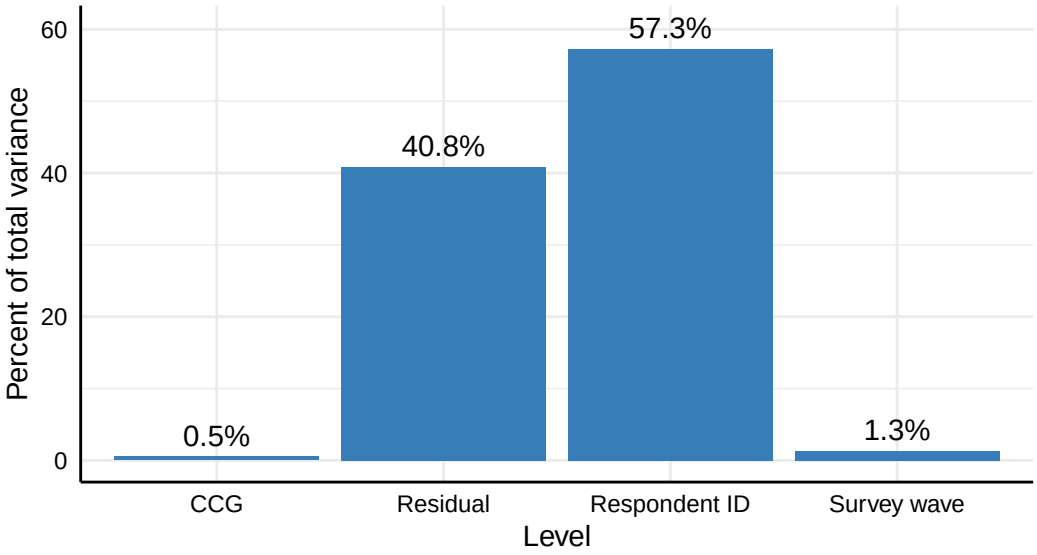

Figure 6.12: Intraclass correlation output plot showing that most variation in NHS evaluations is explained by individual-level factors, rather than at the CCG level or by survey wave

When partitioning the variation by CCG, 57.3% of the variance was at the individual level (between individuals). 1.3% was at the wave level (temporal variation), 0.5% was at the CCG level and 40.8% was not explained by the model. This strongly suggests that variation in NHS

[1]Because of changing CCG boundaries, this test only includes 2018 CCG assignments.

evaluations is not explained by differences in local performance across NHS geographies, but that individual-level factors have a major role, and that in the 2016-2019 period, differences over time between survey waves were minor.

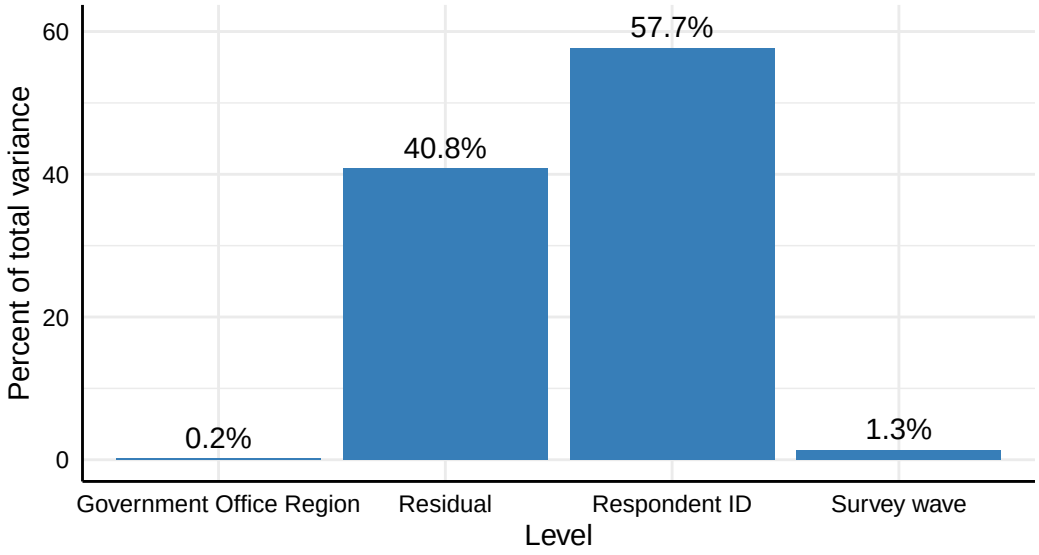

Figure 6.13: Intraclass correlation output plot showing that most variation in NHS evaluations is explained by individual-level factors, rather than at the region level or by survey wave (BES data)

When partitioning the variation by region, 57.7% of the variance was at the individual level (between individuals). 1.3% was at the wave level (temporal variation). 0.2% of variation was at region level and 40.8% was not explained by the model. This shows that as the geography gets wider, the explanatory value of geography gets even smaller.

In summary, the variance components models show that most of the variation in people's evaluations of the NHS in England is explained by individual-level factors (political, demographic and other differences), not by the area in which they live. This was consistent whether at a broad regional level or at Clinical Commissioning Group level. In addition, around a 40% of variation could not be explained by geographical, individual, or temporal factors; given the complexity of social phenomena and ascertaining what factors influence opinion change, this is not surprising. These results support the findings of the geographical variation models and reiterate their theoretical contribution: geographical differences in performance appear to only

be responsible for very small differences in public evaluations of the NHS.

6.8 Discussion

In this chapter, I aimed to test Primary Hypothesis III (*"voters' evaluations of the NHS will be better when service performance metrics are better")*., and help answer the overall research question of whether voters are able to perceive and evaluate variation in public service performance. I did this through a series of models that examined both over-time and geographical variation in performance. Evidence was found that supported all four secondary hypotheses, and therefore the overarching primary hypothesis: both in times and in places where performance metrics are better, voters evaluate the NHS more highly.

What are the implications of the findings of this chapter for the thesis as a whole? The key point is that voters evaluated government performance on the NHS more highly when waiting time performance was better — either compared to other time periods, or to other areas of the country. The electorate's ability to perceive and evaluate government performance is a fundamental, necessary first step that enables retrospective voting to take place and create government accountability.These models suggest that on the issue of healthcare, voters do learn from performance, despite the barriers and limitations to this process that we know exist.

In addition, the models show that this voter responsiveness, while not homogeneous throughout the electorate, was not confined to any particular partisan group. Supporters of both government and opposition parties displayed an association between NHS evaluations and waiting time performance metrics. I will now examine the results from each set of models in more detail and discuss their implications.

In the first set of models, over-time differences in satisfaction were shown to have a substantial association with NHS satisfaction. In years with shorter waiting times, people were more satisfied with the NHS, and in years with longer waiting times they were less satisfied. A standard-deviation change in waiting times was associated with an average of quarter-of-a-step change in satisfaction, which is similar to a typical Conservative/Labour partisan gap in satisfaction

in a particular year. The difference between worst and best waiting times in the whole series is predicted to change satisfaction by 88% of a Likert step for a typical respondent.

I would argue that these are significant associations with important theoretical implications. They are consistent with the idea that national conditions can produce an information environment that is strong enough to induce changes in opinion at the aggregate level.

When modifying the national models to examine the expectation that changes in performance can affect the opinions of people within all partisan groups, I found that, as expected, longer waiting times were associated with lower satisfaction across all partisan groups. This confirmed that the overall responsiveness demonstrated previously was not due to a small subgroup of responsive "floating voters", or supporters of one party, but that voters of any partisanship can have their priors changed by evidence of differences in national performance over time.

In the second set of models, variation in waiting times between different Clinical Commissioning Groups in England was associated with respondents' views about whether the NHS was improving or deteriorating. The magnitude of this association, however, was very small in the years under study. This is still compatible with the 'moderated learning' theory, which does not imply perfect or universal updating of evaluations; these findings are consistent with the existence of barriers and limitations that make it difficult for learning to take place in the absence of a consistent flow of credible performance information. Another element of the geographical results that points to limitations and contingency is the partisan group analysis. This found that, unlike in the national over-time models, an association between local performance and evaluations was not present in all partisan groups; this may indicate that the weaker performance signals associated with local differences did not significantly affect voters' priors in some subgroups of the electorate.

Bringing together the findings from the over-time and geographical analyses, public opinion was more strongly associated with national over-time performance in the three decades studied than geographical differences in performance within a single time period in the late 2010s. As discussed previously, this likely reflects a differential in influence between local and national

media. The difference is compatible with a moderated learning approach, since it suggests that while performance affects opinion, the extent to which it does so is dependent on the information environment, or in other words, the value and precision of the performance signals voters receive.

In summary, these findings strongly support both Primary Hypothesis III's expectation that voters respond to performance, and the overall theoretical expectation that both performance and partisanship play significant roles in influencing opinions about the NHS in England. They show that changes in performance over time are associated with differences in satisfaction with the NHS, both at an aggregate level and within all partisan sub-groups; and that there is also a weaker type of public responsiveness to geographical differences in performance. The influence of performance signals is therefore able to provoke a response in the electorate. This response may be modest in response to short-term, small variations in performance, but it is measurable and important: this type of public response is a fundamental and necessary first condition for the public to be able to hold government to account.

7 Does personal experience change voters' evaluations of health services?

7.1 Introduction

In the theory laid out in Chapter 4, I argued that voters balance both accuracy and directional goals when processing information, by taking account of signals that conflict with their priors, but assigning them less weight. For most people most of the time, their level of satisfaction with public services will be anchored by their partisan priors; however, new information in the form of performance signals can prompt people to update their views to the extent that the new information is judged sufficiently credible. I argued that the most credible and precise type of performance information would be personal experience of services - therefore, people are likely to change their views of services after a new period of contact with them. I then used repeated cross-section survey data from the British Social Attitudes series to find that there is a correlation between waiting times and satisfaction with the health service, suggesting that people are able to perceive and react to signals about service performance.

In this chapter, I aim to test Primary Hypothesis IV:

> PRIMARY HYPOTHESIS IV: RECENT SERVICE USE WILL AFFECT VOTERS' EVALUATIONS OF THE NHS

This chapter therefore tests the idea that personal contact reduces the influence of partisan priors on voters' evaluations of services by giving them credible signals about the government's performance. According to the theory laid out in chapter 4, voters' opinions about

public services are a combination of their partisan-informed priors and an aggregation of the performance signals they have received.

This allows us to make several assumptions: within a partisan group, if recent National Health Service (NHS) service users are more satisfied, this indicates that the performance signals indicated a higher level of service quality than their partisan priors suggested. The inverse holds if recent service users are less satisfied than their copartisans. If there is no difference between recent service users and others, this could indicate that the performance signals were consistent with their partisan priors; or, that this theory is incorrect, and recent contact is not associated with changes in public service satisfaction. This leads me to two secondary hypotheses:

> Secondary Hypothesis 1: people who have recent experience of health services will have a different level of satisfaction than those who do not.
>
> Secondary Hypothesis 2: partisans with recent experience of health services will have a different level of satisfaction than others within their partisan group who do not.

What does the theory predict about how these satisfaction levels will differ between partisan groups (government partisans, non-partisans and opposition partisans)? Because the underlying performance signals are not observable, it is only possible to draw conclusions about them based on differences between people who have more or less exposure to relevant information.

Recent personal experience of services may result in a reduction in differences in satisfaction between partisan groups in several ways. Firstly, if the signals imply a quality of service somewhere between the prior expectations of the partisan groups, both should move towards the signals, resulting in a narrowing of partisan differences. Secondly, if the signals are consistent with one partisan group's views but not the other's, then the group whose views are least consistent with the signals should alter their level of satisfaction in the direction of the signals.

However, if the signals imply a service quality that is higher or lower than the priors for both partisan groups simultaneously, it is possible that the partisan difference could remain static

because of both groups moving in parallel towards the signal ("parallel updating"). This depends on the balance between three factors: the credibility of new information pulling opinions towards it, the strength of the existing partisan prior, and the distance between the prior satisfaction level and the information.

I therefore argue that as a minimal expectation, recent contact should not be associated with a divergence in satisfaction between government and opposition partisans.

> Secondary Hypothesis 3: Differences in satisfaction between government and opposition partisans (the partisan gap) will not be greater among those with recent NHS contact.

What would this theory predict about non-partisans? Extending the minimal expectation in Hypothesis 3, additional information should not cause non-partisans to move in a different direction from both partisan groups.[1] As I will explain below, depending on the exact type of moderated learning that is taking place, we might expect non-partisans to update their views more slowly or more quickly (or at the same rate) as partisans; however, in the moderated learning framework the same information should not cause different partisan groups to revise their views away from each other.

> Secondary Hypothesis 4: Service contact should not result in non-partisans' satisfaction moving in the opposite direction from both other partisan groups.

In chapter 4, I described how the moderated learning framework does not impose specific expectations about non-partisans. From a simple 'partisan screen' perspective, one would expect non-partisans is to have weaker priors because they do not have a specific set of partisan beliefs to accommodate. This implies they should be more receptive to new information, and could react more quickly to changes in the information environment, acting as a 'leading indicator' of the underlying performance of a service. Because government partisans' priors encourage them to always support the governing party and vice versa for opposition partisans, in principle we would expect non-partisan voters to start from the position of being less satisfied

[1] Note that they could appear to diverge from *one of the other groups* if both partisan groups and non-partisans were moving towards a single point between the partisan groups' satisfaction.

than government partisans and more satisfied than opposition partisans when they don't have recent personal experience.

However, in the previous chapter we have seen that non-partisans are less responsive to differences in performance than partisans. They seem to consist of disengaged respondents who tend not to update their views based on new experiences. This suggests that non-partisans either have strong priors, or that they are less attuned to information about performance - meaning that they receive and process fewer performance signals, compared to partisans. This is supported by evidence in the literature that non-partisans are less informed individuals with little interest in politics (Campbell 1960; Converse 1962; Green & Jennings 2017b). By contrast, partisans feel a duty to be engaged (Rau 2022). In the BSA data used for this chapter, I note that non-partisans were less likely than partisans to read morning newspapers three or more times a week, even after controlling for demographic differences and year (see Appendix A2 for details); this provides another small data point consistent with the 'disengaged non-partisan' hypothesis.

In this chapter, I will explore whether this is also the case with personal experience, or whether non-partisans respond to direct performance signals as expected by the 'partisan screen' approach. In line with the evidence from the previous chapter, I therefore expect non-partisans to update their views to the same extent or less than partisans.

> Secondary Hypothesis 5: Recent NHS contact will have the same or lesser effect on non-partisans' satisfaction, compared to that of partisans.

In summary, partisan differences may be reduced by greater information, but in some scenarios they will not be. However, in all circumstances, partisan differences should not diverge following NHS contact, nor should non-partisans' views diverge from those of partisans.

7.2 Research Design

7.2.1 Data Sources

I will use two separate data sources to explore this question, as each has its methodological advantages.

The first is the British Social Attitudes (BSA) survey data, which also used in the previous two chapters. Because this is composed of repeated cross-sectional surveys rather than taking repeated measurements from the same panel of individuals, more assumptions need to be made when making over-time comparisons. The main assumption here is that the demographic and political controls included in the model are sufficient to rule out any unobserved confounding variable.

For example, imagine that chronic illness is correlated both with partisanship and probability of recent NHS treatment. Because chronic illness is not directly measured in the survey data, it would not be controlled for, and therefore could cause a spurious correlation between recent NHS treatment and partisanship to emerge. Since chronic illness is correlated with age, in this case the descriptive models indirectly attempt to control for any inter-group differences in chronic illness, but this may not be sufficient. In addition, there may be unknown confounders involved, or even reverse causality (e.g. if satisfaction leads to change of partisanship, and not just vice versa).

Despite these limitations, the BSA is a rich dataset with a wide variety of variables across many years, so I believe it is still very useful for exploratory and descriptive analysis exploring potential associations. The BSA data includes demographic details including socio-economic group, marital status, region and age, as well as political variables including partisanship. These are discussed in more detail below.

Bearing in mind the limitations of the BSA cross-sectional survey data, I will then look to triangulate my findings using panel data from Understanding Society (the successor to the British Household Panel Survey). Because the same respondents are surveyed repeatedly for

the Understanding Society panel, I can estimate within-individuals models that control for differences at the individual level and compare changes in the same respondents over time. This therefore avoids some of the challenges to identifying trends that are inherent in cross-sectional data. However, the Understanding Society survey only asks about health service contact and NHS evaluations a few times and in a more limited time frame, meaning it is helpful for triangulation of the BSA data, but does not offer the same comprehensive coverage.

In summary, the BSA allows comparison between individuals who were and were not recent inpatients over a long period of time. Understanding Society allows comparison of a set of the same individuals before and after being inpatients, over a shorter period of time.

7.2.2 Outcome variables

In the British Social Attitudes survey data, I use the NHSSAT variable. This is the answer to the question: "All in all, how satisfied or dissatisfied would you say you are with the way in which the National Health Service runs nowadays?". Respondents can choose "Very satisfied" (coded as 5), "Quite satisfied" (coded as 4), "Neither satisfied nor dissatisfied" (coded as 3), "Quite satisfied" (coded as 2), or "Very dissatisfied" (coded as 1). Refusals or "Don't know" responses are treated as missing data.

This response is useful as it captures broad attitudes to the NHS as a public service. Ideally, there would also be questions about whether the service was improving or getting worse, but since the question is asked in all years, trends over time are easily observable. The five-point Likert item response is treated as a continuous variable in the main analysis. Logistic regression models that do not treat NHSSAT as linear are tested in Appendix A2.

Unfortunately, no general NHS satisfaction question is available in the Understanding Society panel survey. The most applicable question, LOCSERB, is part of a question matrix asking the respondent to rate local services. This variable relates specifically to the response about "Medical facilities". This means that the panel data models are specifically measuring the effect of recent NHS treatment on satisfaction with "local medical facilities", which for some

respondents may correspond to their views on the NHS in general. The link between service satisfaction metrics and evaluations of general government performance were previously discussed in chapter 3 (Section 3.2.1). For other respondents, though, they may hold differing views about the medical facilities local to them, and the overall political management of the National Health Service (NHS). The LOCSERB outcome metric changing as a result of NHS contact would imply that views about local NHS services can be changed by NHS contact, but does not necessarily extend to views of the NHS in general. As a result, the Understanding Society data cannot wholly confirm findings from the BSA data, but if the BSA data shows an effect, it implies that an effect should also be seen in the Understanding Society metric (i.e. because people's personal experience of services has affected their view of the NHS as a whole, it should also affect their view of the services they used).

Responses are given on a four-point scale: "Poor", "Fair", "Very good" and "Excellent", which I code as 1-4 respectively. Refusal or "Don't know" answers are treated as missing. As the scale has only four points and is therefore potentially more problematic to treat as continuous, I conduct an additional multinomial regression model in Appendix A2.

7.2.3 Predictors

In order to operationalise the concept of having substantial contact with the NHS, I have chosen to focus on hospital inpatient services. Typically, being a hospital inpatient would involve either an elective or day-case admission following a referral from a GP or dentist and inpatient stays are often followed up on with at least one outpatient appointment. Admission can also be non-elective (emergency) - rather than a GP referral this generally would have included a 999 call and ambulance transport. Therefore, a hospital inpatient spell usually implies multiple contacts with several different staff, potentially in different settings, and a sufficiently serious condition to warrant secondary care use. I argue that this type of interaction can be expected to provide credible information to the individual about how that service is performing.

An alternative measure that I could have used to represent NHS contact is the same question, but for being an outpatient in hospital. An outpatient does not stay overnight, but is

assessed, treated or diagnosed in a clinic appointment (NHS.uk 2020). The number of appointments varies; however, around two thirds of outpatient appointments are 'follow-up' or 'review' appointments, with one-third being the first appointment for a particular issue (NHS Digital 2020). As mentioned above, some outpatient appointments are to follow up on recovery after inpatient treatment, so there is overlap between recent inpatient and outpatient status.

Thinking about the opportunities for receiving performance signals through personal experience, being an inpatient implies a much more significant interaction than being an outpatient alone: an outpatient may have two appointments over the course of a few months, which may each only last ten minutes (Durham & Darlington NHS Foundation Trust 2014), compared to at least one overnight stay for an inpatient. I would therefore expect a larger effect among recent inpatients than recent outpatients.

After conducting the main analysis, I repeat several of the main analyses using recent outpatient status - these analyses are available in Appendix A2. Using outpatient status yields a bigger 'treated' sample size, but the overall sample size is slightly lower. As expected, the estimated treatment effect is slightly smaller when using outpatient status as the predictor, as compared with inpatient status. Since theoretically and empirically, inpatient status gives a clearer signal, I use inpatient status as my predictor in the main body of this chapter.

In the British Social Attitudes survey between 2004 and 2018, respondents are asked "In the last twelve months, have you or a close family member or close friend… been an inpatient in an NHS hospital?". I use this as the main predictor for the descriptive analysis, with responses indicating that the respondent themselves had been in hospital being coded as Treatment, with all other responses coded as Control.

In the BSA dataset, this gives a control group (those not recently inpatients) of 9,561 and treatment group of 3,292. Sample size by year is shown in Figure 7.1 (counting only respondents who answered both the NHS satisfaction and inpatient status questions, not including "Don't Know").

After 2009, the sample size is much smaller. This is because for later years, the question about being an NHS inpatient was only included in one version of the self-completion questionnaire,

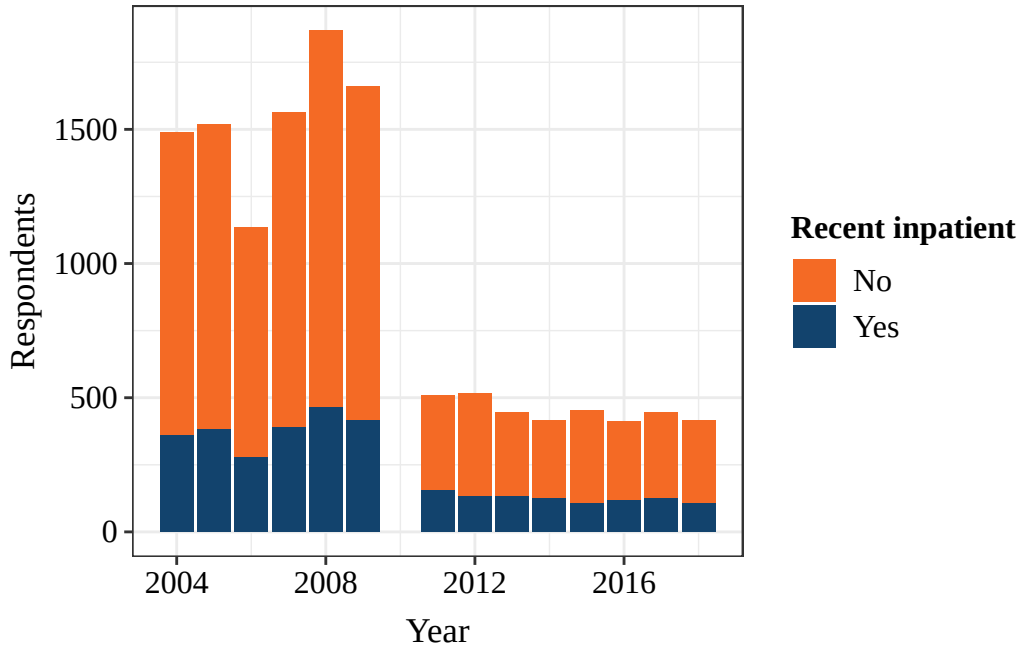

Figure 7.1: Sample size (number of respondents) in the British Social Attitudes survey data by group and year

meaning fewer respondents were asked and fewer completed questionnaires were received. We can therefore expect greater uncertainty for estimates based on single waves of the post-2009 data.

In the Understanding Society data, respondents are asked whether they have been "in hospital or clinic as an in-patient overnight or longer" in 1998, 2003 and 2008[2]. Overall, 1,969 responses were "Yes" (some from the same individuals) and 10,489 responses were "No". Figure 7.2 shows the split by year.

7.2.4 Groups

As well as the overall effect of recent NHS treatment, these hypotheses also relate to the effect of that treatment among different partisan groups.

[2]This question is also asked in two later waves, but not in conjunction with questions evaluating local services.

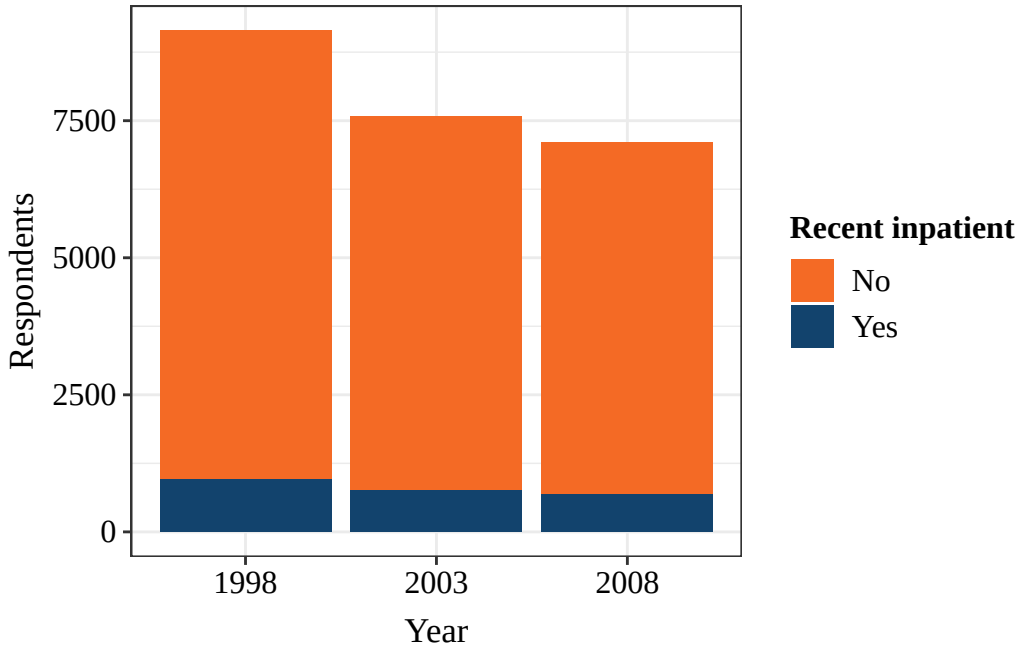

Figure 7.2: Sample sizes (number of respondents) in the Understanding Society panel data by group and year

To operationalise partisanship in the BSA models, I recode the PARTYID2 variable, which is based on the answers to the questions "Generally speaking, do you think of yourself as a supporter of any one political party?", and the follow-up "Which one?". If no preference is given, they are asked "Do you think of yourself as a little closer to one political party than to the others?", and if no preference is given again, the respondent is asked "If there were a general election tomorrow, which political party do you think you would be most likely to support?". No party name is prompted. I then convert this in each year into either "Government partisan" if they support a governing party, "Opposition partisan" if they support a non-governing party, or "Non-partisan" if they did not give a party preference at all[3].

Figure 7.3 shows changes in political party identity over time in the BSA data (among those responding to the NHS inpatient and satisfaction questions only).

[3]This means that during the Conservative/Liberal Democrat coalition years, Liberal Democrat supporters are coded as 'Government Partisan', whereas in all other years they are coded as 'Opposition Partisan'.

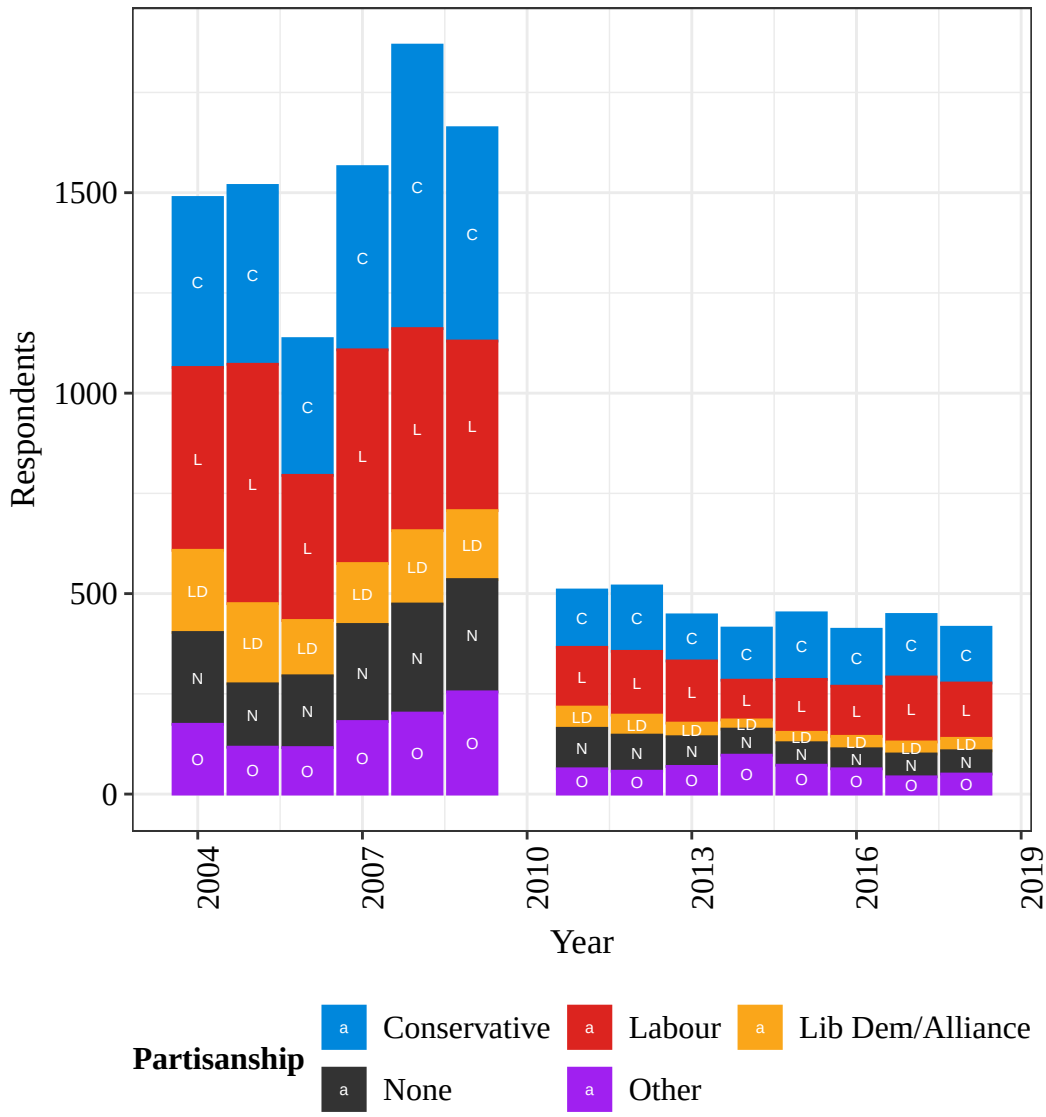

Figure 7.3: Sample size from the British Social Attitudes survey data by party identity and year

7.2.5 Control variables

As described in the "Data Sources" section, using the British Social Attitudes cross-sectional survey data brings with it the possibility that variables other than recent NHS treatment could act as confounders. In order to reduce this risk, these models control for a set of demographic variables that have previously been shown to be influential on NHS satisfaction, or which have been used as controls in similar studies. These are gender (Welch & Thomas 1988), age (Calnan, Almond & Smith 2003), social class and ethnic minority status (Laycock 2009), marital status and region (Calnan, Almond & Smith 2003; Welch & Thomas 1988). These controls were previously used in chapters 5 and 6.

In the Understanding Society panel data, fewer demographic and political variables are available. However, the two-way fixed effects and difference-in-differences designs already account for time-invariant differences between individuals such as demographic factors - either through absorption into the individual fixed effect, or through comparison with pre-treatment outcomes for the same individuals (in the case of difference-in-differences).

7.3 Models and expectations

7.3.1 The overall effect of NHS contact on satisfaction

In order to test Primary Hypothesis IV (*"recent service use will affect voters' evaluations of the NHS"*), I estimate a set of models to test each individual secondary hypothesis. I will firstly test Secondary Hypothesis 1 ("*people who have recent experience of health services will have a different level of satisfaction than those who do not*").

In order to do this, I will estimate OLS (Ordinary Least Squares) regression models with and without demographic and political control variables using British Social Attitudes cross-sectional survey data. For these models (the bivariate model M1 and multivariate model M1a), the predictor is recent contact and the outcome variable is self-reported satisfaction

with the NHS. This will therefore give an overall effect size for the association between recent NHS inpatient experience and satisfaction with the NHS.

Because the data is cross-sectional, this design compares across different individuals, and is therefore potentially adversely affected by unmeasured confounders. As such, we cannot recover causal estimates from this kind of design. However, it can tell us about the associations that exist in the data after controlling for known confounders.

Having shown using cross-sectional data whether respondents who said they were recently NHS inpatients differed in their level of satisfaction from those who did not, I will then seek to triangulate the results using the Understanding Society panel data.

I will use Understanding Society panel data to test the effect of being a recent inpatient on evaluations of local health services. I will do this using a two-way fixed-effects (TWFE) model (M2), indexed for individual and year. The individual fixed effects absorb any individual-specific variation, while the year fixed effects absorb any variation particular to an individual year, leaving only over-time group variation. Although this was until recently a common approach to this kind of problem, there are potential issues with the weighting schema inherent in the TWFE model when used in a differential treatment timing setup, which can introduce bias into the results - for example, by using units that were previously treated and therefore potentially still affected by the treatment, as a control (Imai & Kim 2020).

To allay concerns about these issues, I also use a more recent methodology, Callaway and Sant'Anna's doubly robust difference-in-differences (DID) method (Callaway & Sant'Anna 2021b), to test the effect. Whereas in the TWFE model previously treated individuals could potentially re-enter the control group in future years, this model (M3) only compares treated individuals to those who have not yet been treated or will never be treated. Other potential model misspecification is mitigated by using the doubly robust estimators proposed by Sant'Anna and Zhao (2020).

Difference-in-differences designs can provide an unbiased causal estimate of a treatment effect when certain assumptions hold. In this analysis, the most important assumption is that of parallel trends. The parallel trends assumption is that the control and treatment groups'

trends would have continued in parallel had the treatment not taken place. Differences in the levels of the outcomes are not problematic (either pre- or post-treatment), since the estimand is the difference in trends over time, rather than the differences in the outcome metric itself. Unfortunately, because only three years are available in the data for this model and treatment is staggered, options for checking parallel trends are limited. However, a statistical test of parallel trends pre-treatment will be reported from Callaway and Sant'Anna's software.

The overall average treatment effect will be reported, as well as treatment effects in the first year the respondent reports NHS inpatient treatment, and 5 years later.

7.3.2 NHS contact effects and partisanship

In order to test the remaining hypotheses (2-5), I then return to the BSA cross-sectional data.

> Secondary Hypothesis 2: partisans with recent experience of health services will have a different level of satisfaction than others within their partisan group who do not.
>
> Secondary Hypothesis 3: Differences in satisfaction between government and opposition partisans (the partisan gap) will not be greater among those with recent NHS contact.
>
> Secondary Hypothesis 4: Service contact should not result in non-partisans' satisfaction moving in the opposite direction from both other partisan groups.
>
> Secondary Hypothesis 5: Recent NHS contact will have the same or lesser effect on non-partisans' satisfaction, compared to that of partisans.

I will estimate a multivariate OLS regression model (M4) that includes interactions between treatment status and partisanship. Examining these coefficients, and deriving predicted satisfaction levels from this model, will show whether British Social Attitudes Survey respondents who reported recent NHS inpatient experience had different views of the NHS, as compared to those who did not, but who had the same partisan background. This will also allow me to

Table 7.1: Results from OLS regression results (M1 and M1a) showing that average NHS satisfaction is higher among recent NHS inpatients than others (BSA data)

	Bivariate (M1)	Multivariate (M1a)
(Intercept)	3.312***	2.857***
	[3.288, 3.335]	[2.745, 2.969]
Recent inpatient	0.118***	0.084***
	[0.072, 0.165]	[0.038, 0.131]
Num.Obs.	12 853	12 474

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

estimate whether the difference in satisfaction between government and opposition partisans was greater or not for those who reported recent inpatient experience. Demographic controls and dummy year effects will be included to avoid bias. This model will therefore give evidence to help validate or invalidate Hypotheses 3-5.

7.4 Results

7.4.1 The overall effect of NHS contact on satisfaction

I first estimate two OLS regression models that test how information gained from recent inpatient experience affects satisfaction with the NHS. In these models, the association is modelled between a dummy variable representing whether the respondent reported having been an NHS inpatient in the previous year, and their reported satisfaction with how the NHS is being run (on a five-point scale). Table 7.1 shows the estimated coefficients for the main effects and intercept of a simple bivariate regression (model 1) and a model that also includes demographic controls (model 1a).

The results from the bivariate model (M1) indicate that recent inpatients' NHS satisfaction rating is, on average, 12% of one step on the Likert scale higher than others. This falls slightly to 8% when demographic and political factors have been controlled for. The full model coefficients and 95pc confidence intervals are shown in Figure Figure 10.1 in Appendix A2.

This result supports Hypothesis 1: there is a statistically significant association between recent NHS treatment and higher satisfaction with the NHS. The association is robust to the inclusion of controls that account for political and demographic differences between respondents.

This supports the central concept behind this chapter: the idea that having personal experience of the NHS is a good source of new information, and processing this information causes people to update their level of satisfaction with public service performance. It is notable that this treatment effect emerges among people who have had experience of being an inpatient in the past twelve months: there will inevitably be people in the 'control group' who have had numerous GP appointments and/or outpatient appointments, which could be diluting the effect size. Nevertheless, having spent at least one night in hospital, even months before, has a measurable impact on voters' satisfaction with the health service.

The difference between treated and control respondents (as estimated by the coefficient from the multivariate model) is approximately 0.07 standard deviations in the NHSSAT outcome variable. Although this seems small, this is because there is considerable variation at the individual level; however, comparing this difference to other between-group differences in means makes clear that its magnitude is meaningful. To contextualise the size of the association between recent inpatient status and satisfaction, it is approximately 74% as large as the average difference in overall public satisfaction between two consecutive years (0.11 points on the 1-5 scale), and 42% as large as the average difference in satisfaction between government partisans and opposition partisans (0.2 points on the Likert scale). Given the prominence of the role of partisanship in discussions of the politics of public service provision both within the political science literature and in non-academic commentary, and the importance of over-time differences in political change, these comparisons suggest a smaller, but meaningful role for contact with public services in updating public opinion.

Continuing the analysis of the same model, Figure 7.4 shows the predicted NHS satisfaction ratings for those with recent inpatient experience and those without, for a hypothetical 'typical' respondent with demographic characteristics set at their mode (a 55-67 year old married woman, from an intermediate non-manual employment background, in South East England, in

2008). Note that although the confidence intervals overlap between the two groups, the 8.4ppt contrast (predicted difference) between treated and control groups is statistically significant at the p = 0.05 level (95% CI: 3.8% - 13% of a step on the Likert scale).

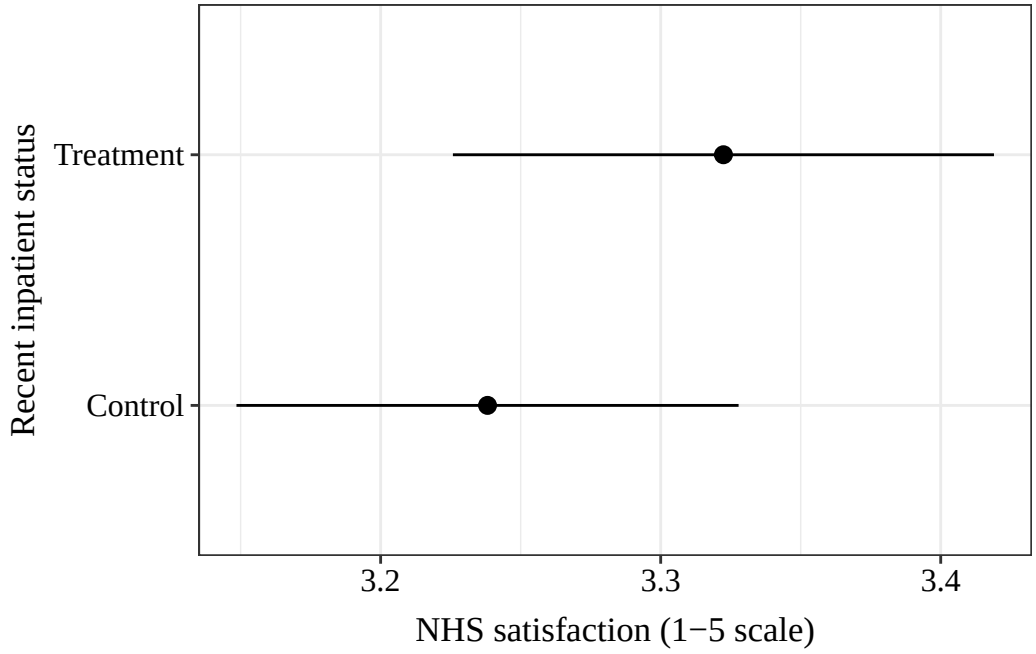

Figure 7.4: Predictions from multivariate regression (M1a) showing that recent NHS inpatients are more satisfied with the NHS than others in the BSA data

I now attempt to triangulate my findings using the Understanding Society panel data by looking at variation over time within the same individuals. I first estimate a Two-Way Fixed-Effects (TWFE) model (M2), indexed for individual and year. Results are shown in Table 7.2.

As the results indicate, being a recent inpatient had a statistically significant effect at the 95pc level, after accounting for individual differences and differences between years. The estimated coefficient is 0.11, indicating a difference of around 7% of one step of Understanding Society's four-point Likert scale measuring satisfaction with local medical facilities[4].

Because of the possibility of bias due to longer treatment effects, I also perform a difference-in-differences analysis on Understanding Society panel data using the doubly-robust Callaway-

[4]For an alternative multinomial logistic model that avoids treating this Likert scale as continuous, see Appendix A2.

Table 7.2: Results from two-way fixed effects model (M2), showing that recent inpatients evaluated their local medical facilities more highly than others in the Understanding Society data

	Two-way fixed effects model (M2)
Recent inpatient	0.105*** [95pc CI: 0.048, 0.162]
Num.Obs.	9032
Std.Errors	by: pidp
FE: pidp	X
FE: year	X

Sant'Anna method (M3). The parallel pre-trends test yielded a p-value of greater than 0.05, suggesting a lack of evidence for significant differences in pre-treatment trends in satisfaction. Figure 7.5 also indicates that there was no statistically significant treatment effect 5 years before treatment.

In Table 7.3, the left column shows an average treatment effect on the treated across both groups (those first treated in 2003 and those first treated in 2008), shown as "ATT (Average)". This is the overall effect of treatment, but assumes that treatment effects remain (e.g. that someone who was a recent inpatient in the 2003 data would still be a recent inpatient in some sense in 2008). In this model, it was estimated as approximately 11% of a step on the Likert scale, with 95pc confidence intervals ranging from around 3-19% on the same scale.

The right column shows the average treatment effect on the treated by time since treatment. Here, "ATT(0)" refers to the effect when the survey was taken within a year of being an inpatient; "ATT(-5)" is five years pre-treatment, and "ATT(5)" is five years after treatment. This shows that when only considering the immediate effect of NHS treatment, the effect size is estimated slightly higher than the overall average, at 14% of a Likert step (95pc CI from 3-25%). The effect five years later is estimated at 9% of a Likert step, but as the 95pc confidence intervals cross zero, we can't be sure this effect is meaningful. This is demonstrated graphically in Figure 7.5.

Returning to the idea of benchmarking the size of this association against other comparisons in

Table 7.3: Results from doubly robust difference-in-differences estimation (M3), showing that people rate their local medical facilities more highly after inpatient treatment, compared to untreated individuals (Understanding Society data)

	Overall DID from M3	Time Since Treatment from M3
ATT(Average)	0.108	
	[0.025, 0.191]	
ATT(-5)		0.045
		[−0.123, 0.212]
ATT(0)		0.136
		[0.032, 0.241]
ATT(5)		0.111
		[−0.021, 0.243]
Num.Obs.	3607	3607
Std.Errors	by: pidp	by: pidp
type	group	dynamic

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

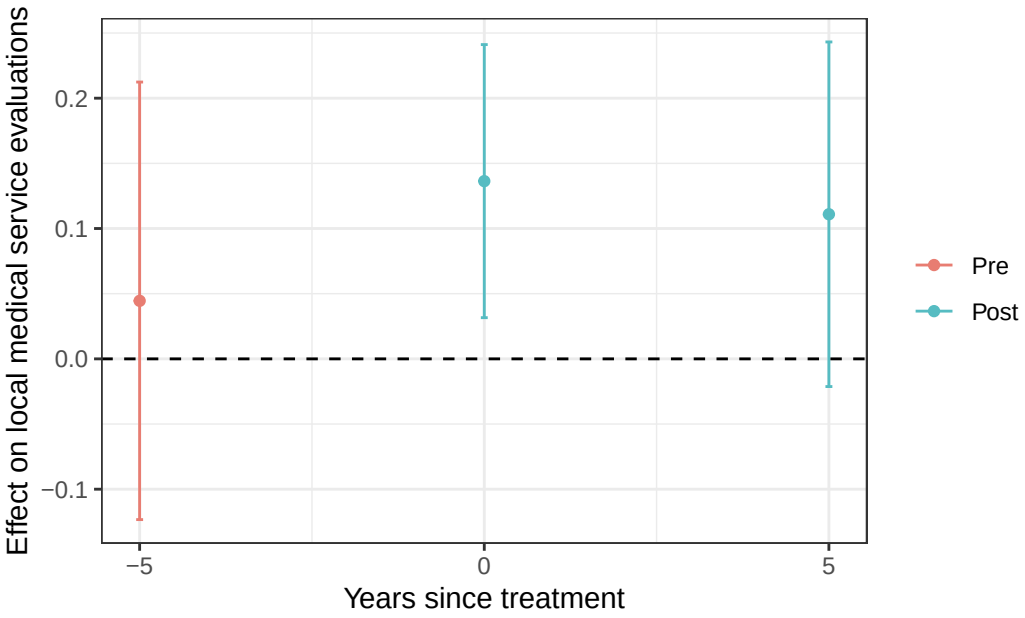

Figure 7.5: Estimated effect in different time periods of hospital treatment on evaluations of local medical facilities from doubly robust difference-in-differences estimation (M3) on the Understanding Society data

the data, I can now compare the estimated difference-in-differences in the year after inpatient treatment against other variation in the data.

The estimated ATT (0.14) is approximately 3.86 times larger than the average partisan satisfaction gap between government partisans and opposition partisans in the Understanding Society data. By comparison, this contact effect is much larger than the equivalent contact effect for the BSA data, however, the partisan gap in the Understanding Society dataset is very small to begin with (0.04). This is likely because the Understanding Society survey waves used here were all from a period of time when the partisan gap on this issue was historically small, as well as the more 'local' nature of the question, which may activate fewer partisan biases, and be more closely linked to local performance rather than national performance - meaning that the two effect sizes are not well suited for direct comparison.

In terms of within-person variance, the average respondent's standard deviation for their satisfaction responses is 0.69. The estimated difference-in-differences is 51 % of that mean individual standard deviation. This suggests that changes in opinion associated with recent NHS inpatient experience are meaningful, but that other over-time factors are at least as important in influencing respondents' evaluations of local services.

7.4.2 NHS contact effects and partisanship

If NHS contact leads people to update their prior evaluations of government performance on the NHS, this will be more effective as a transmission mechanism for performance signals if it is effective among all partisan groups. There is therefore an expectation in the moderated learning theory that learning will take place among all groups, on the condition that there is a consistent flow of performance signals. The theory also posits that contact with services will allow people to gain more information about public service performance and update their evaluations; this process may result in a narrowing of partisan gaps, although for reasons discussed earlier in this chapter, it may be the case that partisan groups move in parallel. The minimal expectation here is that partisan groups (including non-partisans) will not move away from each other (i.e., partisan gaps will not increase).

Table 7.4: Multivariate OLS model (M4) results showing that people with recent inpatient experience were more satisfied with the NHS than others within their partisan group, across all party identities (BSA data)

	Multivariate OLS (M4)
(Intercept)	3.055*** [95pc CI: 2.937, 3.172]
Recent inpatient	0.121** [95pc CI: 0.042, 0.201]
Non-partisan	−0.204*** [95pc CI: −0.279, −0.129]
Opposition partisan	−0.187*** [95pc CI: −0.240, −0.135]
Recent inpatient x Non-partisan	−0.169* [95pc CI: −0.313, −0.025]
Recent inpatient x Opposition partisan	−0.023 [95pc CI: −0.125, 0.080]
Num.Obs.	12 469

These expectations led to Secondary Hypothesis 2 in relation to government and opposition partisans:

> "*partisans with recent experience of health services will have a different level of satisfaction than others within their partisan group who do not"*

and to Secondary Hypotheses 4 and 5 in relation to non-partisans:

> *"Service contact should not result in non-partisans' satisfaction moving in the opposite direction from both other partisan groups" "Recent NHS contact will have the same or lesser effect on non-partisans' satisfaction, compared to that of partisans*"

In order to test these, I return to the British Social Attitudes (BSA) data. This is because the question in the BSA survey is more closely linked to national government performance than the Understanding Society question, and the Understanding Society sample size is much smaller. I first estimate the effect of NHS inpatient experience, mediated by partisanship, using a multivariate OLS model (M4), with interacted treatment terms for NHS inpatient status and partisanship, demographic controls, and year dummies. The main results are shown in Table 7.4.

The Treatment variable is statistically significant at the $p = .05$ level, with a coefficient of 0.12, indicating that government partisans with recent NHS inpatient experience were around 12% of a point on the Likert scale more satisfied than those without. The interacted treatment

term 'Treatment x Non-Partisan' is statistically significant, which means that the association between recent NHS experience and satisfaction differs between government partisans (the reference category) and non-partisans.

For a more understandable view of the inter-group differences, group-average marginal effects (G-AME) are calculated and shown in Figure 7.6. These estimates represent the partisan group-level average predicted increase in NHS satisfaction associated with being a recent inpatient.

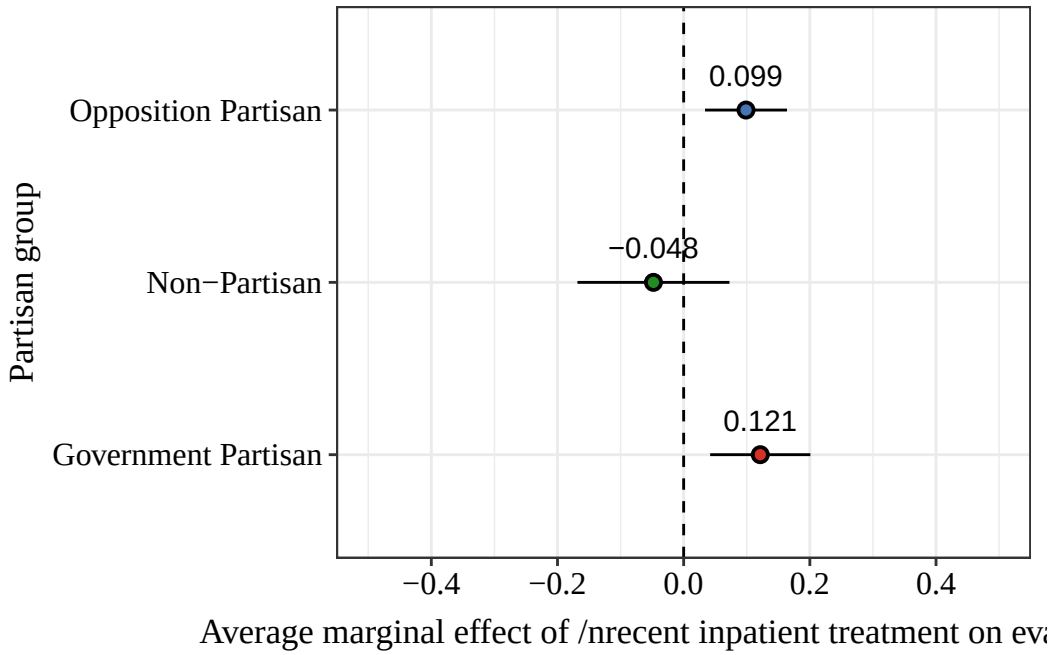

Figure 7.6: Contrasts from OLS regression (M4) showing the effect of recent inpatient stay by party identity group in the BSA data

As Figure 7.6 shows, among both government partisans and opposition partisans, predicted satisfaction is higher for respondents with recent NHS inpatient treatment. As the regression results showed, the average marginal effect of recent treatment among government partisans is around 12% of a step on the 5-point Likert scale (95% CI: 4% - 20%). The average marginal effect among opposition partisans is approximately 10% (95% CI: 3% - 16%). Non-partisans with recent NHS inpatient experience had a 4ppt lower predicted satisfaction than those without (95% CI: -17% to +7%), but the difference was not statistically significant at the $p = .05$ level.[5]

[5]A very similar analysis using the Understanding Society data had the same substantive findings, but with much

These results suggest that the treatment effect of NHS contact shown in the previous model is highly likely to be operative among partisans. The model therefore adds evidence for Secondary Hypothesis 2 ("*partisans with recent experience of health services will have a different level of satisfaction than others within their partisan group who do not").*

However, the results are unclear in relation to non-partisans. In particular, there is no clear evidence for Secondary Hypothesis 4 ("*Service contact should not result in non-partisans' satisfaction moving in the opposite direction from both other partisan groups*"). While the effect estimate for both government and opposition partisans is positive, meaning that partisans with recent inpatient experience were more satisfied with the NHS than partisans without recent inpatient experience, the estimate for non-partisans is negative but not statistically significant. This means that non-partisans with recent NHS contact may be more satisfied, less satisfied or just as satisfied as those without.

7.4.3 Differences in partisan gaps among people with recent NHS inpatient experience

One of the assumptions in the moderated learning theory set out in chapter 4 of this thesis was that additional information would not lead to partisans updating their evaluations in different directions. This was set out in Secondary Hypothesis 3 (*"Differences in satisfaction between government and opposition partisans (the partisan gap) will not be greater among those with recent NHS contact"*). In order to test whether this was the case, I use the partisanship-based BSA model (M4) to compute the contrast between a typical respondent with their partisanship set to "government partisan" and one with their partisanship set to "opposition partisan" in the model, both with recent NHS inpatient experience. I perform the same contrast between partisan groups where the respondents are in the control (not recent inpatient) condition. Figure 7.7 shows the estimated partisan gap between government and opposition partisans for each of the treatment groups.

lower precision - see Appendix A2 for details.

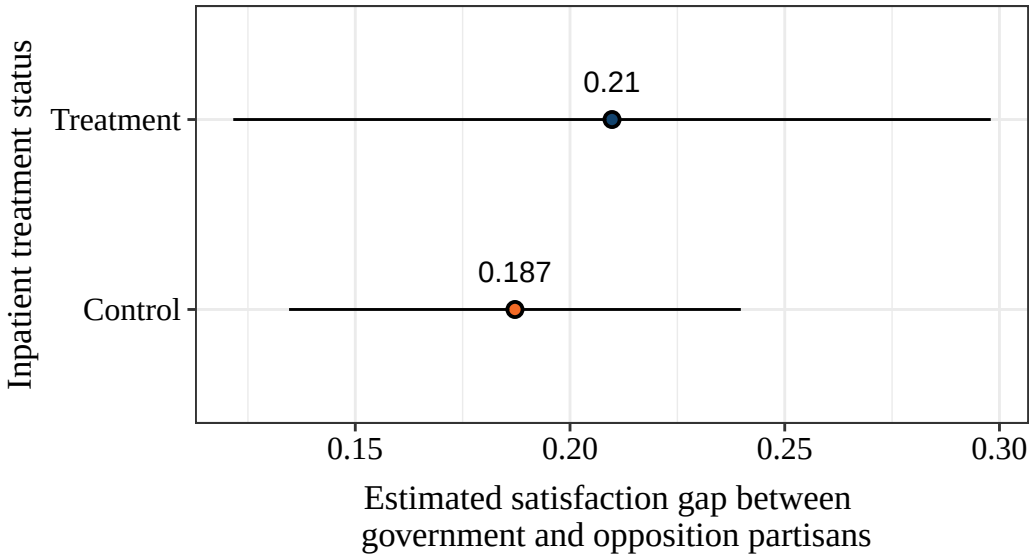

Figure 7.7: Contrasts from M4 showing the difference in partisan gap among treated and non-treated respondents. It shows that the gap is slightly wider among treated individuals; however, subsequent statistical hypothesis testing shows that the difference in partisan gap is not statistically significant at the .05 level (BSA data)

As Figure 7.7 indicates, the model predicts that the partisan gap is bigger among recent NHS inpatients than among others. However, subsequent statistical testing (plots of which are found in the Appendix) shows that the estimated 2% difference in partisan gaps between treatment groups (difference-in-differences) is not statistically significant (95% CI: -12% to +7%). This means we cannot be sure if recent NHS contact was associated with an increase or decrease in the partisan gap.[6] As such, I do not find evidence to support Secondary Hypothesis 3.

One of the aims of this chapter was to test whether non-partisans were less responsive to performance information gleaned from personal experience than partisans - similarly to how the previous chapter showed that non-partisans were less responsive to differences in waiting time performance generally. Secondary Hypothesis 5 posits that "*recent NHS contact will have the same or lesser effect on non-partisans' satisfaction, compared to that of partisans".* The results of Figure 7.6 are ambiguous on this point, since the 95% confidence intervals for the effect of recent NHS contact on non-partisans overlap with those of its effect on partisans. To test

[6]Given that the sample size in the British Social Attitudes survey is too small to provide a precise estimate of this complex parameter, I do not repeat the exercise on the much smaller Understanding Society dataset.

whether non-partisans were indeed less responsive to NHS contact, I conduct two difference-in-differences hypothesis tests, which compare the group-average marginal effect (G-AME) of recent inpatient experience among non-partisans with the equivalent group-average marginal effects among government and opposition partisans.

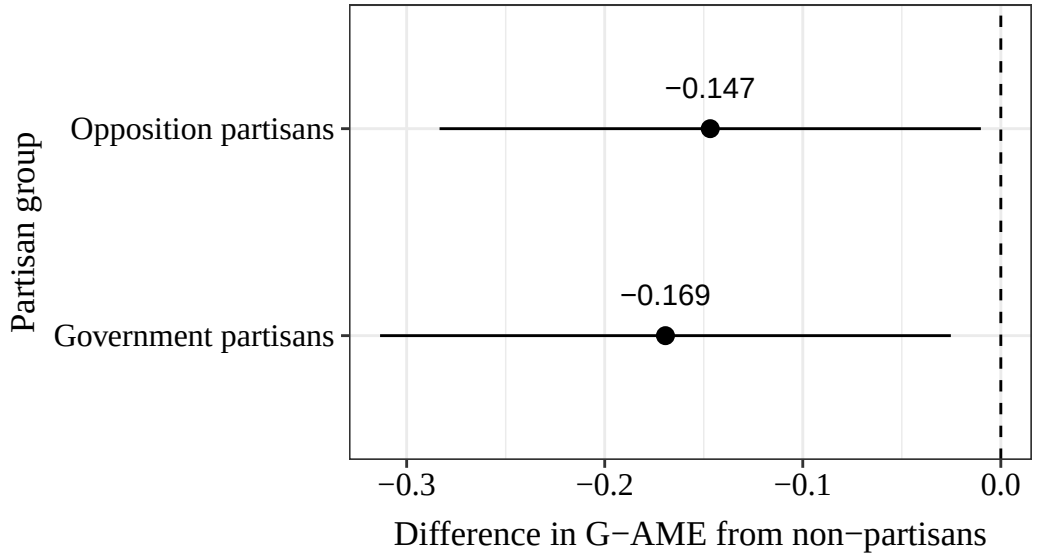

Figure 7.8: Hypothesis tests showing that the group-average marginal effect (G-AME) of recent NHS inpatient experience is smaller among non-partisans than among government or opposition partisans (BSA data)

Figure 7.8 shows that the group-average marginal effects for non-partisans were lower than those for opposition and government partisans, as expected. However, the 95% confidence intervals suggest that these differences were statistically significant at the $p = .05$ level. This is compatible with the idea that that non-partisans are less likely to update their views a result of personal experience of NHS services compared with people who identified with a particular political party, and echo the findings about non-partisans from the previous chapter.

7.5 Discussion

This chapter aimed to explore the extent to which voters change their evaluations of public service performance in response to information gained through personal experience of services,

and in so doing test Primary Hypothesis IV (*"recent service use will affect voters' evaluations of the NHS"*). Using two sets of survey data, one cross-sectional and one panel-based, I tested secondary hypotheses that broke this broader hypothesis down further into questions about a) the overall effect of recent NHS treatment, and b) how this interacts with partisanship and non-partisanship. The 'Results' section was structured according to the type of analysis being used; in this section, I will first address the overall effect of NHS contact, then discuss how this varies by partisanship, and then focus on non-partisans separately.

In general, there was support for the theory and Primary Hypothesis IV in particular. Recent inpatient experience was associated with higher satisfaction with the NHS and local medical facilities, indicating that health service utilisation yields information about performance that allows voters to update their views, even among most partisans. The results about non-partisans again suggested that they are less responsive to information about performance than either government or opposition partisans.

7.5.1 Overall effect of recent NHS treatment

The analyses in this chapter found that there was an association between recent NHS inpatient treatment and being more satisfied with health service performance, which suggests that people do update their views about public service performance based on their own experiences. In every model estimated, the overall effect of treatment was estimated to be positive. In both single-level OLS models (the 'overall' OLS model and the 'partisanship' OLS model), the effect of treatment was statistically significant, and in the panel data models this was also the case. While partisans' views differ depending on treatment status, there was no statistically significant difference between non-partisans with recent inpatient experience and non-partisans without.

Given that several different types of model and two different data sets all yield similar effect estimates for recent NHS treatment, there is considerable evidence here to suggest that having recent experience of NHS hospital inpatient treatment results in higher satisfaction with the

health service, at least among people with certain levels of partisanship. The difference-in-differences analysis on panel data suggest that this effect fades over time, as the treatment effect is positive but non-significant after five years.

This supports the theory in that it suggests that performance signals do change people's views, but that partisan priors continue to play a considerable role, even among those with recent experience of services. It also confirms evidence previously presented in the literature that people with recent NHS experience are more satisfied with the health service than others (Edwards 2006). In the context of this thesis, it suggests that personal experience is another mechanism by which people receive signals about performance of services, helping to enable government accountability for delivering positive outcomes.

7.5.2 How these treatment effects differ by partisanship

It is also important to understand the extent to which partisanship moderates how voters learn from personal experience of public services. If learning about government performance by personal experience is only present in some partisan or non-partisan groups, that reduces the extent to which this mechanism can play a role in retrospective evaluation. In that case, inducing accountability among governments would be more difficult.

I therefore tested the following secondary hypotheses:

> Secondary Hypothesis 2: partisans with recent experience of health services will have a different level of satisfaction than their co-partisans who do not.

> Secondary Hypothesis 3: Differences in satisfaction between government and opposition partisans (the partisan gap) will not be greater among those with recent NHS inpatient experience.

Both of these hypotheses were supported by the evidence: both government and opposition partisans, on the whole, were more satisfied with the NHS when they had recently been inpa-

tients, compared to their own partisan group. In addition, there was a slight narrowing of the partisan divide on satisfaction among the treated group compared to the controls.

This suggests that, as predicted by the theory, even partisans' views can also be altered by personal experience, through a moderated learning process using new information about performance. The direction of change was positive for both sets of partisans, which could result from two different explanations.

The first potential explanation is that, rather than result from evaluation of performance signals, this change would *always* be positive because it reflects greater personal salience or appreciation for the NHS. Personal contact may, as speculated, activate voters' views in a less political and more affective way than indirect information about performance and therefore result in a political mental framing not being activated by the survey question. This explanation does not explain why such an effect would be operative among partisans but not, as my modelling showed, non-partisans; however, the lack of a contact effect among non-partisans may have been due to statistical chance, since the estimate was an imprecise null.

An alternative explanation is that both partisan groups revised their satisfaction upwards because the performance signals they picked up on through their experience were very positive - even more positive than government partisans' priors. Expectations of public services may have been depressed by media coverage's tendency to highlight negative outcomes, as previously argued by Klein (2007). This would imply that indirect performance information underestimates the true performance levels of services, and personal contact provides a more accurate update to voters' opinions. One possible way to discern between these explanations would be to analyse detailed survey data that combines questions on survey use and satisfaction *during a period of particularly poor performance*, since in that situation the second explanation would likely not apply.

7.5.3 How NHS treatment affects non-partisans

Earlier in the chapter, I outlined how expectations of non-partisans differ between specific theories within the broader 'moderated learning' approach. A focus on the 'partisan screen' account could lead one to expect non-partisans to have weaker priors and therefore be more willing to change their minds, compared to committed partisans. However, the evidence from this thesis so far suggests that, as has been pointed out in other literature, non-partisan voters are less politically engaged and are less likely to receive performance information and re-evaluate their views accordingly.

In this chapter, two specific hypotheses were tested in relation to non-partisans:

> Secondary Hypothesis 4: Service contact should not result in non-partisans' satisfaction moving in the opposite direction from both other partisan groups.

> Secondary Hypothesis 5: Service contact should not result in non-partisans' satisfaction diverging from those of both other partisan groups.

In line with the discussion above, the results showed that there was a weaker relationship between recent NHS treatment and satisfaction among non-partisans than among partisans. This is again consistent with a view of the non-partisan group in which they are less engaged and responsive to performance signals, regardless of their source.

8 Do voters respond to planned health service cuts?

8.1 Introduction

In chapter 4 of this thesis, I hypothesized that voters update their views about public services using a combination of their partisan-informed priors and new information - obtained either directly through personal experience or indirectly through information they hear or read. This new information provides signals about the quality of a service's performance. Later, in chapter 6, I used descriptive regression modelling to explore how public opinion of public healthcare provision has changed in response to different levels of waiting time performance; this showed that both local and national performance are correlated with people's evaluations of the health service, suggesting that people do adjust their views based on new information about healthcare. In chapter 7, I then then narrowed my focus and used a difference-in-differences design, with panel data, to measure the average effect of hospital inpatient treatment on satisfaction with local health services. This showed that information signals from direct contact with hospital services results in an average improvement in evaluations of the National Health Service (NHS).

In this chapter, I again adopt a quasi-experimental design with panel data, but in this case I test whether the electorate votes retrospectively on the issue of local health service provision in England[1]. By measuring changes in opinions and vote choice in response to planned reduc-

[1]Only pre-intervention values are used as controls - because if voters changed their allegiances in response to retrenchment, this would introduce post-treatment bias, since it would be both a control and an outcome of the

tions (retrenchment) in nearby health services, I can test how a particular piece of information about the NHS affected the first three steps of the four-step accountability chain (perception of change, attribution of credit or blame, retrospective voting, change in politicians' behaviour). If voters changed their opinions and were less likely to vote for the incumbent Conservative party, this would be a strong indication that retrospective voting was operative in this example, with the associated benefits for government accountability. Finding that opinions changed but vote choice remained unchanged as a result of the new information would indicate that while voters could evaluate performance information, other factors prevented retrospective voting from taking place.

This study takes place around the 2017 UK general election. During 2016, NHS commissioning bodies in England created Sustainability and Transformation Plans (STPs) with the aim of improving financial viability through reconfiguration of services. Five acute hospitals were proposed to be downgraded, as well as twenty-four Accident and Emergency (A&E) departments (Milmo, Day, Mort, *et al.* 2017; Milmo 2017).

The plans were published on 15th December 2016 (although in some cases early drafts had already been made public in the press). This allows me to measure any local effect on voting intention by measuring voters' opinions and voting intention from before the plans were public, and comparing these to their post-publication opinions and reported 2017 vote. If these changes are different for people living in areas affected by planned hospital closures, this can be interpreted as the information from the plans causing voters to re-evaluate their views.

8.1.1 Planned changes to hospital services and retrospective voting

In this section, I provide a brief recap of the idea of retrospective voting as described more fully in chapter 2, and how it relates to accountability. I discuss whether we can expect to observe retrospective voting in the particular circumstances of the case under study. The principal motivation for exploring the factors affecting public opinion about the NHS in this thesis is

process (King 2010).

because of the role of accountability. Accountability is the idea that public preferences can influence government policy (Manin, Przeworski & Stokes 1999). One simple mechanism that is proposed as a way that accountability can emerge, is retrospective voting. In the retrospective voting model, voters can choose to reward governments for implementing policies that yielded good outcomes, or punish them for failures (Healy & Malhotra 2013). In this way, politicians are encouraged to respond to public opinion for fear of being replaced. If retrospective voting occurs on the issue of public service management, then governments will necessarily be incentivised to respond to public opinion and maintain good performance standards in those services.

We can see retrospective voting as requiring a three-step process (Anderson 2007). In the first, voters recognise outcomes; in the second, they make a judgement about those outcomes, including attribution of credit or blame when appropriate, and in the third, final step they express that judgement at the ballot box.

There are reasons to believe that we should expect to see retrospective voting following the 2016 proposals of hospital downgrades and closures. Firstly, in chapter 5 I found that there is a (relatively weak) relationship between local waiting time performance and public evaluations of the NHS. This suggests that voters have some awareness of local public service conditions. In this chapter, a different type of challenge to the quality of service provision is being tested, namely a planned hospital closure or downgrade; this should arguably be more noticeable (i.e. generate more performance signals) than a general local difference in waiting times, because it involves changes to the tangible, physical estate rather than a deterioration in value - per Pierson's (1996) argument that tangible losses are felt more keenly than intangible ones.

Secondly, the welfare state literature has generally found that retrenchment is most likely to face opposition and fail when it is highly visible (Lindbom 2014) and targets 'concentrated interests' (Pierson 1996) that are able to organise and rally opposition, such as NHS staff networks and unions. In this case, we know that in the run-up to the 2017 election the opposition Labour party campaigned to stop the plans (Kelso 2017). All of these factors indicate a high level of salience and high likelihood of provoking a 'punishment' response from the electorate.

However, there are also reasons why voters' priors may not have been shifted by the publication of the Sustainability and Transformation Plans. The first is that the information environment may have been saturated. During 2016 and 2017, information about public services competed for voters' attention with the unusually high salience of the fallout from the Brexit referendum (Hobolt 2018), including a change in Prime Minister and negotiations with the EU, controversy over Conservative proposals to change social care funding that were dubbed a "dementia tax" (Asthana & Elgot 2017; Dorey 2017), controversy about and rising support for Jeremy Corbyn's leadership of the Labour Party (Mellon, Evans, Fieldhouse, *et al.* 2018), etc.. While local issues can affect elections, I would argue that national issues tend to be more salient, partly because national media plays a greater role than local media, and because general elections determine who leads the government - unlike, for example, mid-term by-elections.

In addition, the likelihood of NHS cuts may have already been accounted for in voters' prior expectations. In the welfare state literature, it has been argued that governments can successfully reduce services if they are able to convince voters that spending cuts are needed to secure broad financial viability, and that conservatives may be ideologically in favour of retrenchment anyway (Giger 2012; Pierson 1996).

In summary, based on the moderated learning theory set out in this thesis, we would expect to see differences in opinion and voting intention among people living in affected areas, contingent on the hospital change plans being sufficiently salient, voters placing enough weight on local issues, and local cuts not being accounted for in voters' priors. I will therefore set out to test the final primary hypothesis described in chapter 4's theoretical framework:

> PRIMARY HYPOTHESIS V: HEALTH SERVICE RETRENCHMENT WILL RESULT IN RETROSPECTIVE VOTING AGAINST INCUMBENT POLITICIANS

In order for retrospective voting to take place, there first needs to be a perception of performance, and then a judgement made of that performance, and an attribution of credit or blame. Therefore, if retrospective voting emerges as a result of planned hospital retrenchment, then we should also observe a shift in attitudes among affected individuals.

I therefore propose to test Primary Hypothesis V by testing two specific hypotheses about changes in public opinion resulting from planned health service retrenchment:

> Secondary Hypothesis 1: The publication of plans to cut hospital services resulted in worse opinions of government management of the NHS among people living near affected hospitals
>
> Secondary Hypothesis 2: The publication of plans to cut hospital services resulted in lower Conservative vote share among people living near affected hospitals

I will set out in detail in the Methods section how pre-existing differences between people living in affected and unaffected areas will be accounted for, to allow a fair counterfactual comparison and to enable me to make causal claims about changes in opinion resulting from the plans' publication.

8.2 Methods

8.2.1 Research design and modelling assumptions

In this study I use a difference-in-differences design with panel data to estimate the effect of planned local hospital reconfigurations on local people's opinions and vote choice. I do this by tracking changes in individuals' opinions and voting intention before and after the publication of the Sustainability and Transformation Plans (STPs) in 2016 in the British Election Study Internet Panel (Fieldhouse, Green, Evans, *et al.* 2019). I then compare these changes in responses to the relevant questions between people living in constituencies with at least one affected hospital or Accident and Emergency unit and those living in other constituencies. The publication of the STPs in the lead-up to the 2017 General Election provides a good opportunity to use a difference-in-differences design, because the local nature of the plans means that many respondents - those unaffected by the plans - can act as a control group for those living near affected hospitals, providing a counterfactual view of how opinions would have evolved within the treatment group had the treatment not taken place (Angrist & Pischke 2014).

As described in the chapter on personal contact with the NHS, the difference-in-differences design is able to recover an unbiased estimate of the average treatment effect on the treated (ATT) conditional on some assumptions being met, notably those of parallel trends (Goodman-Bacon 2021) and stable unit treatment value (Clarke 2017). Studies claiming a treatment effect are most convincing when a clearly parallel pre-treatment trend and post-treatment divergence can be established (Angrist & Pischke 2008). In this section, I discuss how the setup of this study meets those assumptions.

To assess whether pre-treatment trends were parallel between the treatment and control groups, I will use visual comparison and regression testing. To give greater confidence that the assumption of parallel trends is met, I will re-weight the control group using entropy balancing. This technique finds weights for individual units that result in balance between the treatment and control groups on a set of covariates (Hainmueller 2012). It has been used to establish common pre-treatment trends in a difference-in-differences context in studies that measured the effect of newspaper endorsements in the 1997 General Election (Hainmueller 2012; Ladd & Lenz 2009) and of flooding on political participation (Rudolph & Kuhn 2018); as well as in various contexts within the economics literature (Freier, Schumann & Siedler 2015; Makridis 2019). Entropy balancing also appears to cope better with model misspecification than alternatives such as inverse probability weighting (Zhao & Percival 2016). By estimating the models both with and without entropy balancing, we can have greater confidence that post-treatment differences in trends between the two groups are due to the treatment itself, rather than innate differences between the respondents in the control and treatment groups. I will also follow Ladd and Lenz (2009) in specifying a multivariate regression using controls as an alternative method for reducing differences between control and treatment groups.[2]

As well as the parallel trends assumption, this design is based on on the stable unit treatment value assumption (SUTVA). This is the assumption that every unit (respondent) can be clearly categorised as either "control" or "treatment" in both pre- and post-treatment periods. There

[2]Only pre-intervention values are used as controls - because if voters changed their allegiances in response to retrenchment, this would introduce post-treatment bias, since it would be both a control and an outcome of the process (King 2010).

are two ways that this assumption could be violated in this design: firstly, in terms of time - if some of the sample were aware of the plans during the designated pre-treatment period, that would add a negative bias to the estimate of the treatment effect (Clarke 2017). Secondly, inaccurately allocating respondents to the control or treatment group based on their geographical location could add bias in either direction, if the treatment has a wider or narrower geographical footprint than expected.

To avoid the first of these potential issues, the pre-treatment period ends after May 2016, more than 6 months before the official publication. While some STPs may have codified change programmes that were already in place, given the time-scales it seems unlikely that a meaningful number of voters would be aware of specific plans to downgrade or close hospital services at that point. This therefore excludes the period of time that is ambiguous in terms of treatment status.

Similarly, a clear distinction between areas that are 'affected' and 'unaffected' by local hospital changes is needed. This is not straightforward because hospitals do not have catchment areas, and what is 'local' to one person may not be to another. As parliamentary constituencies are relatively similar in population size, and because in an electoral context parliamentary candidates would mobilize opposition within their constituency, I decided to classify people living in the same constituency as the affected hospital or A&E as being 'treated'. I perform robustness checks on this assumption in Section 8.3.5 by re-running the analysis using two other specifications, one with a broader treatment definition, and one designed to minimize 'spillover' of treatment effects into the control group by removing people from constituencies neighbouring those with affected hospitals from the analysis.

8.2.2 Data Sources

Details of planned hospital and A&E closures were taken from from the Johnston Press Investigation's findings published in the i newspaper (Milmo, Day, Mort, *et al.* 2017; Milmo 2017), checked against local Sustainability and Transformation Plans where clarification was needed.

Only acute hospitals were included; community hospitals, which are generally smaller facilities for rehabilitation of elderly patients after surgery (Pitchforth, Nolte, Corbett, *et al.* 2017), were not included on the basis that their throughput is much smaller and public awareness of these hospitals is lower.

Postcode details for affected hospitals were manually gathered from Google Maps and entered into a Parliamentary lookup tool to determine which constituency they were in (Google 2018; UK Parliament 2018). Panel study respondents were then categorised intro treatment and control groups on the basis of their stated parliamentary constituency and if it contained an affected hospital.

I used waves 3, 4, 6, 7 and 12 of the British Election Study Internet Panel 2014-2019 (Fieldhouse, Green, Evans, *et al.* 2019). Dates for the waves and how they are used in the analysis are shown in the table below:

Table 8.1: Dates for BESIP waves

Wave	Dates	Role
3	September - October 2014	For checking parallel trends
4	March 2015	For checking parallel trends ("NHS change" variable only)
6	May 2015	For checking parallel trends (attitude to cuts and Conservative vote intention variables)
7	May 2016	Pre-treatment wave
12	May - June 2017	Post-treatment wave

Waves prior to Wave 12 are considered 'pre-treatment' waves, with the treatment occurring between May 2016 and May 2017, because the first draft STPs were produced for June 2016, and published in December 2016. Due to different questions being asked in different waves, each variable has three pre-treatment measures, but these differ (for some, wave 4 is used, for others wave 6 is used). In each case the main difference-in-differences comparison is between waves 7 and 12. This allows the study to account for shifts in public opinion from late 2014 until shortly before STPs were developed, and for the main analysis to compare changes over approximately one year ending with the 2017 General Election.

8.2.3 Treatment and Control Groups

The treatment group consists of British Election Study Internet Panel respondents living in the 39 constituencies affected by planned hospital changes who responded to survey questions in the main pre- and post- treatment waves. The control group is respondents who lived in the remaining 494 English constituencies and also responded to those waves. Figure 8.1 shows which constituencies were affected.

Table 8.2 shows the number of respondents in each group; respondents who did not complete the survey in either the main pre-intervention or post-intervention wave were excluded, as were those who answered "Don't Know" in either survey wave. This was done to ensure a balanced panel over time. Effective Sample Size is also shown for the entropy balanced models.

Table 8.2: Sample size and effective sample size (ESS) by question, from the British Election Studies Internet Panel data

n (Control)	n (Treatment)	ESS (Control)	ESS (Treatment)	Variable
12938	983	10561.389	983	Evaluation of NHS change
11398	854	9349.938	854	Attitude to local cuts
12342	936	10169.042	936	Attitude to NHS cuts
13244	968	11064.849	968	Conservative vote

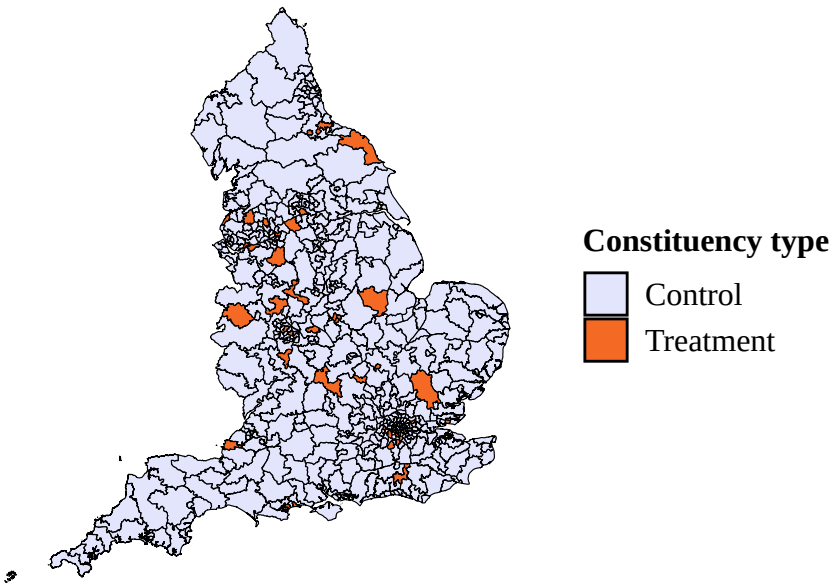

Figure 8.1: Map showing which parliamentary constituencies were affected by planned hospital changes

8.2.4 Outcomes: evaluations, attitudes and vote choice

In order to measure changes in evaluations, attitudes and vote choice that new information about planned hospital cuts could engender, I draw on four questions from the panel survey.

The first of these variables has been analysed in previous chapters. The 'NHS change' evaluation question captures the extent to which the survey respondent believes that the NHS is improving or getting worse. It is part of a question grid with the heading: "Do you think that each of the following are getting better, getting worse or staying about the same?". There are then three rows, the order of which is randomized. This variable measures the response to "The NHS" on a five-point Likert scale of "Getting a lot worse", "Getting a little worse", "Staying about the same", "Getting a little better" and "Getting a lot better", with "Don't Know" as an alternative answer.

This variable is distinct in that it attempts to elicit an objective evaluation of the direction of travel in the NHS. If new information about a local hospital or A&E department being down-

graded causes the respondent to revise their evaluations of health service provision in general, we would expect to see the treatment group become relatively more negative on this indicator compared to the control group. This would indicate that local conditions were being used as a heuristic to understand the general state of the NHS.

The second and third of these variables ('Local Cuts' and 'NHS Cuts') describe the extent to which the survey respondent believes that cuts to local services and the NHS, respectively, have gone too far, or not far enough. They are part of a question grid with the heading: "Do you think that each of these has gone too far or not far enough?". There are then five randomized rows. Responses are given on a five-point Likert scale of "Not gone nearly far enough", "Not gone far enough", "About right", "Gone too far" or "Gone much too far" with "Don't Know" as an alternative answer. The 'Local Cuts' variable is the response given to the question about "Cuts to local services in my area". For the 'NHS cuts' variable, the respondent is randomly assigned to be asked about either 'Cuts to NHS spending', or 'cuts to NHS services'.

Rather than evaluating the health service, these questions ask about voters' attitudes to service retrenchment, one focusing on the NHS (which could be interpreted to refer to national or local conditions) and one on 'local' services (which could also refer to council services such as libraries or social care). Because of the reference to 'cuts', these questions may be more likely to prompt the respondent to factor information about hospital changes into their response; they also invoke the respondent's views on the extent to which cuts are necessary, rather than simply ask about direction of travel.

The final variable ("Conservative vote") tests whether the voter intends to vote Conservative in that wave, and in the post-treatment wave, whether they did vote Conservative in the 2017 general election. It is intended to capture retrospective voting by showing whether differences in attitudes and evaluations that may emerge in the treatment group led to a change in voting patterns. If, however, attitudes and opinions change without a change in voting intention, this could signal that concerns about the local hospital were outweighed by other electoral considerations.

8.2.5 Parallel trend testing and entropy balancing

To establish whether pre-intervention trends were parallel in the raw data, four graphs were created showing average scores by treatment group and wave for each outcome variable. In each case, three pre-treatment waves are shown.

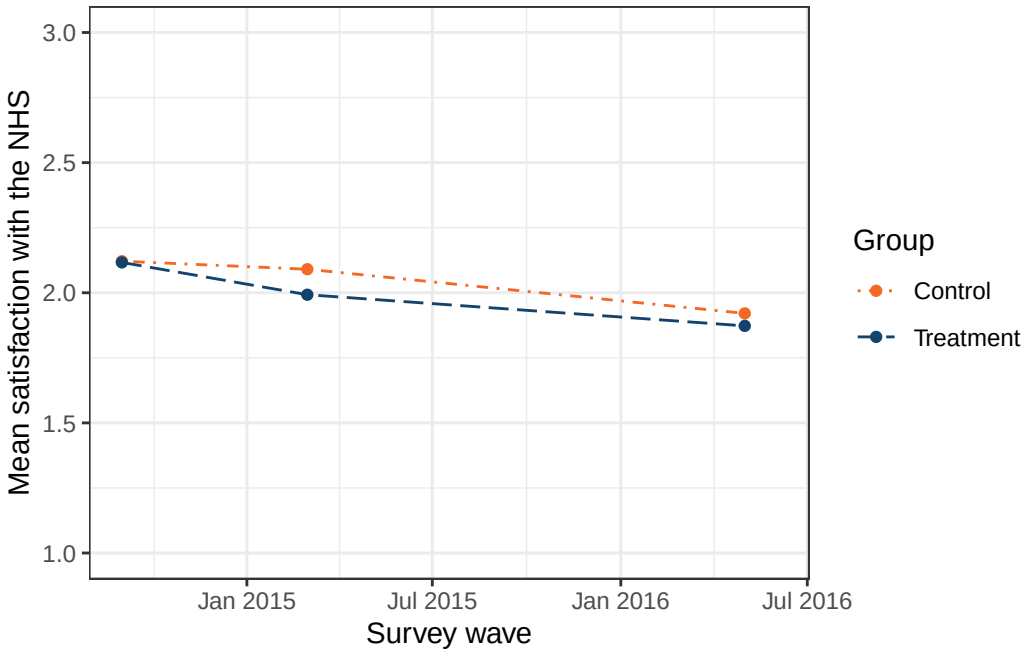

Figure 8.2: Pre-intervention trends in control and treatment groups for the NHS change outcome

Figure 8.2 shows that the average response for both control and treatment groups was close to 2 ("Getting a little worse"); there was a very small decline over time and a slight divergence between the groups.

For the Local Cuts outcome, Figure 8.3 shows that there was a small gap between the treatment group and control group emerged and then closed again during the pre-treatment period. The mean is closest to the response that cuts had "gone too far" (2 on the 5-point scale).

A similar trend is barely visible for the NHS cuts variable in Figure 8.4: the control group starts off at the same level as the treatment group and both reduce between survey waves, but the treatment group reduces slightly further. The mean view is again that cuts had gone too far.

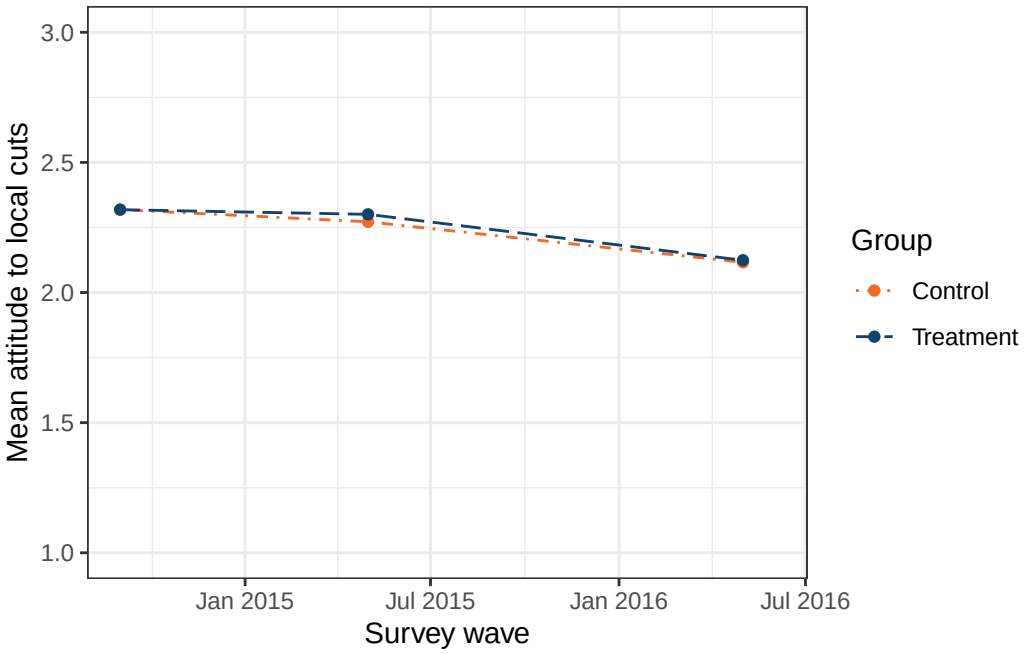

Figure 8.3: Pre-intervention trends in control and treatment groups for the local cuts outcome

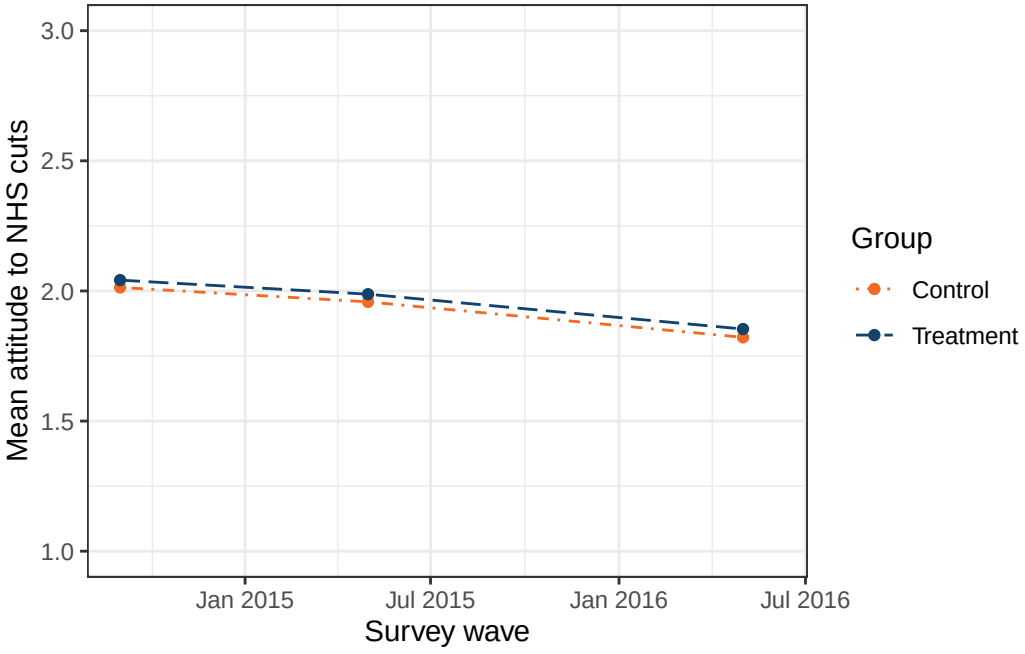

Figure 8.4: Pre-intervention trends in control and treatment groups for the NHS cuts outcome

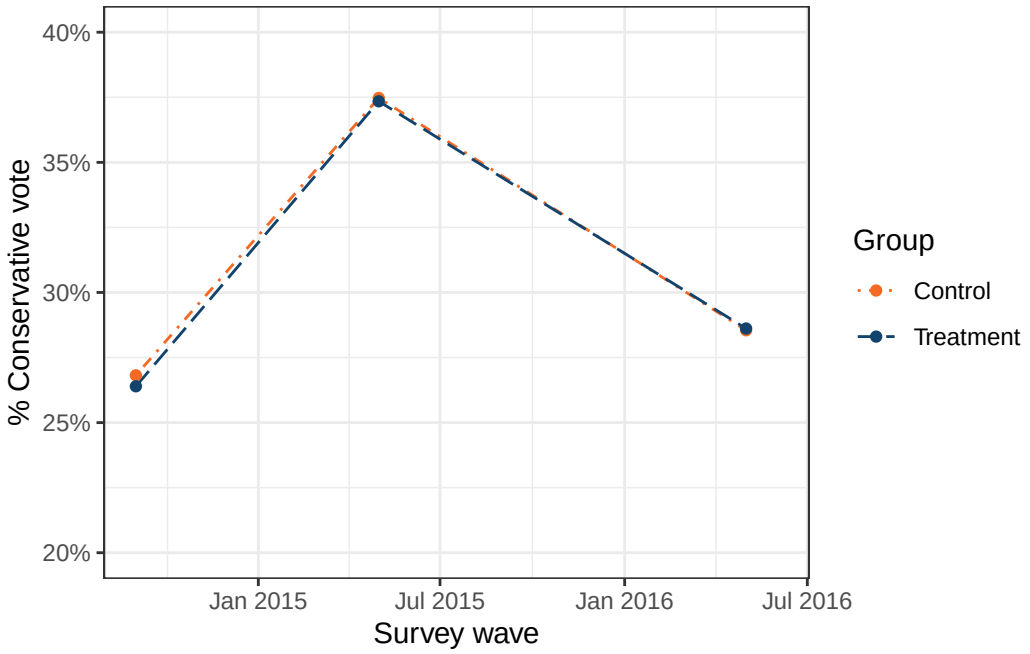

Figure 8.5: Pre-intervention trends in control and treatment groups for the Conservative vote outcome

The Conservative voting intention also had a slight difference in trends, with the control group generally slightly more Conservative, but the difference was greatest in wave 3, then wave 7, whereas in wave 6 both groups were almost identical.

As well as visual comparison, linear regression can be used to determine whether these differences in trends were statistically significant. Table 8.3 shows the results of regressing the interaction between treatment group membership and a binary dummy indicating that the observation was in wave 7 (the final pre-treatment wave), with the second-to-last pre-treatment wave (either wave 6 or wave 4, depending on the outcome variable) acting as reference category.

As the results in Table 8.3 show, the differences in pre-treatment trends ("Treatment group x time") ranged from small to very small; the largest was for the NHS change variable, and this was a difference of 5% of a step on the Likert scale. In all cases, the differences were not statistically significant at the $p = .05$ level.

Table 8.3: Results from parallel trend testing, showing there were no statistically significant differences detected between trends before the intervention

	NHS change	Local cuts	NHS cuts	Conservative vote
(Intercept)	2.090***	2.272***	1.958***	0.375***
	[2.070, 2.111]	[2.246, 2.298]	[1.934, 1.982]	[0.362, 0.388]
Treatment group	−0.098*	0.029	0.030	−0.001
	[−0.185, −0.011]	[−0.073, 0.131]	[−0.043, 0.102]	[−0.048, 0.045]
Time	−0.170***	−0.156***	−0.136***	−0.089***
	[−0.189, −0.151]	[−0.174, −0.137]	[−0.153, −0.119]	[−0.098, −0.080]
Treatment group x time	0.050	−0.021	0.002	0.002
	[−0.040, 0.140]	[−0.094, 0.052]	[−0.061, 0.065]	[−0.031, 0.035]
Num.Obs.	23 710	20 783	22 532	24 085
Std.Errors	by: onscode	by: onscode	by: onscode	by: onscode

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Although the statistical tests showed that the differences were not significantly different from zero, in order to add further confidence to the method, entropy balancing was used to establish parallel trends. A set of covariates were chosen for use in the entropy balancing process that have been established in the literature as influencing opinion about the NHS. These are: age (Calnan, Almond & Smith 2003), gender (Welch & Thomas 1988), ethnicity (Laycock 2009), social class operationalised as social grade of main income earner (Laycock 2009), voting intention (Edwards 2006) and party identity (Laycock 2009).

I also include EU referendum vote, probability of voting Conservative, and Government Office Region (GOR). The EU vote variable, while not indicated by any literature on retrenchment, was included because the political influence of the Brexit referendum in 2016 could act as a confounding factor, and Euroscepticism is not necessarily aligned with the left-right axis. Region is included as differences in service provision across geographies may affect opinion. Self-reported Conservative vote probability was included to ensure that a pre-treatment political balance was established between the control and treatment group.[3] Finally, in each analysis I also include the pre-intervention outcome variable. This establishes parallel trends by directly

[3]Except in the case of the Conservative vote share model, which already includes a binary Conservative vote share variable as the pre-intervention outcome variable to balance on (see following paragraph).

re-weighting the control group to match the treatment group's average opinions during the pre-intervention period (Cefalu, Vegetabile, Dworsky, *et al.* 2020; Hazlett & Xu 2018).

8.3 Results

8.3.1 Perception of changes in the NHS

To determine whether there was a significant difference in perception of the NHS between people living near hospitals that were planned to close under Sustainability and Transformation Plans and people living in other parts of England following the publication of these plans, the difference-in-differences estimator was calculated for control and treatment groups between the April 2016 (pre-intervention) and May 2017 (post-intervention) waves, using the unweighted dataset. Figure 8.6 also shows the previous data point for comparison (March 2015). Any 'treatment effect' – a change in perception only among those living near the affected hospitals – should be visible as a negative divergence from previous trends only among the treatment group.

Visually, there is very little difference in trends between the control and treatment groups. If anything, it appears that the control group became slightly less satisfied with the NHS relative to the treatment group. Figure 8.7 shows the difference in differences for the entropy balanced data for the same outcome, which appears to show no difference in perceived NHS change between the groups following the intervention.

To ascertain if this difference was statistically significant, a set of regression models were used to estimate the significance of the DID estimator (i.e. difference in trends). Following Ladd and Lenz (2009), I estimate a simple bivariate model, using the raw unweighted data, where the interaction between treatment status ('Retrenchment') and a dummy indicating the post-intervention survey wave ('time') produces an estimate of the difference-in-differences. Like Ladd and Lenz, I also estimate a multivariate model that includes a set of demographic and

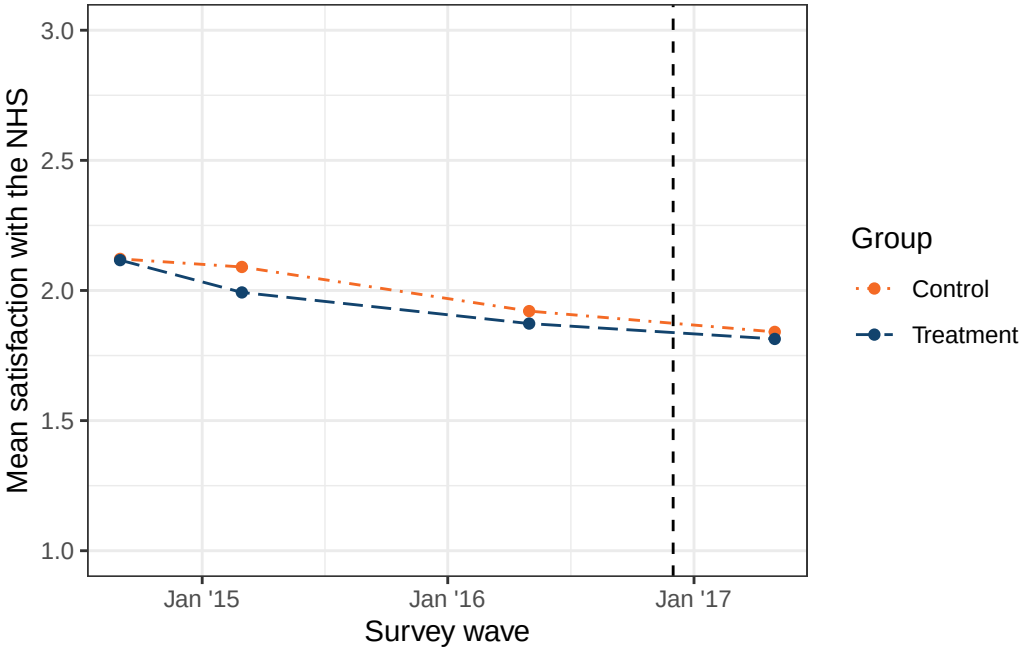

Figure 8.6: Difference-in-differences between control and treatment groups for the NHS change outcome, showing no significant treatment effect

political controls for the raw unweighted data; see the Methods section (Section 8.2) for the full list.[4]

Because of the importance of the parallel trends assumption for difference-in-differences analysis, I also estimate a bivariate model using the entropy balanced dataset.

Visual inspection of post-intervention trends and the regression results both indicate that there was no statistically significant deviation of trends between the control and treatment groups. The DID estimator varied between -0.007 for the entropy balanced model and 0.021 for the bivariate model, where 1 would indicate a difference in trend of one full category on the Likert scale. These estimates were not statistically significant at the p =.05 level in any of the three model specifications.

Figure 8.8 shows the estimated coefficients and confidence intervals for the key variables in

[4]I also performed the analysis using individual-level fixed effects, but this yielded very similar results without narrowing confidence intervals, so this method was dropped from the analyses.

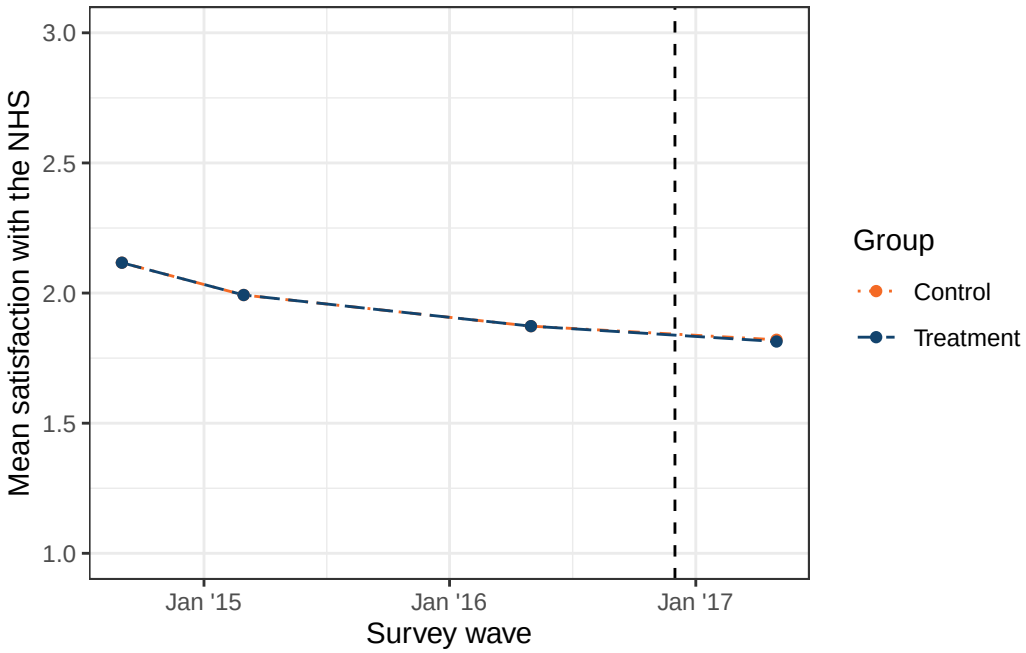

Figure 8.7: Entropy balanced difference-in-differences between control and treatment groups for the NHS change outcome using entropy balanced data, showing no significant treatment effect

Table 8.4: Outputs from difference-in-differences regression models, showing no significant treatment effect on views about whether the NHS is getting better or worse

	Bivariate	Multivariate	Entropy balanced
(Intercept)	1.921***	2.310***	1.873***
	[1.905, 1.936]	[2.205, 2.416]	[1.856, 1.890]
Treatment group	−0.048	−0.052+	0.000
	[−0.111, 0.015]	[−0.110, 0.005]	[−0.063, 0.063]
Time	−0.080***	−0.080***	−0.052***
	[−0.095, −0.065]	[−0.095, −0.065]	[−0.069, −0.036]
Treatment group x time	0.021	0.021	−0.007
	[−0.035, 0.077]	[−0.035, 0.077]	[−0.063, 0.050]
Num.Obs.	27 842	27 842	27 842
Std.Errors	by: onscode	by: onscode	by: onscode

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

each model, as well as three controls for the multivariate model. These were selected to facilitate comparison of the magnitude of the estimated difference-in-differences with other factors - for example, Labour party identity. The estimated coefficient for this control is -0.334, indicating that Labour-identifying voters voters were around 31% of a point on the five-point Likert scale less positive about the direction of change in the NHS, compared to the reference category (Conservative identifiers).

In the bivariate and multivariate models, the difference-in-differences coefficient estimate equates to 2% of a point on the Likert scale, with 95% confidence intervals from -3% of a step to 8% of a step. The estimate in the entropy balanced model is less than -1% of a step, with 95% confidence intervals from -6% to 5% of a step. The null result is therefore not entirely precise; in chapter 6, the effect of a one-week difference in national waiting times was estimated to be 4-5% of a step on the Likert scale, which is similar in magnitude to the outer limits of the confidence intervals for the difference-in-differences, depending on model choice. It is therefore difficult to be sure that there is no effect at all, but the results do indicate that if there is a true effect in either direction, it is very small.

8.3.2 Attitude to local cuts

To determine whether voters living in areas with hospitals threatened with closure noticed and reacted to the plans, I repeated these analyses on the survey question asking whether voters believed that cuts to local services had gone too far. The difference-in-differences was estimated for control and treatment groups between April 2016 and May 2017. Figure 8.9 and Figure 8.10 show the trends for the raw data, followed by the entropy balanced data.

Consistent with the results for the NHS change outcome measure, the difference in post-treatment trend between control and treatment groups appears negligible. This is confirmed by the regression results as shown in Table 8.5

None of the models yielded a statistically significant result at the p =.05 level. The DID estimate ranged between -0.002 (less than one percent of one step on the Likert scale) and zero; 95%

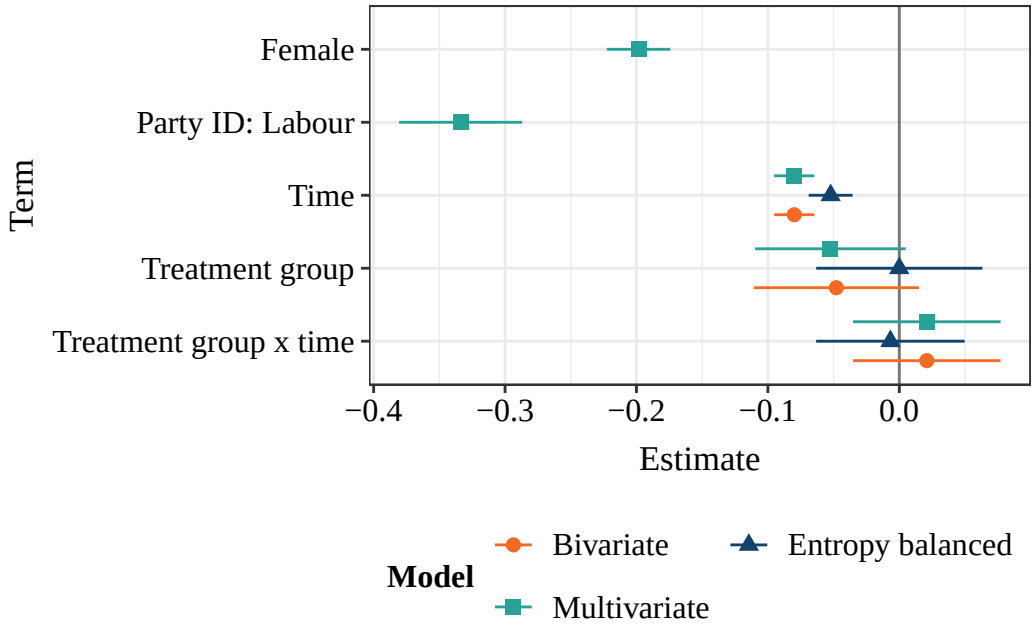

Figure 8.8: DID regression outputs showing estimated treatment effect for the 'NHS change' outcome, compared to selected covariates

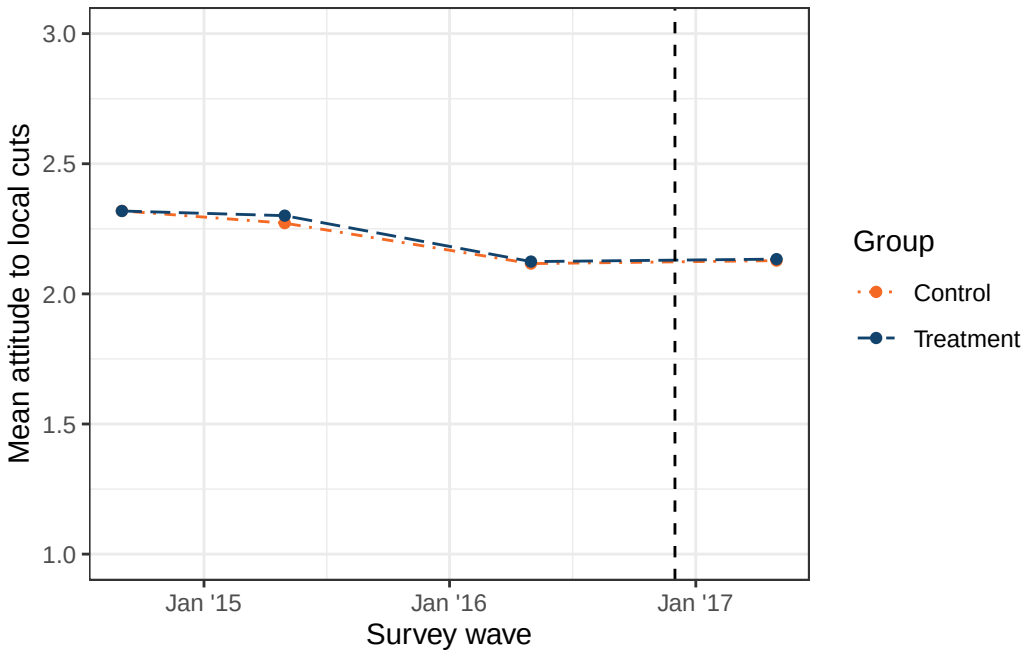

Figure 8.9: Difference-in-differences between control and treatment groups for the 'local cuts' outcome, showing no significant treatment effect

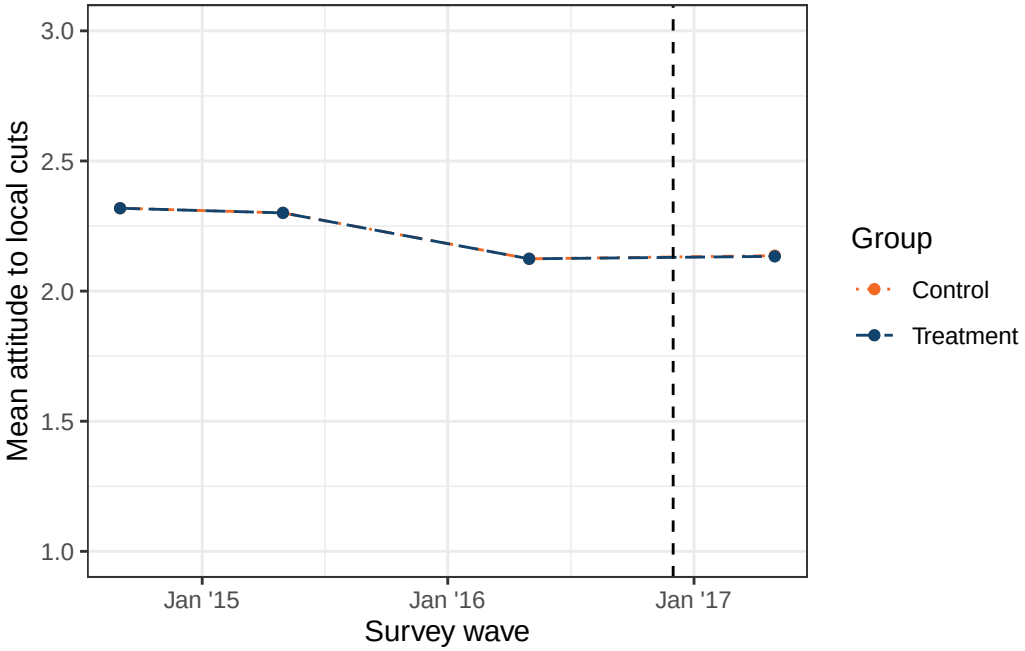

Figure 8.10: Entropy balanced difference-in-differences between control and treatment groups for the 'local cuts' outcome, showing no significant treatment effect

Table 8.5: Outputs from difference-in-differences regression models, showing no significant treatment effect on views about cuts to local services

	Bivariate	Multivariate	Entropy balanced
(Intercept)	2.116***	2.536***	2.124***
	[2.094, 2.138]	[2.406, 2.665]	[2.100, 2.148]
Treatment group	0.008	0.021	0.000
	[−0.067, 0.083]	[−0.035, 0.078]	[−0.075, 0.075]
Time	0.011	0.011	0.012
	[−0.003, 0.026]	[−0.003, 0.026]	[−0.005, 0.029]
Treatment group x time	−0.002	−0.002	−0.003
	[−0.059, 0.055]	[−0.059, 0.055]	[−0.060, 0.055]
Num.Obs.	24 504	24 504	24 504
Std.Errors	by: onscode	by: onscode	by: onscode

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

confidence intervals were between approximately -6% and 5% of a step on the Likert scale in all models. The regression effects plot gives a visual representation of the results along with selected covariates (Age, female, and point 10 on the left-right scale).

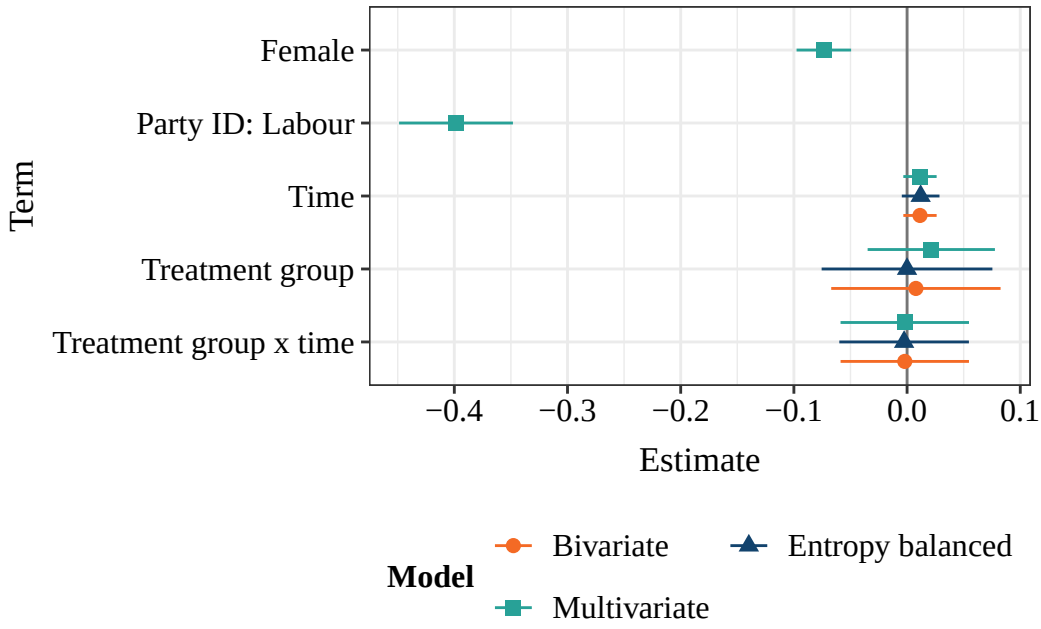

Figure 8.11: DID regression outputs showing estimated treatment effect for the 'local cuts' outcome, compared to selected covariates

8.3.3 Attitude to NHS cuts

Thirdly, to determine whether voters living in areas with hospitals threatened with closure noticed and reacted to the plans, I carried out the difference-in-differences analyses on the survey question asking whether voters believed that cuts to NHS services had gone too far was analysed. The difference in differences was estimated for control and treatment groups between April 2016 and May 2017. Figure 8.12 and Figure 8.13 show the unweighted and then entropy balanced trends.

Consistent with the previous outcome measures, the difference between control and treatment groups appears minor. This is confirmed by the regression results as shown in Table 8.6. None

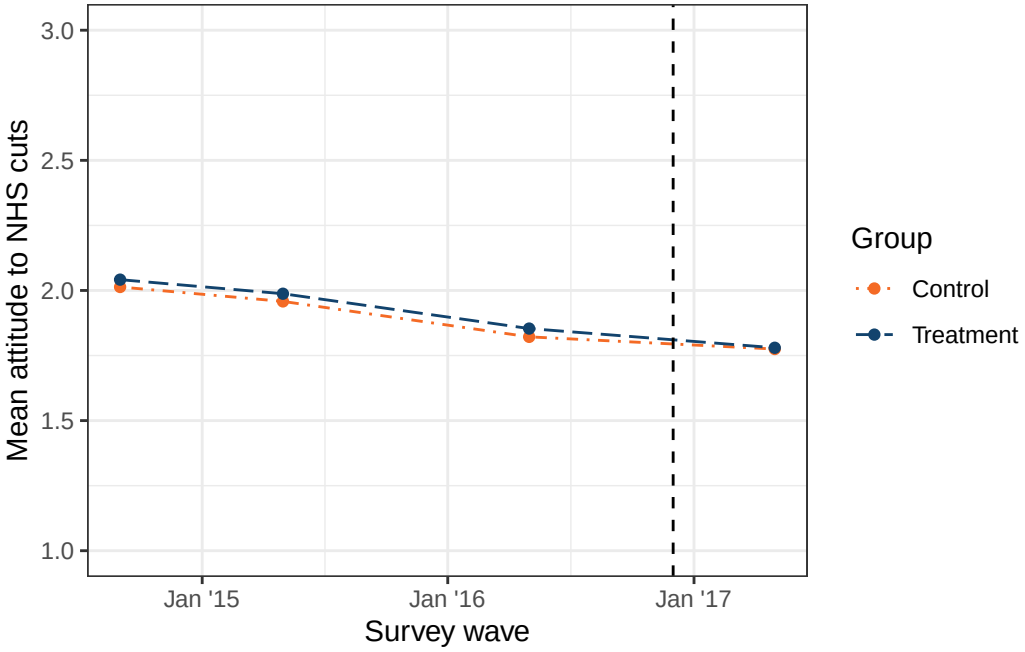

Figure 8.12: Difference-in-differences between control and treatment groups for the 'NHS cuts' outcome, showing no significant treatment effect

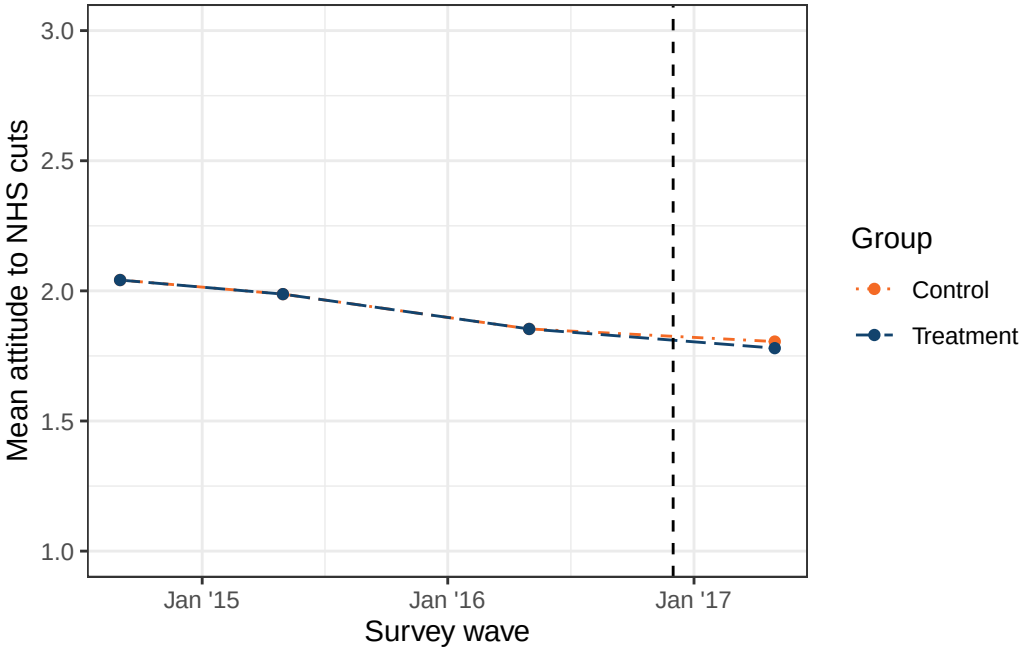

Figure 8.13: Entropy balanced difference-in-differences between control and treatment groups for the 'local cuts' outcome, showing no significant treatment effect

Table 8.6: Outputs from difference-in-differences regression models, showing no significant treatment effect on views about cuts to the NHS

	Bivariate	Multivariate	Entropy balanced
(Intercept)	1.822***	2.498***	1.854***
	[1.802, 1.841]	[2.378, 2.618]	[1.831, 1.876]
Treatment group	0.032	0.030	0.000
	[−0.037, 0.101]	[−0.032, 0.092]	[−0.070, 0.070]
Time	−0.047***	−0.047***	−0.049***
	[−0.060, −0.034]	[−0.060, −0.034]	[−0.064, −0.033]
Treatment group x time	−0.027	−0.027	−0.025
	[−0.079, 0.025]	[−0.079, 0.025]	[−0.077, 0.027]
Num.Obs.	26 556	26 556	26 556
Std.Errors	by: onscode	by: onscode	by: onscode

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

of the models yielded a statistically significant result at the p =.05 level. The DID estimate varied between -0.025 in the entropy balanced model and -0.027 in the bivariate model. The lower 95% confidence intervals around the DID estimate were -7% of a step on the Likert scale in all models, and the upper intervals were 2-3% of a step. The regression effects plot in Figure 8.14 gives a visual representation of the results along with selected covariates (Age, female, and Labour party identity).

Taking this result together with the results for previous questions, we have not found evidence for Secondary Hypothesis 1 ("*The publication of plans to cut hospital services resulted in worse opinions of government management of the NHS among people living near affected hospitals*").

8.3.4 Conservative voting intention

In order to determine whether people living near hospitals and A&E departments that were threatened with downgrade or closure changed their voting intention in response to the publication of Sustainability and Transformation Plans, the difference-in-differences estimator was then calculated for control and treatment groups between the April 2016 (pre-intervention) wave and voters' recalled vote following the 2017 general election, using a binary measure

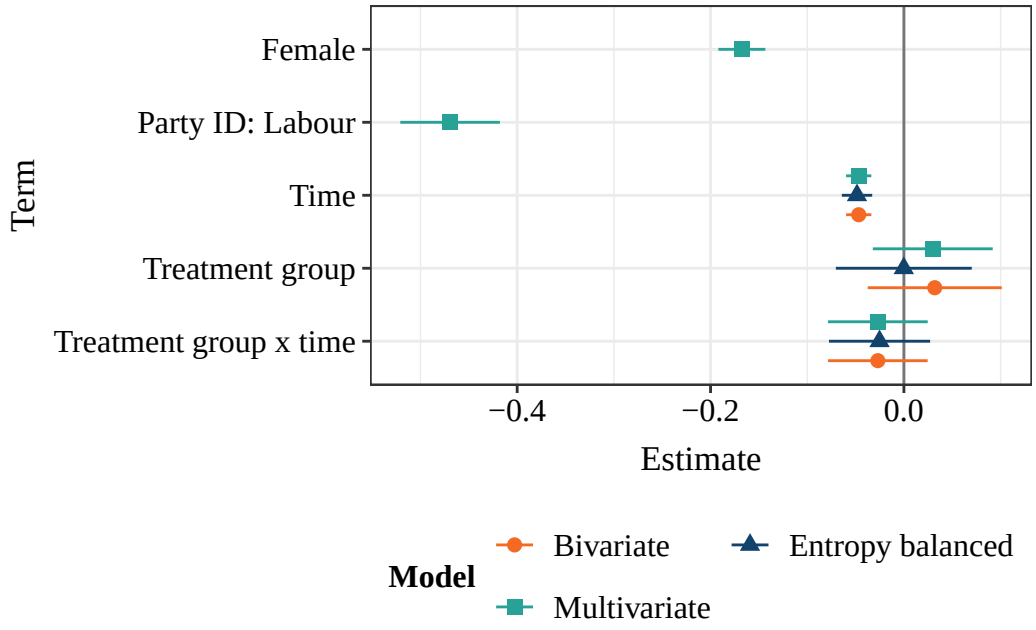

Figure 8.14: DID regression outputs showing estimated treatment effect for the 'NHS cuts' outcome, compared to selected covariates

of Conservative voting intention as the outcome measure. Unweighted and entropy balanced trends are shown in Figure 8.15 and Figure 8.16, respectively.

In both the raw and entropy balanced graphs, there appears to be a very small difference in trends between the control and treatment groups, with a lower percentage of the treated group intending to vote Conservative.

To ascertain if this difference was statistically significant, regression models were estimated; their outputs are shown in Table 8.7.

Visual inspection of post-intervention trends and the regression results both indicate that there was no significant deviation of trends between the control and treatment groups following the publication of the hospital plans for this outcome.

The DID estimate was not statistically significant at the p =.05 level in any of the three model specifications; the point estimate was a vote share difference of less than 0.1%, and the 95% confidence intervals were between -3% and 3-4%. As with the previous results, while a more

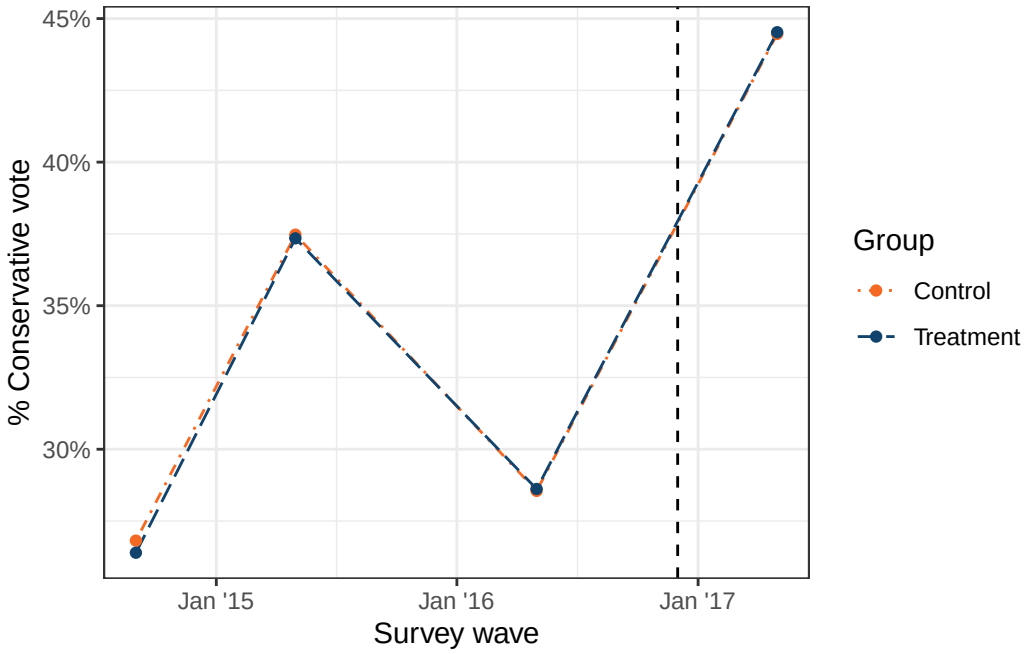

Figure 8.15: Difference-in-differences between control and treatment groups for the 'Conservative vote' outcome, showing no significant treatment effect

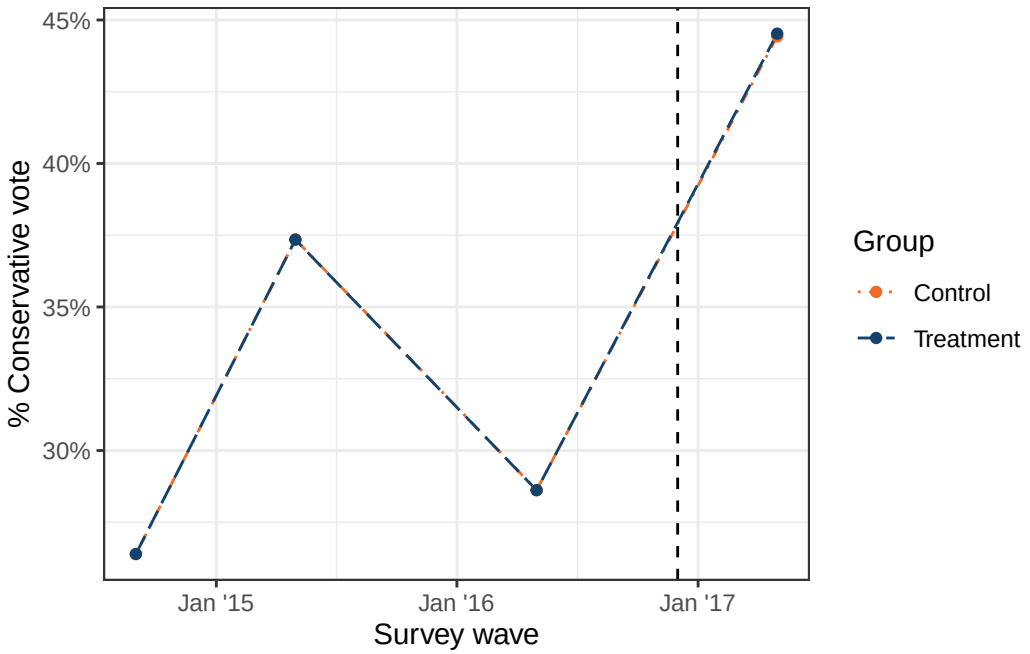

Figure 8.16: Entropy balanced difference-in-differences between control and treatment groups for the 'local cuts' outcome, showing no significant treatment effect

Table 8.7: Outputs from difference-in-differences regression models, showing no significant treatment effect on Conservative voting intention

	Bivariate	Multivariate	Entropy balanced
(Intercept)	0.285***	0.025	0.286***
	[0.276, 0.295]	[−0.014, 0.065]	[0.275, 0.297]
Treatment group	0.001	0.000	0.000
	[−0.036, 0.037]	[−0.036, 0.036]	[−0.037, 0.037]
Time	0.159***	0.159***	0.158***
	[0.150, 0.168]	[0.150, 0.168]	[0.148, 0.168]
Treatment group x time	0.000	0.000	0.001
	[−0.034, 0.034]	[−0.034, 0.034]	[−0.034, 0.036]
Num.Obs.	28 424	28 424	28 424
Std.Errors	by: onscode	by: onscode	by: onscode

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

precise estimate would have been preferable, there is no indication of a strong treatment effect. Figure 8.17 shows the estimated coefficients and confidence intervals for the key variables and selected controls for the multivariate model.

These results do not support Secondary Hypothesis 2 ("*The publication of plans to cut hospital services resulted in lower Conservative vote share among people living near affected hospitals*").

8.3.5 Robustness Checks

This section shows results from the robustness checks that were used to give greater confidence that the null results were not dependent on a particular model specification. As discussed in the Methods section (Section 8.2), one potential pitfall with using a difference-in-differences design on observational data is the risk of inaccurately defining the treatment group and therefore violating the stable unit treatment value assumption. In the context of this study, this would occur if people outside the 'treated' constituencies were also affected by the proposed closures, resulting in the 'control' group's trends being partially affected by the treatment.

To mitigate against this potential spillover effect, I recreated the main models using a broader definition of treatment. In this 'broad retrenchment' specification, I include the 39 constituen-

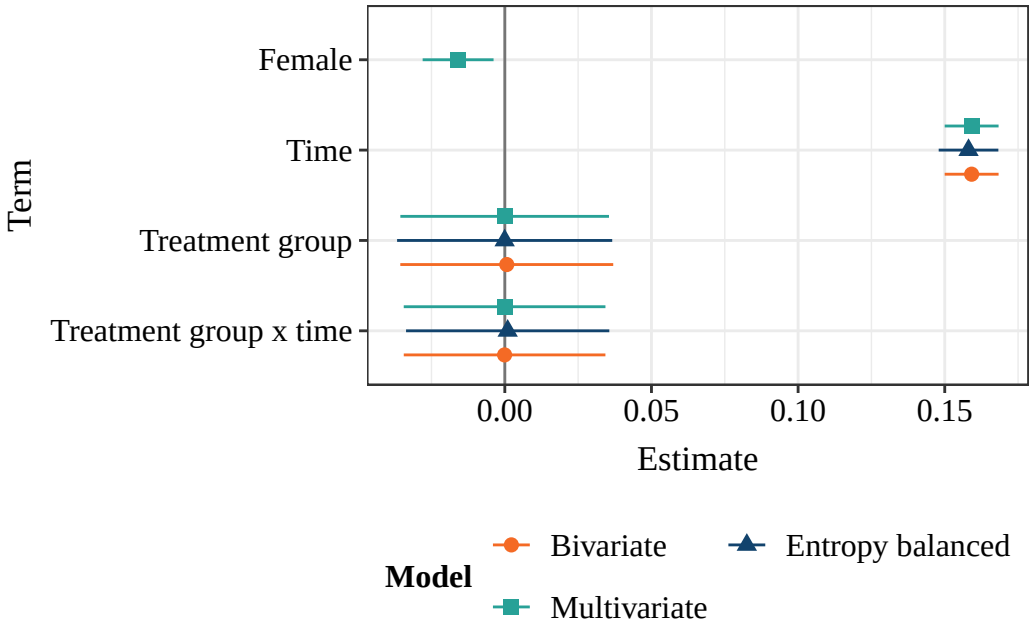

Figure 8.17: DID regression outputs showing estimated treatment effect for the 'Conservative vote' outcome, compared to selected covariates

cies with a hospital that was under threat, but also added any constituencies that share a land border with those 39. This definition assumes, therefore, that people living within a much larger radius of the hospital would be affected by its planned closure. This results in a total of 173 constituencies being counted as subject to the retrenchment treatment (32% of the English total).

However, it also seems reasonable to assume that different hospitals would have a differently sized 'footprint' in terms of where their patients live. Hospitals in England do not have defined 'catchment areas' - patients can choose where to initiate several kinds of non-urgent care . Given this, some people living in some 'neighbouring' constituencies may be affected, but others would not. Because there isn't a reliable way to differentiate between the two, I also run the analysis using the original single-constituency definition of the treatment group, but excluding anyone living in the neighbouring constituencies from either the control or the treatment group. If the effects of retrenchment are sometimes wider than one constituency but not always, this alternative specification reduces the risk of wrongly assigning people into

the control or treatment groups. I call this the 'no neighbours' treatment specification. 134 constituencies border at least one constituency with planned retrenchment and are therefore excluded from the 'no neighbours' analysis. Figure 8.18 shows which constituencies fall into which categories.

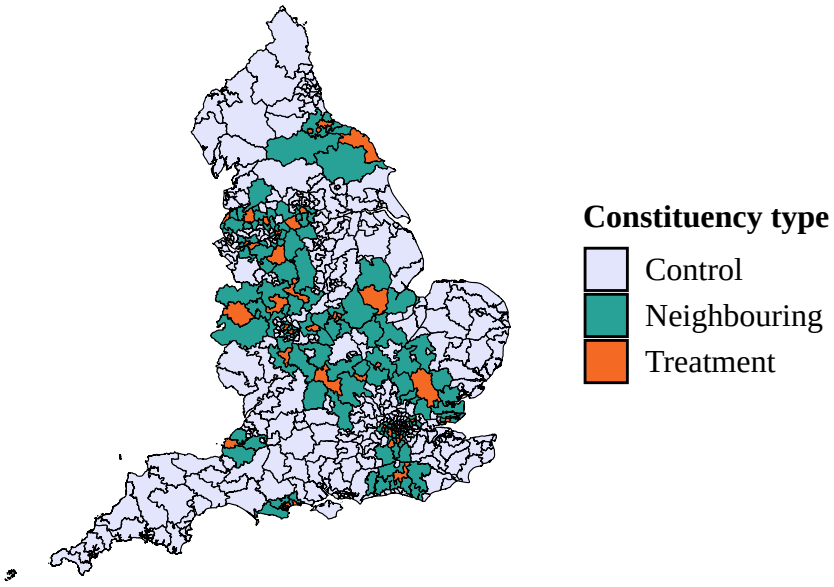

Figure 8.18: Map showing constituencies with planned hospital changes in the alternative specifications

As Figure 8.19 shows, these alternative specifications yielded results that were consistent with the findings of the main analysis. In each facet of the plot, difference-in-difference point estimates and 95% confidence intervals (clustered at constituency level) are displayed for each of the three model types (bivariate, multivariate, and entropy balanced).

To facilitate comparison across different specifications, models and variables, Figure 8.20 displays the estimates and confidence intervals shown above, ranked by point estimate.

Regardless of the specification, no point estimate for the treatment effect reached 0.05 or -0.05, which would equate to an average difference in trends of 5% of one step on a 5-point Likert scale, or in the case of the Conservative vote variable, a 5% difference in Conservative vote

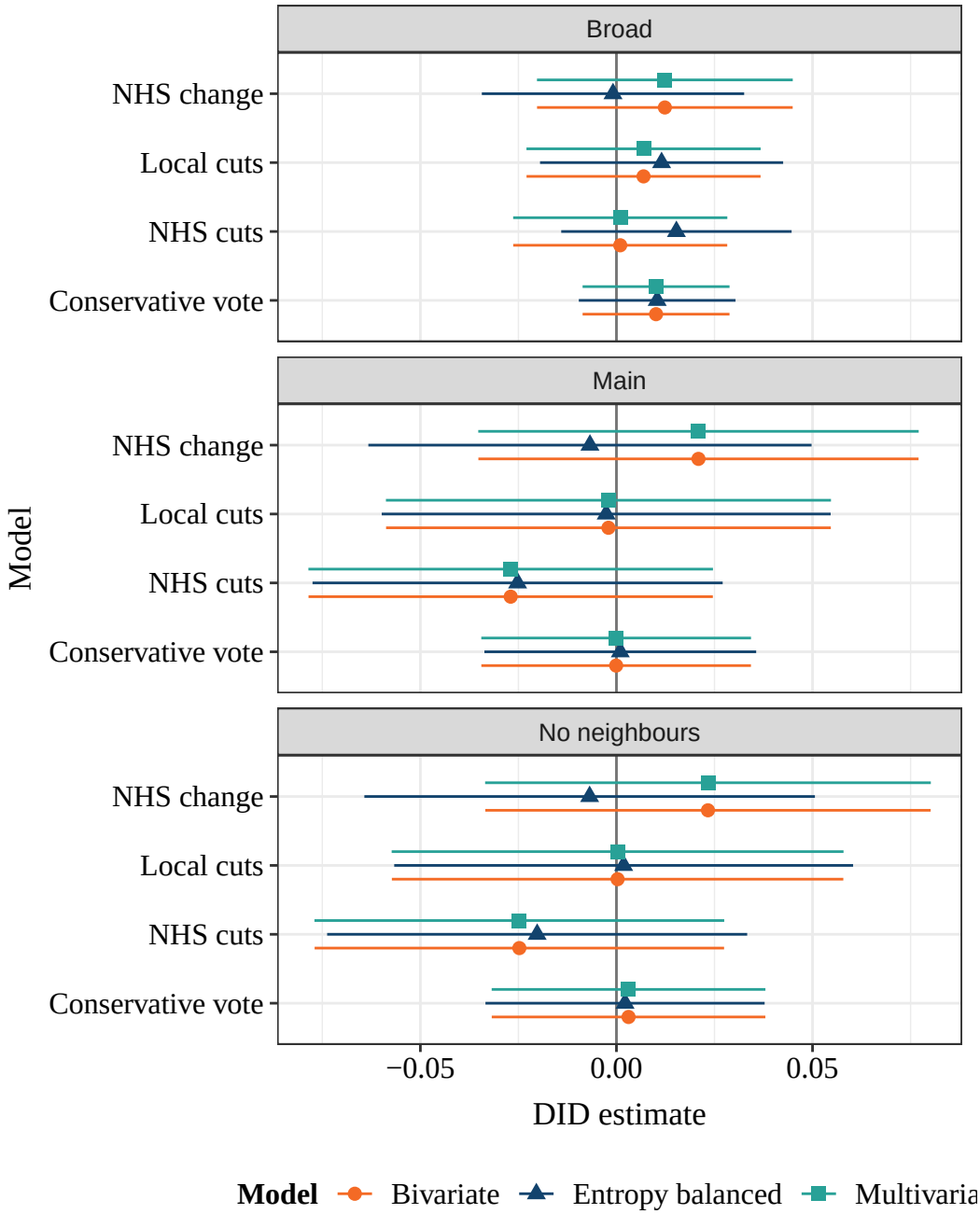

Figure 8.19: Results from difference-in-differences regressions for all models and specifications(faceted), showing no treatment effect estimate was statistically significant

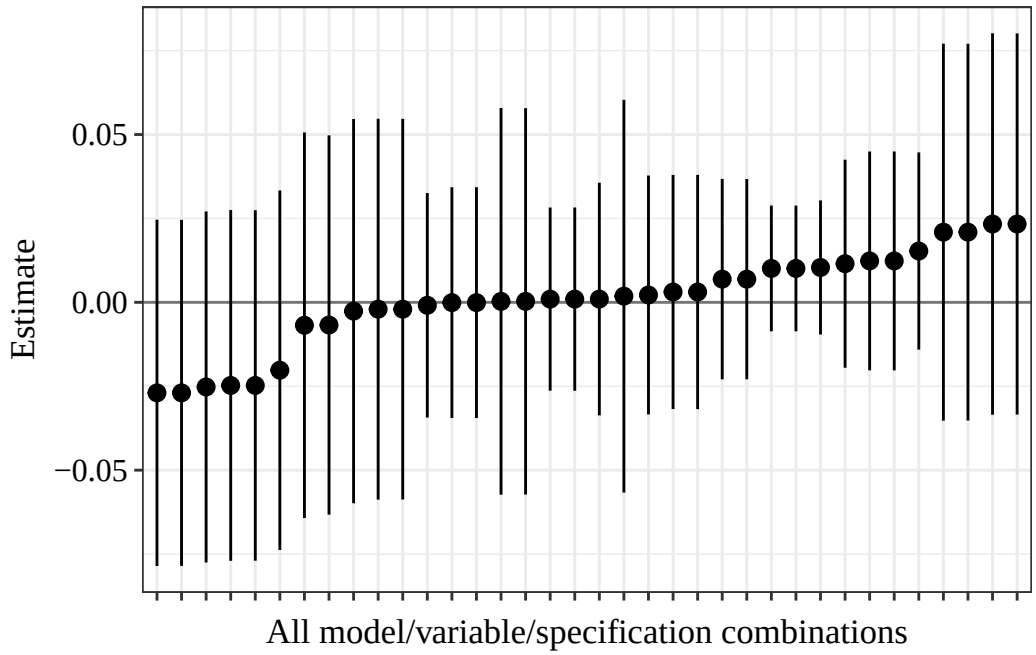

Figure 8.20: Results from difference-in-differences regressions for all models and specifications (ordered by estimated effect size), showing no treatment effect estimate was statistically significant

share. No upper or lower bound for the 95% confidence intervals of any estimate reached 0.1 or -0.1 (10% of one step on a 5-point scale or a 10% change in vote share). All estimates failed to achieve statistically significant at the $p = 0.05$ level. I therefore fail to reject the null hypothesis that there was no difference in trends between the control and treatment groups, regardless of specification.

The effect of introducing the different specifications tended to vary with the outcome variable under analysis. In all cases, the 'broad retrenchment' specification resulted in narrower confidence intervals, which is to be expected as the treatment group was considerably larger in that specification. The broad retrenchment specification also resulted in all of the estimates being closer to zero for the NHS Cuts and NHS Change variables, but this was not uniformly the case across the Conservative Vote and Local Cuts models. This may be because of random variation, or it may indicate lower geographical variation in the NHS-related outcome variables.

The 'no neighbours' specification, since it retained the same treatment group (but reduced the numbers in the control group), usually gave very similar estimates and confidence intervals to the standard specification used in the main analysis.

8.4 Discussion

This chapter aimed to increase our understanding of whether and how voters evaluate public services, and then, having assigned credit or blame, vote retrospectively; it also aimed to give us more information about the national-local dimension by testing the results of planned local service retrenchment. The design therefore tested Primary Hypothesis V (*"Health service retrenchment will result in retrospective voting against incumbent politicians"*), which was broken down into two more specific secondary hypotheses:

> HYPOTHESIS 1: People living in areas affected by planned hospital cuts will be more negative about government management of the NHS than others, after accounting for pre-STP differences and trends.

> HYPOTHESIS 2: People living in areas affected by planned hospital cuts voted Conservative at a lower rate than others, after accounting for pre-STP differences and trends.

This analysis was unable to detect any meaningful difference in trends between survey respondents living near hospitals affected by Sustainability and Transformation Plans and those elsewhere in England in terms of their NHS evaluations, attitudes towards cuts, or vote choice, across the treatment period. This was the case in simple bivariate difference-in-difference models, as well as when either adding control variables or using entropy balancing to re-weight the control group. The finding was also robust to changes in the geographical definition of the treatment group. In some cases the point estimates had the opposite sign than predicted by the theory - as would be expected if the true effect was zero.

If the assumptions of the research design hold, this means we can have a moderately high level of confidence that the plans' publication did not have a statistically significant effect on these views among local people generally. As a result, we would not expect there to be any effect on vote choice, since the first step in the retrospective voting process was not present (perception of difference in performance). This chapter therefore does not contribute to our understanding of the later stages of the process, other than to point to the difficulty of the prerequisite conditions being present (i.e. the first two stages of the process taking place). There are several potential explanations why these null results may have emerged, both technical and theoretical; the result could also be the result of a combination of several factors.

In terms of modelling practice, it may be that a small treatment effect existed, but that it was too small to be detected with effective sample sizes of less than 1000 in the treatment groups (in the main specification). This effect could be up to around 3-4% of a step on the Likert scale, depending on the outcome measure and model type. An effect of this size would be comparable to the effect of geographic variation in waiting times. Unfortunately, no alternative panel data exists with a bigger sample size for these research questions and time period. As an example, looking at the NHS cuts question using the main (same-constituency) specification, the 95% confidence intervals for the difference-in-differences in the entropy balanced data were

between -0.077 and 0.027, meaning the difference in trends between control and treatment groups was most likely between -8% and +3% of one step on the five-point Likert scale. This leaves open the possibility that hospital closure plans had a very small local treatment effect, similar in scale to the association between longer local waiting times and NHS evaluations found in a previous chapter.

A second model-related reason why a null effect may have emerged is that using constituency boundaries to represent which hospital is 'local' could be problematic. While better than the alternatives, it is possible that people could identify with a hospital in a neighbouring constituency or one further away, particularly in densely populated areas such as Greater London. This would mean that people are incorrectly categorized as to their treatment status, introducing bias into the results. However, anticipating this problem, the robustness checks demonstrated, that the null finding holds even when excluding neighbouring constituencies from the dataset.

Alternatively, it may be that there was no true effect, even a small one. Earlier in the chapter I wrote that validation of these hypotheses would be contingent on the hospital change plans being sufficiently salient, voters placing enough weight on local issues, and local cuts already being accounted for in voters' priors. Given that previous analyses in this thesis have found that there is a very small association between local performance and public opinion locally, and a much larger association between changes in national performance over time and public opinion, it seems consistent to imagine that any of these three factors could have reduced or eliminated any possible change in opinion that may have resulted from the announcement of the local hospital changes. Only a small proportion of voters may have been aware of the changes, they may have placed more weight on their view of the health service nationally, and they may have already 'priced in' the possibility of hospital cuts in their area.

I would argue that the assumptions made in this chapter's research design are reasonable. Therefore, we can be fairly confident that proposed changes to local hospital services in 2016 either had no effect, or a very small effect, on local public opinion at the 2017 General Election. The findings in this chapter therefore do not support Primary Hypothesis V (*"Health*

service retrenchment will result in retrospective voting against incumbent politicians"). Rather, it adds further support to the view that while differences in public service performance can alter people's satisfaction with those services, the extent to which this takes place depends on the context, with consistent and clear information about performance being necessary to . This implies that governments can downgrade public service provision on a local level without risking electoral punishment in those areas. The broader implications for accountability, as well as my findings from the previous empirical chapters, will be discussed further in chapter 9.

9 Conclusions and questions: How voters evaluate public services

In this final thesis chapter, I bring together key findings from across the empirical chapters and show how they relate to each other and to previously produced research, as well as discussing some areas that could benefit from being addressed in future research. I firstly reflect on the overall findings from this thesis and how they fit within the existing literature. Following this, I return to the list of primary hypotheses that I aimed to test, and summarise my findings for each one. I will then discuss some cross-cutting themes that were present in multiple thesis chapters, and finally, close the thesis by outlining some remaining questions and possible avenues for future study.

9.1 Looking back: locating my theory within the literature

At the beginning of this thesis, I wrote about how accountability for public service performance could be created through a process of retrospective voting. Governments - and the individual politicians who comprise them - should make choices with a view to maintaining good performance within the services they manage, because they know that if they do not, voters will recognise the downward trend in service performance, attribute blame to the government, and remove them from office at elections.

Is this simplified model reflective of reality? In the fourth chapter of this thesis, I described the theoretical underpinnings and empirical support for what I and others have termed the social-

psychological approach to voting behaviour. In this sociologically-informed view of politics, the voter's perception of social reality is not only informed, but *shaped* by their political identity. When an individual sees themselves as rooted in and belonging to a particular social group within society, and that social group is linked to a particular political party and set of values, it becomes very important that they maintain a coherent world-view that is compatible with belonging to that political-social coalition. As a result, far from acting as dispassionate consumers selecting politicians based on past performance, voters' perceptions of how governments are performing - including on the issue of public service provision - is largely a function of whether the government is being led by "their" party. As discussed in chapter 4, partisanship can not only result in voters paying attention to different facts, it can also cause them to process information about objective reality in different ways - and this is evidenced by numerous experimental studies (Bullock, Gerber, Hill, *et al.* 2013; Wagner, Tarlov & Vivyan 2014). From this perspective, therefore, it is our political views that influence our satisfaction with government policy, rather than vice versa.

While the social-psychological literature points out barriers to voters' ability to evaluate government performance, there is evidence from a range of contexts that shows government performance does, nevertheless, result in different political outcomes. While there is no universal causal relationship between performance and public opinion, the valence literature argues that voters take overall government competence into account when deciding their vote. The use of heuristics - cognitive short-cuts - allows even low-information voters to be able to take account of differences in the outcomes of government policy when voting.

In chapter 4, I outlined a theoretical framework that accepted that, for the most part, people's views on public service performance are strongly influenced by their partisanship. I called this their 'partisan-informed priors', which essentially means that government partisans will be more satisfied than opposition partisans, because of the partisan assumption that one's preferred party will deliver better outcomes than the alternative. However, I also accepted the contention from the valence voting approach that opinion change, and therefore retrospective voting, is possible under certain circumstances. Those circumstances arise when evidence that contradicts voters' partisan priors is credible and consistent enough that it cannot be dismissed

via motivated reasoning. The 'moderated learning' theory therefore acknowledges the roles of both partisanship and performance, and described how more credible routes for information to be communicated (such as personal experience of services) may be more likely to result in opinion change compared to low-credibility sources (such as media known to be critical of the voter's preferred party). Five primary hypotheses were spelled out; below, I recap these hypotheses and summarise the thesis' findings in relation to each.

9.2 Key empirical findings

> PRIMARY HYPOTHESIS I: GOVERNMENT PARTISANS WILL EVALUATE THE NHS MORE HIGHLY THAN OPPOSITION PARTISANS.

On the whole, the first primary hypothesis was supported by the evidence. This was tested in chapter 5, which focused on the role of partisanship, and found through survey data from 1983-2019 that supporters of the government of the day (government partisans) were more likely to be satisfied with how the National Health Service (NHS) was being managed than other respondents. The partisan gap in satisfaction differed over time, reducing from around 50% of a step on the British Social Attitudes (BSA) survey's 5-point Likert satisfaction scale to around 25% of a step post-1997. This confirmed that, while its role seems to have reduced over time, partisanship remains a key factor associated with people's evaluations of the NHS.

> PRIMARY HYPOTHESIS II: NON-PARTISANS' EVALUATIONS OF THE NHS WILL BE LOWER THAN THOSE OF GOVERNMENT PARTISANS, BUT HIGHER THAN THOSE OF OPPOSITION PARTISANS.

The second primary hypothesis was partially supported: before 1997, non-partisans' satisfaction was between that of government and opposition partisans. However, in the latter part of the BSA time series (1997-2019), people who did not identify with any party (non-partisans) tended to be relatively dissatisfied with the NHS, usually at the same level but sometimes even more dissatisfied than opposition partisans. Expectations around non-partisans were also

subverted in chapters 6 and 7, where non-partisans were also found to be less responsive to both differences in performance and to recent inpatient experience than partisan respondents (whether government or opposition). As flagged up in the literature review, some studies have found that non-partisan voters, rather than being dispassionate "floating voters", are typically less informed about politics and struggle to interpret links between specific policy positions and group identity (Achen & Bartels 2017; Converse 1962). These findings add evidence to that alternative hypothesis.

PRIMARY HYPOTHESIS III: VOTERS' EVALUATIONS OF THE NHS WILL BE BETTER WHEN SERVICE PERFORMANCE METRICS ARE BETTER

The third primary hypothesis relates to service performance metrics. To test this hypothesis, in chapter 6, I used median waiting times for hospital treatment from 1988 - 2019 to represent 'performance' of health services. I performed two tests: the first estimated whether people were more satisfied with health services in years where waiting times were shorter and less satisfied in years when they were longer. This test found that there was, as expected, a positive association between better waiting time performance and satisfaction with the health service. The second test was geographical: using a smaller set of panel data for 2016-2019 from the British Election Study (which has more detailed geographical information than the BSA), I matched panel respondents to their local NHS commissioning area and tested whether people living in areas with better waiting times evaluated the health service more highly than those living in areas with worse waiting times. I again found that there was an association in the expected direction, although in this case it was a substantively very small association. As a result, I was able to corroborate the third primary hypothesis and determine that voters do indeed give better evaluations of the NHS when service performance metrics are better; this implies that voters are able to perceive and understand signals in the information environment about NHS performance.

PRIMARY HYPOTHESIS IV: RECENT SERVICE USE WILL AFFECT VOTERS' EVALUATIONS OF THE NHS

My theory of moderated learning posited that personal contact with services would be a credible source for information about public service performance. As such, there was a theoretical expectation that recent experience of NHS inpatient services would result in voters re-evaluating their evaluations of service performance. Because services might be better or worse than expected, it was not clear whether NHS contact would reduce partisan gaps or merely move people in parallel towards the true underlying value of performance. In fact, in both cross-sectional (BSA) and panel (Understanding Society) data, having recent inpatient experience was associated with better service evaluations (of the NHS and local medical services, respectively) among partisans, though not necessarily among non-partisans. The use of panel data and within-person comparisons allows us to be more confident that this difference is a result of having recent experience of the health service, rather than simply reflecting existing differences between individuals; this is a significant methodological improvement on previous analyses that were based on repeated cross-section survey data. Interestingly, there was some evidence that non-partisans were less likely than partisans to update their views in response to recent inpatient experience - which is perhaps suggestive of partisan differences in engagement leading to less responsiveness among non-partisans.

While these findings may indicate that voters updated their views towards a higher 'true' performance value, it is also possible that they indicate that recent health service experience is framed differently - triggering more social or emotional mental processes, rather than partisan political ones. In order to truly understand whether voters with recent service experience are picking up on performance signals earlier or simply expressing gratitude for the help they received, more panel data from periods with different levels of underlying NHS performance is needed. The fourth primary hypothesis is therefore supported, but the theoretical implications are slightly ambiguous.

> PRIMARY HYPOTHESIS V: HEALTH SERVICE RETRENCHMENT WILL RESULT IN RETROSPECTIVE VOTING AGAINST INCUMBENT POLITICIANS

The final of the five primary hypotheses laid out in chapter 4 was that *provision of services*, as well as differences in performance metrics such as waiting lists, could lead to opinion change

and ultimately retrospective voting against incumbents. Drawing on findings from the welfare state literature, I discussed why the case of England in 2016-17 could potentially have yielded an example of punishment voting on the issue of health service provision. In late 2016, plans were published to downgrade or consolidate hospital services in some parts of the country; this was soon followed by the 2017 UK general election. Using panel data, I was able to estimate whether people living in areas affected by these plans updated their opinions (e.g. becoming more negative about cuts, becoming more negative about the NHS' direction of travel) or voting intention, compared to people living in unaffected areas. Using a quasi-experimental difference-in-differences approach to the question, I found that on none of the four metrics (two questions about attitudes, one NHS evaluations question, and a voting intention measure) was there any statistically significant effect among people living near the affected hospital sites. This means that the four-step accountability process (perception of change, attribution of credit or blame, change in vote choice, change in politicians' behaviours) may have broken down at the initial perception stage: voters living in areas affected by the plans were no more likely than anyone else to believe that local cuts or cuts to NHS services had gone too far. While it is possible that this result is because of modelling choices or statistical imprecision, I consider that it is more likely that voters did not perceive and process the information about the changes, perhaps as a result of the dominance of national news and other highly salient political issues.

9.3 Connecting the dots across chapters

What are the common threads running through these hypotheses and the other secondary findings outlined in this thesis? It is clear that although some secondary hypotheses did not find support (sometimes due to sample size limitations), there was support for the idea that voters can and do evaluate healthcare performance in England. To do this, they take in information signals from a range of sources - including their partisan beliefs, as well as personal experience and information about service performance. The influence of partisanship, while important, does not 'crowd out' the influence of performance metrics or recent service contact.

Arguably the most important data visualisation in this thesis is Figure 6.6 (which is reproduced below as Figure 9.1), because it demonstrates this balance plays out at an aggregate party group level over time. The visualisation compares average satisfaction within party groups from the British Social Attitudes survey with a relative measure of waiting times (2009-indexed median waiting times averaged from January to June in each year). Two features in particular show this:

Firstly, there is a visible association between NHS performance (operationalised as median waiting times) and average satisfaction with the NHS. It is not a one-to-one relationship, but the trends are very similar, showing:

- a rise and then a fall in the 1990s,
- a sustained, consistent rise during the 2000s;
- and a gradual small deterioration after 2010.

This general association was shown to be robust to the inclusion of demographic controls in chapter 5, meaning that the relationship was very unlikely to be an artefact of demographic shifts within parties, but rather a real change in opinions. This over-time association was complemented by the finding of a cross-geographical relationship between performance and satisfaction, albeit on a smaller scale.

Secondly, partisanship is clearly associated with satisfaction throughout the period. While each party group is clearly influenced by national waiting time performance, it is evident that the most satisfied group in each government period is government partisans - a finding confirmed by multivariate regression analysis. After the mid-2000s, non-partisans are often less satisfied than opposition partisans, perhaps reflecting less engagement and responsiveness to change among this group.

There are also some features of the graph that suggest other changes in the information environment or political context are important in ways not fully explained by this thesis. In particular, the reduction in the partisan gap after 2000 may have been caused by the improvement

in waiting times, but this is not clear since there is no previous period of higher performance to compare it to.

The difference between non-partisans and partisans (whether pro-government or pro-opposition) was a common theme across several of the analyses in this thesis. As Figure 9.1 and chapter 5's regression results showed, while non-partisans' evaluations of the NHS were between those of the main two partisan groups in the 1990s, after that point they have similar dissatisfaction levels to opposition partisans. This is contrary to a simple 'partisan screen' interpretation, which would have predicted that non-partisans would be more satisfied than opposition partisans, since opposition partisans would downplay positive news about government performance. In chapters 6 and 7, I found that while non-partisans are responsive to differences in performance, personal experience of health services did not have a statistically significant effect on their opinions of the NHS - whereas partisan voters with recent inpatient experience were more satisfied than those without.

As discussed in the relevant chapters, this is not entirely unexpected based on previous empirical findings that non-partisans tend to be less informed and less engaged in the political process (Achen & Bartels 2017; Converse 1962). Based on these analyses, we should therefore think of non-partisans as being a separate voter group that is less engaged and may be less responsive to information about performance, compared to partisans. It is unclear at which stage or stages of the four-step accountability chain this breakdown happens. It may be that non-partisans receive and perceive fewer performance signals (e.g. because of being less inclined to consume news); it may also be because they do not have the necessary subject knowledge to make effective links between the conditions they observe and political choices that enables the attribution of credit or blame in the second step of the accountability chain.

In summary, this thesis offers evidence that, consistent with theories of moderated learning, voters draw on information from personal experience and other sources to evaluate public service performance. They update their views based on what they view as credible information, which has the result that while partisan gaps persist, there is an overall association between better performance, especially at the national level, and higher satisfaction with health services.

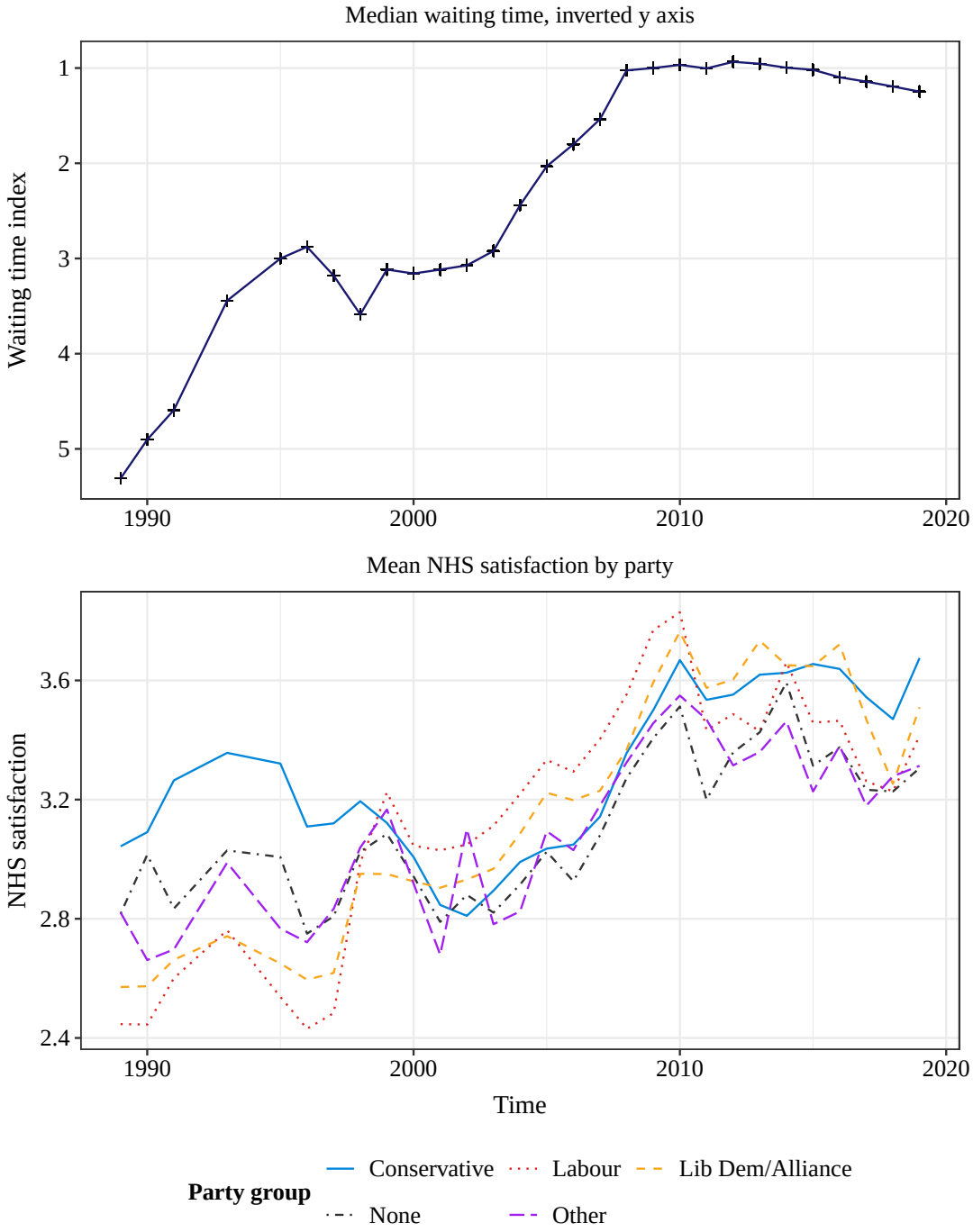

Figure 9.1: Line graphs comparing mean NHS satisfaction by party from BSA data and median waiting times in each year.

The circumstances in which this results in retrospective voting against governments with poor public service performance remains an open question.

9.4 Looking forward: questions to consider

In the discussion above I have reached some conclusions that I hope provide answers to the most important questions posed in this thesis, and add to the evidence base around how voters evaluate public services. There are, however, also some questions that arise from the findings in this research. I outline these below.

Firstly, the role of personal contact with public services in evaluations of public services remains unclear. As previous literature shows and chapter 7 confirmed, voters with recent experience of public services tend to be more satisfied with services than those without, even within the same partisan groupings; however, the reason why this is the case is not evident. There are two general explanations that I have considered: one is that personal experience is processed in a way that is qualitatively different from other performance information. It may trigger a more emotional, less political framing that both pushes past the partisan 'perceptual screen' and results in uniformly more positive views. The other explanation is that personal experience is a good source of performance information, but that the analysis in this thesis was primarily based on data from a period of time (the late 2000s) when performance was improving. If people with personal experience act as a 'canary in the coal-mine', picking up on changes in performance before others, then it follows logically that they would be more satisfied when performance is improving. Unfortunately, due to the relevant questions not being asked of many respondents in later years, there is not enough statistical power to determine conclusively whether contact results in lower satisfaction when performance is deteriorating.

Future research into this could therefore be fruitful if sufficient data is gathered on both service use and satisfaction in a context of worsening NHS performance. If such a study found that voters with recent experience of services were more satisfied than others, this would lend support to the first explanation. If not, this would lend support to the second explanation -

which would be another piece of evidence for a retrospective voting approach to public opinion in this arena.

Chapter 8 attempted to measure differences in trends in attitudes, evaluations and vote choice among people affected by local hospital retrenchment plans and others. In all cases, across various model designs and specifications, treatment effects were estimated as not statistically different from zero. This indicated a lack of evidence for the thesis that voters living near hospitals threatened with closure would engage in retrospective punishment voting by perceiving the change, attributing blame to incumbent politicians, and voting against them.

At the end of that chapter I outlined some potential reasons why this null finding may have emerged. I argued that the modelling assumptions and statistical set-up of the study were reasonable and would have found a small-to-moderate treatment effect if one existed, and therefore, the reasons are more likely to be substantive than technical. In particular, voters may have been unaware of the plans due to the dominance of the national news paradigm, and the unusually high levels of political salience during the post-referendum, pre-Brexit period.

As shown in chapter 2, the consensus in the literature is that retrospective voting is a contingent phenomenon that emerges in some contexts, but not others; and the most recently published research continues to show a mixed picture. To give two examples from papers published in 2023: a study with a similar research design exploring potential punishment effects for school and hospital closures in Denmark found that, while mayors were punished in local elections for school closures in some areas, there was no punishment voting against national governments for hospital closures (Nyholt 2023); by contrast, a study in Turkey found that introducing more local health services increased incumbents' vote share in those areas (Adiguzel, Cansunar & Corekcioglu 2023). Taking the evidence from chapter 8 together with the positive findings on performance in chapter 6 suggests that voters are more likely to be able to evaluate health service conditions at a national level than at a local level. Further evidence from the English context would be helpful to determine whether this is also the case in relation to other types of public services.

References

Achen, C.H. and Bartels, L.M. (2017) *Democracy for realists*. [Online]. Princeton University Press. Available at: https://www-degruyter-com.ezphost.dur.ac.uk/document/doi/10.1515/9781400888740/html.

Adiguzel, F.S., Cansunar, A. and Corekcioglu, G. (2023) Out of sight, out of mind? Electoral responses to the proximity of health care. *The Journal of Politics*. [Online] 85 (2), 667–683. Available at: doi:10.1086/722040.

Ahmed, F., Burt, J. and Roland, M. (2014) Measuring patient experience: Concepts and methods. *The Patient-Patient-Centered Outcomes Research*. [Online] 7 (3), 235241. Available at: http://link.springer.com/article/10.1007/s40271-014-0060-5.

Anderson, C.D. (2006) Economic voting and multilevel governance: A comparative individual-level analysis. *American Journal of Political Science*. [Online] 50 (2), 449463. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2006.00194.x/full.

Anderson, C.J. (2007) The end of economic voting? Contingency dilemmas and the limits of democratic accountability. *Annual Review of Political Science*. [Online] 10 (1), 271–296. Available at: doi:10.1146/annurev.polisci.10.050806.155344.

Angrist, J.D. and Pischke, J.-S. (2014) *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Angrist, J.D. and Pischke, J.-S. (2008) *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Appleby, J. and Roberts, C. (2012) *Health: Satisfaction with the NHS* [Online]. Available at: https://www.bsa.natcen.ac.uk/latest-report/british-social-attitudes-30/health/satisfaction-with-the-nhs.aspx.

Arel-Bundock, V. (2023a) R package version 0.14.0. *Marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests.* [Online]. Available at: https://vincentarelbundock.github.io/marginaleffects/.

Arel-Bundock, V. (2023b) *The marginal effects zoo (0.14.0) - 25 standard errors* [Online]. Available at: https://marginaleffects.com/articles/uncertainty.html.

Arel-Bundock, V., Gassen, J., Eastwood, N., Huntington-Klein, N., et al. (2023) *Modelsummary: Summary tables and plots for statistical models and data: Beautiful, customizable, and publication-ready.* [Online]. Available at: https://cran.r-project.org/web/packages/modelsummary/index.html.

Asthana, A. and Elgot, J. (2017) Theresa may ditches manifesto plan with 'dementia tax' u-turn. *The Guardian.* [Online] Available at: https://www.theguardian.com/society/2017/may/22/theresa-may-u-turn-on-dementia-tax-cap-social-care-conservative-manifesto.

Bailey, J. (2019) The fact remains: Party ID moderates how voters respond to economic change. *Electoral Studies.* [Online] 61102071. Available at: doi:10.1016/j.electstud.2019.102071.

Barer, D. (2016) The NHS: national religion or national football? *BMJ.* [Online] 352i1023. Available at: doi:10.1136/bmj.i1023.

Bartels, L.M. (2002) Beyond the running tally: Partisan bias in political perceptions. *Political Behavior.* [Online] 24 (2), 117150. Available at: https://www.jstor.org/stable/1558352.

Batchelor, C., Owens, D.J., Read, M. and Bloor, M. (1994) Patient satisfaction studies: Methodology, management and consumer evaluation. *International Journal of Health Care Quality Assurance.* [Online] 7 (7), 2230. Available at: http://www.emeraldinsight.com/doi/abs/10.1108/09526869410074720.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software.* [Online] 67 (1), 1–48. Available at: doi:10.18637/jss.v067.i01.

Bavel, J.J.V. and Pereira, A. (2018) The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences.* [Online] 22 (3), 213–224. Available at:

doi:10.1016/j.tics.2018.01.004.

Bélanger, É. and Meguid, B.M. (2008) Issue salience, issue ownership, and issue-based vote choice. *Electoral Studies*. [Online] 27 (3), 477–491. Available at: doi:10.1016/j.electstud.2008.01.001.

Bendor, J. (2003) HERBERTA. SIMON: Political Scientist. *Annual Review of Political Science*. [Online] 6 (1), 433–471. Available at: doi:10.1146/annurev.polisci.6.121901.085659.

Bennett, A.J. (2013) *1992: "It's the economy, stupid!"*. In: [Online]. Palgrave Macmillan US. pp. 123–159. Available at: doi:10.1057/9781137268600_5.

Bennett, G. (1996) "Camera, Lights, Action!": The British General Election 1992 as Narrative Event. *Folklore*. [Online] 107 (1-2), 94–97. Available at: doi:10.1080/0015587x.1996.9715921.

Berge, L., Krantz, S. and McDermott, G. (2023) *Fixest: Fast fixed-effects estimations*. [Online]. Available at: https://cran.r-project.org/web/packages/fixest/index.html.

Bernardi, L. (2018) Policy Responsiveness and Electoral Incentives: A (Re)assessment. *Political Behavior*. [Online] Available at: doi:10.1007/s11109-018-9490-4.

Black, N. (1992) Jennifer's ear: airing the issues. *Quality and Safety in Health Care*. [Online] 1 (4), 213–214. Available at: doi:10.1136/qshc.1.4.213.

Bleich, S.N., Özaltin, E. and Murray, C.J. (2009) How does satisfaction with the health-care system relate to patient experience? *Bulletin of the World Health Organization*. [Online] 87271–278. Available at: doi:10.1590/S0042-96862009000400012.

Books, J. and Prysby, C. (1999) Contextual Effects on Retrospective Economic Evaluations The Impact of the State and Local Economy. *Political Behavior*. [Online] 21 (1), 1–16. Available at: doi:10.1023/A:1023371328591.

Bowers, M.R., Swan, J.E. and Koehler, W.F. (1994) What attributes determine quality and satisfaction with health care delivery? *Health care management review*. [Online] 19 (4), 4955. Available at: http://journals.lww.com/hcmrjournal/Abstract/1994/01940/What_attributes_determine_quality_and_satisfaction.6.aspx.

Boyd, A., Cornish, R., Johnson, L., Simmonds, S., et al. (2017) Understanding hospital episode statistics (HES). *London, UK: CLOSER*.

Bullock, J.G., Gerber, A.S., Hill, S.J. and Huber, G.A. (2013) *Partisan bias in factual beliefs about*

politics

Callaway, B. and Sant'Anna, P.H.C. (2021a) *Did: Difference in differences* [Online]. Available at: https://bcallaway11.github.io/did/.

Callaway, B. and Sant'Anna, P.H.C. (2021b) Difference-in-Differences with multiple time periods. *Journal of Econometrics*. [Online] 225 (2), 200–230. Available at: doi:10.1016/j.jeconom.2020.12.001.

Calnan, M., Almond, S. and Smith, N. (2003) Ageing and public satisfaction with the health service: An analysis of recent trends. *Social science & medicine*. [Online] 57 (4), 757762. Available at: http://www.sciencedirect.com/science/article/pii/S027795360300128X.

Campbell, A. (1960) *The american voter*. New York ; London, John Wiley.

Campbell, A. (1954) *The voter decides*. [Online]. Row, Peterson. Available at: http://hdl.handle.net/2027/mdp.39015005319135.

Carr-Hill, R.A. (1992) The measurement of patient satisfaction. *Journal of Public Health*. [Online] 14 (3), 236–249. Available at: http://jpubhealth.oxfordjournals.org/content/14/3/236.

Castles, F.G., Leibfried, S., Lewis, J., Obinger, H., et al. (2012) tex.ids= castles2012. *The oxford handbook of the welfare state*. OUP Oxford.

Cefalu, M., Vegetabile, B.G., Dworsky, M., Eibner, C., et al. (2020) *Reducing bias in difference-in-differences models using entropy balancing*. [Online] Available at: doi:10.48550/ARXIV.2011.04826.

Chai, W.J., Abd Hamid, A.I. and Abdullah, J.M. (2018) Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Front. Psychol.* [Online] 9. Available at: doi:10.3389/fpsyg.2018.00401.

Cheng, S.-H. and Song, H.-Y. (2004) Physician performance information and consumer choice: A survey of subjects with the freedom to choose between doctors. *BMJ Quality & Safety*. [Online] 13 (2), 98–101. Available at: doi:10.1136/qshc.2003.006981.

Clarke, D. (2017) *Estimating Difference-in-Differences in the Presence of Spillovers*

Clarke, H.D., Sanders, D., Stewart, M.C. and Whiteley, P. (2009) lccn: CUP. *Performance politics and the British voter*. [Online]. Cambridge, Cambridge University Press. Available at:

http://ezphost.dur.ac.uk/login?url=https://dx.doi.org/10.1017/CBO9780511596872.

Clarke, H.D., Sanders, D., Stewart, M. and Whiteley, P. (2004) *Political choice in britain*. Oxford University Press.

Clarke, H., Sanders, D., Stewart, M. and Whiteley, P. (2011) Valence politics and electoral choice in britain, 2010. *Journal of Elections, Public Opinion and Parties*. [Online] 21 (2), 237–253. Available at: doi:10.1080/17457289.2011.562614.

Clemence, M. (2020) *Ipsos issues index - 2019 year in review* [Online]. Available at: https://www.ipsos.com/en-uk/ipsos-issues-index-2019-year-review.

Clemence, M. (2018) *Ipsos issues index: 2017 in review* [Online]. Available at: https://www.ipsos.com/en-uk/ipsos-issues-index-2017-review.

Clemence, M. (2019) *Ipsos issues index: 2018 in review* [Online]. Available at: https://www.ipsos.com/en-uk/ipsos-issues-index-2018-review.

Clemence, M. and Skinner, G. (2023) *Ipsos issues index: September 2023* [Online]. Available at: https://www.ipsos.com/en-uk/ipsos-issues-index-september-2023.

Coleman, C. (2023) *Health care in rural areas*. [Online] Available at: https://lordslibrary.parliament.uk/health-care-in-rural-areas/.

Converse, P.E. (1962) Information flow and the stability of partisan attitudes. *Public Opinion Quarterly*. [Online] 26 (4), 578. Available at: doi:10.1086/267129.

Curtice, J., Hudson, N. and Montagu, I. (2020) [Online]. London, The National Centre for Social Research. Available at: https://www.bsa.natcen.ac.uk/media/39401/bsa37_technical-details.pdf.

Dahlstrom, C. and Lapuente, V. (2009) Explaining Cross-Country Differences in Performance-Related Pay in the Public Sector. *Journal of Public Administration Research and Theory*. [Online] 20 (3), 577–600. Available at: doi:10.1093/jopart/mup021.

Dancey, L. and Sheagley, G. (2013) Heuristics Behaving Badly: Party Cues and Voter Knowledge. *American Journal of Political Science*. 57 (2), 312–325.

Dassonneville, R. (2016) Volatile voters, short-term choices? An analysis of the vote choice determinants of stable and volatile voters in great britain. *Journal of Elections, Public Opinion and Parties*. [Online] 26 (3), 273–292. Available at: doi:10.1080/17457289.2016.1158181.

De Vries, C.E., Hobolt, S.B. and Tilley, J. (2018) Facing up to the facts: What causes economic perceptions? *Electoral Studies*. [Online] 51115–122. Available at: doi:10.1016/j.electstud.2017.09.006.

Dekavalla, M. (2010) Tax, war and waiting lists: The construction of national identity in newspaper coverage of general elections after devolution. *Discourse & Society*. [Online] 21 (6), 638–654. Available at: doi:10.1177/0957926510381222.

Department of Health (2006) *Choice matters* [Online]. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20130107105354/http://www.dh.gov.uk/assetRoot/04/13/55/42/04135542.pdf.

Department of Health (2023) *Statistical press notice: NHS referral to treatment (RTT) waiting times data july 2012* [Online]. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20121204124531/http://mediacentre.dh.gov.uk/?p=16929.

Department of Health (2012) *Waiting times and list statistics: hospitals* [Online]. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20130104155640mp_/http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Performancedataandstatistics/HospitalWaitingTimesandListStatistics/index.htm.

Digital, Culture, Media and Sport Committee (2023) *Sustainability of local journalism: Seventh report of session 2022-23* [Online]. Available at: https://publications.parliament.uk/pa/cm5803/cmselect/cmcumeds/153/report.html.

Dorey, P. (2017) Jeremy Corbyn confounds his critics: explaining the Labour party's remarkable resurgence in the 2017 election. *British Politics*. [Online] 12 (3), 308–334. Available at: doi:10.1057/s41293-017-0058-4.

Dorussen, H. and Taylor, M. (2001) The political context of issue-priority voting: Coalitions and economic voting in the netherlands, 1970–1999. *Electoral Studies*. [Online] 20 (3), 399–426. Available at: doi:10.1016/S0261-3794(00)00026-3.

Downs, A. (1957) *An economic theory of democracy*.

Durham, C. and Darlington NHS Foundation Trust (2014) *Your outpatient visit* [Online]. Available at: https://www.cddft.nhs.uk/patients-and-visitors/choosing-an-appointment-to-

suit-you-(choose-book)/your-outpatient-visit.aspx.

Economic Cooperation, O. for and Development (2019) *Country health profiles 2019* [Online]. Available at: https://www.oecd.org/health/country-health-profiles-eu.htm.

Edwards, N. (2006) *Lost in translation: why are patients more satisfied with the NHS than the public?* [Online] Available at: http://www.nhsconfed.org/resources/2008/12/lost in translation.

Evans, G. and Chzhen, K. (2016a) Beyond Valence: Estimating Models of Party Choice Without Resort to Ecological Fallacy or Unfounded Causal Assumptions. A Reply to Whiteley et al.*. *Political Science Research and Methods*. [Online] 4 (1), 241–247. Available at: doi:10.1017/psrm.2015.31.

Evans, G. and Chzhen, K. (2016b) Re-evaluating the Valence Model of Political Choice*. *Political Science Research and Methods*. [Online] 4 (1), 199–220. Available at: doi:10.1017/psrm.2013.11.

Exley, S. and Thomson, K. (2002) *Examining the reliability of SOC90 and related socio-eonomic classifications after the switch to SOC2000* [Online]. Available at: http://doc.ukdataservice.ac.uk/doc/4615/mrdoc/pdf/a4615bab.pdf.

Fieldhouse, E., Green, J., Evans, G., Schmitt, H., et al. (2019) *British election study internet panel: Waves 1-16*. [Online] Available at: http://www.britishelectionstudy.com/.

Fiorina, M.P. (1981) lccn: 324.973 FIO. *Retrospective voting in American national elections*. New Haven, Yale University Press.

Fisher, S.D. and Hobolt, S.B. (2010) Coalition government and electoral accountability. *Electoral Studies*. [Online] 29 (3), 358–369. Available at: doi:10.1016/j.electstud.2010.03.003.

Freeman, S. (2000) Deliberative Democracy: A Sympathetic Comment. *Philosophy & Public Affairs*. [Online] 29 (4), 371–418. Available at: doi:10.1111/j.1088-4963.2000.00371.x.

Freier, R., Schumann, M. and Siedler, T. (2015) The earnings returns to graduating with honors — Evidence from law graduates. *Labour Economics*. [Online] 3439–50. Available at: doi:10.1016/j.labeco.2015.03.001.

Gelman, A. and Hill, J. (2007) *Data analysis using regression and hierarchical/multilevel models*. Cambridge University Press: Cambridge, UK.

Giger, N. (2012) Is Social Policy Retrenchment Unpopular? How Welfare Reforms Affect Government Popularity. *European Sociological Review*. [Online] 28 (5), 691–700. Available at: doi:10.1093/esr/jcr039.

Gomez, B.T. and Wilson, J.M. (2006) Cognitive Heterogeneity and Economic Voting: A Comparative Analysis of Four Democratic Electorates. *American Journal of Political Science*. [Online] 50 (1), 127–145. Available at: doi:10.1111/j.1540-5907.2006.00174.x.

Gomez, B.T. and Wilson, J.M. (2001) Political sophistication and economic voting in the american electorate: A theory of heterogeneous attribution. *American Journal of Political Science*. [Online] 45 (4), 899914. Available at: doi:10.2307/2669331.

Goodman-Bacon, A. (2021) Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. [Online] 225 (2), 254–277. Available at: doi:10.1016/j.jeconom.2021.03.014.

Google (2018) Google Maps *Google Maps*.

Green, D.P., Palmquist, B. and Schickler, E. (2002) *Partisan Hearts and Minds*. New Haven, CT: Yale University Press.

Green, J. and Jennings, W. (2017a) Party Reputations and Policy Priorities: How Issue Ownership Shapes Executive and Legislative Agendas. *British Journal of Political Science*. [Online] 1–24. Available at: doi:10.1017/S0007123416000636.

Green, J. and Jennings, W. (2017b) *The Politics of Competence: Parties, Public Opinion and Voters*. Cambridge University Press.

Greifer, N. (2023a) *Cobalt: Covariate balance tables and plots*. [Online]. Available at: https://cran.r-project.org/web/packages/cobalt/index.html.

Greifer, N. (2023b) *WeightIt: Weighting for covariate balance in observational studies*. [Online]. Available at: https://cran.r-project.org/web/packages/WeightIt/index.html.

Grynaviski, J.D. (2006) A Bayesian Learning Model with Applications to Party Identification. *Journal of Theoretical Politics*. [Online] 18 (3), 323–346. Available at: doi:10.1177/0951629806064352.

Guntermann, E., Lenz, G.S. and Myers, J.R. (2021) The Impact of the Economy on Presidential Elections Throughout US History. *Political Behavior*. [Online] 43 (2), 837–857. Available at:

doi:10.1007/s11109-021-09677-y.

Hainmueller, J. (2012) Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*. [Online] 20 (1), 25–46. Available at: doi:10.1093/pan/mpr025.

Hanefeld, J., Powell-Jackson, T. and Balabanova, D. (2017) Understanding and measuring quality of care: Dealing with complexity. *Bulletin of the World Health Organization*. [Online] 95 (5), 368–374. Available at: doi:10.2471/BLT.16.179309.

Hart, W., Albarracín, D., Eagly, A.H., Brechan, I., et al. (2009) Feeling validated versus being correct:a meta-analysis of selective exposure to information. *Psychological bulletin*. [Online] 135 (4), 555–588. Available at: doi:10.1037/a0015701.

Hazlett, C. and Xu, Y. (2018) *Trajectory Balancing: A General Reweighting Approach to Causal Inference With Time-Series Cross-Sectional Data* [Online]. Available at: https://papers.ssrn.com/abstract=3214231.

Healy, A. and Lenz, G.S. (2014) Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*. 58 (1), 3147.

Healy, A. and Malhotra, N. (2013) Retrospective voting reconsidered. *Annual Review of Political Science*. [Online] 16 (1), 285306. Available at: doi:10.1146/annurev-polisci-032211-212920.

Hill, S.J. (2017) Learning together slowly: Bayesian learning about political facts. *The Journal of Politics*. 79 (4), 14031418.

Ho, K., Clarke, H.D., Chen, L.-K. and Weng, D.L.-C. (2013) Valence politics and electoral choice in a new democracy: The case of taiwan. *Electoral Studies*. [Online] 32 (3), 476481. Available at: https://www.sciencedirect.com/science/article/pii/S0261379413000607.

Hobolt, S.B. (2018) Brexit and the 2017 UK General Election. *Journal of Common Market Studies*. [Online] 56 (S1), 39–50. Available at: https://heinonline.org/HOL/P?h=hein.journals/jcmks56&i=1758.

Imai, K. and Kim, I.S. (2020) On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. *Political Analysis*. [Online] 29 (3), 405–415. Available at: doi:10.1017/pan.2020.33.

Ipsos (2023) *GP patient survey* [Online]. Available at: https://gp-patient.co.uk/.

James, O. and John, P. (2007) Public Management at the Ballot Box: Performance Information and Electoral Support for Incumbent English Local Governments. *Journal of Public Administration Research and Theory*. [Online] 17 (4), 567–580. Available at: doi:10.1093/jopart/mul020.

Jenkinson, C., Coulter, A., Bruster, S., Richards, N., et al. (2002) Patients' experiences and satisfaction with health care: Results of a questionnaire study of specific aspects of care. *Quality and safety in health care*. [Online] 11 (4), 335339. Available at: http://qualitysafety.bmj.com/content/11/4/335.short.

Johns, R., Mitchell, J., Denver, D. and Pattie, C. (2009) Valence Politics in Scotland: Towards an Explanation of the 2007 Election. *Political Studies*. [Online] 57 (1), 207–233. Available at: doi:10.1111/j.1467-9248.2008.00762.x.

Johnston, R.J. and Pattie, C.J. (2001a) 'It's the economy, stupid' - but which economy? Geographical scales, retrospective economic evaluations and voting at the 1997 british general election. *Regional Studies*. [Online] 35 (4), 309–319. Available at: doi:10.1080/00343400123348.

Johnston, R.J. and Pattie, C.J. (1997) The region is not dead: Long live the region—personal evaluations and voting at the 1992 british general election. *Space and Polity*. [Online] 1 (1), 103–113. Available at: doi:10.1080/13562579708721757.

Johnston, R. and Pattie, C. (2001b) Dimensions of Retrospective Voting. *Party Politics*. [Online] 7 (4), 469–490. Available at: doi:10.1177/1354068801007004004.

Johnston, R., Pattie, C., Dorling, D., MacAllister, I., et al. (2000) Local Context, Retrospective Economic Evaluations, and Voting: The 1997 General Election in England and Wales. *Political Behavior*. [Online] 22 (2), 121–143. Available at: doi:10.1023/A:1006655300380.

Kahan, D.M. (2013) Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*. [Online] 8 (4), 407–424. Available at: doi:10.1017/S1930297500005271.

Kahan, D.M. (2015) The politically motivated reasoning paradigm. *Emerging Trends in Social & Behavioral Sciences, Forthcoming*.

Kelso, P. (2017) Labour to end NHS restructuring over hospital closure fear. *Sky News*

[online].

Key, J.V.O. (2013) *The Responsible Electorate: Rationality in Presidential Voting, 1936–1960.* [Online]. Harvard University Press. Available at: doi:10.4159/harvard.9780674497764.

Kinder, D.R. and Kiewiet, D.R. (1981) Sociotropic politics: The american case. *British Journal of Political Science.* [Online] 11 (2), 129–161. Available at: http://www.jstor.org/stable/193580.

King, G. (2010) A hard unsolved problem? Post-treatment bias in big social science questions. In: *Hard Problems in Social Science" Symposium, April.* 2010

Kirk-Wade, E., Mirza-Davies, J., Garratt, K. and Harker, R. (2023) *The NHS workforce in england.* [Online] Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-9731/.

Klein, R. (2007) The new model NHS: performance, perceptions and expectations. *British Medical Bulletin.* [Online] 81-82 (1), 39–50. Available at: doi:10.1093/bmb/ldm013.

Knobloch-Westerwick, S. and Meng, J. (2009) Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research.* [Online] 36 (3), 426–448. Available at: doi:10.1177/0093650209333030.

Ladd, J.M. and Lenz, G.S. (2009) Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media. *American Journal of Political Science.* [Online] 53 (2), 394–410. Available at: doi:10.1111/j.1540-5907.2009.00377.x.

Lagarde, M., Erens, B. and Mays, N. (2015) Determinants of the choice of GP practice registration in England: Evidence from a discrete choice experiment. *Health Policy.* [Online] 119 (4), 427–436. Available at: doi:10.1016/j.healthpol.2014.10.008.

Lau, R.R. and Redlawsk, D.P. (2006) *How voters decide: Information processing during election campaigns.* Cambridge studies in public opinion and political psychology. Cambridge, Cambridge University Press.

Laycock, S. (2009) Experience versus perceptions: Accounting for the NHS "perception gap". *Journal of Elections, Public Opinion and Parties.* [Online] 19 (4), 449472. Available at: http://www.tandfonline.com/doi/abs/10.1080/17457280903275295.

Lazarsfeld, P.F., Berelson, B. and Gaudet, H. (1968) *The people's choice.* [Online]. Columbia

University Press. Available at: doi:10.7312/laza93930.

Lewis-Beck, M.S. and Nadeau, R. (2011) Economic voting theory: Testing new dimensions. *Electoral Studies*. [Online] 30 (2), 288294. Available at: doi:10.1016/j.electstud.2010.09.001.

Lewis-Beck, M.S. and Paldam, M. (2000) Economic voting: An introduction. *Electoral studies*. [Online] 19 (2-3), 113121. Available at: https://www.academia.edu/download/6179351/s4._economic_voting_an_introduction__lewis-beck_paldam_2000_.pdf.

Lewis-Beck, M.S. and Stegmaier, M. (2007) *Economic models of voting*. [Online] Available at: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199270125.001.0001/oxfordhb-9780199270125-e-027.

Lewis-Beck, M.S. and Stegmaier, M. (2019) *Economic voting*. In: Roger D. Congleton, Bernard Grofman, and Stefan Voigt (eds.). [Online]. Oxford University Press. p. 0. Available at: doi:10.1093/oxfordhb/9780190469733.013.12.

Lewis-Beck, M.S. and Stegmaier, M. (2008) The Economic Vote in Transitional Democracies. *Journal of Elections, Public Opinion and Parties*. [Online] 18 (3), 303–323. Available at: doi:10.1080/17457280802227710.

Lilburn, C. (2023) *Public support for government's handling of the NHS reaches new low, polling finds* [Online]. Available at: https://www.health.org.uk/news-and-comment/news/public-support-for-government-s-handling-of-the-nhs-reaches-new-low.

Lindbom, A. (2014) Waking up the giant? Hospital closures and electoral punishment in sweden. *How Welfare States Shape the Democratic Public: policy feedback, participation, voting, and attitudes*. 156178.

Little, A.T. (2022) *Detecting motivated reasoning*. [Online] Available at: doi:10.31219/osf.io/b8tvk.

Lobo, M.C. (2013) Dimensions of the economic vote: Valence, positional and patrimony voting in portugal's 2011 elections. *Electoral Studies*. [Online] 32 (3), 460464. Available at: https://www.sciencedirect.com/science/article/pii/S0261379413000693.

Makridis, C.A. (2019) Do Right-to-Work Laws Work? Evidence on Individuals' Well-Being and Economic Sentiment. *The Journal of Law and Economics*. [Online] 62 (4), 713–745. Available at: doi:10.1086/707081.

Manin, B., Przeworski, A. and Stokes, S. (1999) *Elections and representation.* In: Cambridge University Press. pp. 29–54.

Mannion, R. and Braithwaite, J. (2012) Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Internal Medicine Journal.* [Online] 42 (5), 569–574. Available at: doi:10.1111/j.1445-5994.2012.02766.x.

Marsh, M. and Tilley, J. (2010) The attribution of credit and blame to governments and its impact on vote choice. *British Journal of Political Science.* [Online] 40 (01), 115134. Available at: http://journals.cambridge.org/abstract_S0007123409990275.

Maybin, J. (2010) *A high-performing NHS?* [Online]. Available at: http://www.kingsfund.org.uk/publications/high-performing-nhs.

McIntyre, D. and Chow, C.K. (2020) Waiting time as an indicator for health services under strain: A narrative review. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*. 570046958020910305.

Mellon, J., Evans, G., Fieldhouse, E., Green, J., et al. (2018) Brexit or corbyn? Campaign and inter-election vote switching in the 2017 UK general election. *Parliamentary Affairs.* [Online] 71 (4), 719–737. Available at: doi:10.1093/pa/gsy001.

Milmo, C. (2017) The full list of the 24 A&E units marked for closure *iNews.* [Online]. Available at: https://inews.co.uk/nhs/full-list-24-aes-marked-closure/.

Milmo, C., Day, A., Mort, D., Burn, C., et al. (2017) The full list of the 19 hospitals facing closure *iNews.* [Online]. Available at: https://inews.co.uk/news/health/great-nhs-gamble-full-list-19-hospitals-facing-closure/.

Morris, J., Shlepper, L., Dayan, M., Jefferies, D., et al. (2023) *Public satisfaction with the NHS and social care in 2022* [Online]. Available at: https://www.kingsfund.org.uk/publications/public-satisfaction-nhs-and-social-care-2022.

Nadeau, R., Niemi, R.G. and Yoshinaka, A. (2002) A cross-national analysis of economic voting: Taking account of the political context across time and nations. *Electoral Studies.* [Online] 21 (3), 403423. Available at: http://www.sciencedirect.com/science/article/pii/S0261379401000026.

Nannestad, P. and Paldam, M. (1997) The grievance asymmetry revisited: A micro study of economic voting in Denmark,1986–1992. *European Journal of Political Economy*. [Online] 13 (1), 81–99. Available at: doi:10.1016/s0176-2680(96)00034-1.

National Audit Office (2001) *Inpatient and outpatient waiting in the NHS* [Online]. Available at: https://www.nao.org.uk/reports/inpatient-and-outpatient-waiting-in-the-nhs/.

NHS Digital (2020) *Hospital Outpatient Activity 2019-20* [Online]. Available at: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-outpatient-activity/2019-20.

NHS England (2023) *Referral to treatment (RTT) waiting times* [Online]. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/.

NHS.uk (2020) *Outpatients and day patients* [Online]. Available at: https://www.nhs.uk/nhs-services/hospitals/going-into-hospital/outpatients-and-day-patients/.

Nir, L. (2011) Motivated reasoning and public opinion perception. *Public Opinion Quarterly*. 75 (3), 504532.

Nyholt, N. (2023) Left behind: Voters' reactions to local school and hospital closures. *European Journal of Political Research*. [Online] n/a (n/a). Available at: doi:10.1111/1475-6765.12622.

Office for National Statistics (2022) *Census 2021 geographies* [Online]. Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/census2021geographies#middle-layer-super-output-areas-msoas.

Office for National Statistics (2021) *Clinical commissioning group population estimates (national statistics)* [Online]. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/clinicalcommissioninggroupmidyearpopulationestimates.

Office for National Statistics (2023a) *Clinical Commissioning Groups Boundaries in England (April 2016, April 2017, April 2018, April 2019)* [Online]. Available at: https://geoportal.statistics.gov.uk/search?collection=Dataset&sort=name&tags=all(BDY_CCG).

Office for National Statistics (2023b) *MSOA (dec 2011) boundaries super generalised clipped (BSC) EW V3* [Online]. Available at: https://geoportal.statistics.gov.uk/datasets/ons::msoa-dec-2011-boundaries-super-generalised-clipped-bsc-ew-v3/explore.

Organisation for National Statistics (2020) *Population estimates for the UK, england and wales, scotland and northern ireland* [Online]. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates.

Pattie, C. and Johnston, R. (1998) Voter turnout at the British General Election of 1992: Rational choice, social standing or political efficacy? *European Journal of Political Research.* [Online] 33 (2), 263–283. Available at: doi:10.1023/A:1006854900327.

Pebesma, E., Bivand, R., Racine, E., Sumner, M., et al. (2023) *Sf: Simple features for r.* [Online]. Available at: https://cran.r-project.org/web/packages/sf/index.html.

Pierson, P. (1996) The new politics of the welfare state. *World Politics.* [Online] 48 (2), 143–179. Available at: http://www.jstor.org/stable/25053959.

Pitchforth, E., Nolte, E., Corbett, J., Miani, C., et al. (2017) *Community hospitals and their services in the NHS: Identifying transferable learning from international developments – scoping review, systematic review, country reports and case studies.* Health Services and Delivery Research. Southampton (UK), NIHR Journals Library.

Prendergast, T. (2023) *Healthcare expenditure, UK health accounts* [Online]. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2021#government-healthcare-expenditure.

R Core Team (n.d.) *R: A language and environment for statistical computing.* [Online]. Vienna, Austria, R Foundation for Statistical Computing. Available at: https://www.R-project.org.

Rahn, W.M. (1993) The Role of Partisan Stereotypes in Information Processing about Political Candidates. *American Journal of Political Science.* [Online] 37 (2), 472–496. Available at: doi:10.2307/2111381.

Rau, E.G. (2022) Partisanship as Cause, Not Consequence, of Participation. *Comparative Political Studies.* [Online] 55 (6), 1021–1058. Available at: doi:10.1177/00104140211047406.

Redlawsk, D.P., Civettini, A.J.W. and Emmerson, K.M. (2010) The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"? *Political Psychology.* [Online] 31 (4), 563–593. Available at: doi:10.1111/j.1467-9221.2010.00772.x.

Rudolph, L. and Kuhn, P.M. (2018) Natural Disasters and Political Participation: Evidence

from the 2002 and 2013 Floods in Germany. *German Politics*. [Online] 27 (1), 1–24. Available at: doi:10.1080/09644008.2017.1287900.

Sanders, D. (1999) Conservative incompetence, labour responsibility and the feelgood factor: Why the economy failed to save the conservatives in 1997. *Electoral Studies*. [Online] 18 (2), 251–270. Available at: doi:10.1016/S0261-3794(98)00053-5.

Schaffner, B.F. and Roche, C. (2017) Misinformation and motivated reasoning: Responses to economic news in a politicized environment. *Public Opinion Quarterly*. [Online] 81 (1), 86–110. Available at: doi:10.1093/poq/nfw043.

Schaffner, B.F. and Streb, M.J. (2002) The partisan heuristic in low-information elections. *The Public Opinion Quarterly*. [Online] 66 (4), 559–581. Available at: https://www.jstor.org/stable/3078708.

Stevens, D. (2019) *Satisficing in political decision making*. [Online] Available at: doi:10.1093/acrefore/9780190228637.013.1020.

Stiers, D. (2022) Performance voting, retrospective voting, and economic voting. Conceptual clarity and empirical testing. *Social Science Quarterly*. [Online] Available at: doi:10.1111/ssqu.13125.

Stokes, D.E. (1963) Spatial models of party competition. *The American Political Science Review*. [Online] 57 (2), 368–377. Available at: doi:10.2307/1952828.

Strickland, A.A., Taber, C.S. and Lodge, M. (2011) Motivated reasoning and public opinion. *Journal of health politics, policy and law*. 36 (6), 935944.

Taber, C.S., Cann, D. and Kucsova, S. (2009) The motivated processing of political arguments. *Political Behavior*. 31 (2), 137155.

Taber, C.S. and Lodge, M. (2006) Motivated skepticism in the evaluation of political beliefs. *American journal of political science*. 50 (3), 755769.

Taber, C.S., Lodge, M. and Glathar, J. (2001) *The motivated construction of political judgments*. In: [Online]. Cambridge University Press. pp. 198–226. Available at: doi:10.1017/cbo9780511896941.010.

Tan, S., Erens, B., Wright, M. and Mays, N. (2015) Patients' experiences of the choice of GP practice pilot, 2012/2013: A mixed methods evaluation. *BMJ Open*. [Online] 5 (2), e006090.

Available at: doi:10.1136/bmjopen-2014-006090.

The King's Fund (2017) *Sustainability and transformation plans (STPs) explained* [Online]. Available at: https://www.kingsfund.org.uk/topics/integrated-care/sustainability-transformation-plans-explained.

Tilley, J., Neundorf, A. and Hobolt, S.B. (2018) When the Pound in People's Pocket Matters: How Changes to Personal Financial Circumstances Affect Party Choice. *The Journal of Politics*. [Online] 80 (2), 555–569. Available at: doi:10.1086/694549.

UK Parliament (2018) MPs *UK Parliament*.

Wagner, M., Tarlov, J. and Vivyan, N. (2014) Partisan Bias in Opinion Formation on Episodes of Political Controversy: Evidence from Great Britain. *Political Studies*. [Online] 62 (1), 136–158. Available at: doi:10.1111/j.1467-9248.2012.01002.x.

Wänström, J., Karlsson, M. and Wänström, L. (2012) Citation Key: wanstromRiskfylldaBeslutSkolnedlaggelsers2012 tex.ids= wänström2012. *Riskfyllda beslut? : Skolnedläggelsers avtryck i kommunala valresultat*.

Welch, S. and Thomas, S. (1988) Explaining the gender gap in british public opinion. *Women & Politics*. [Online] 8 (3-4), 2544. Available at: http://www.tandfonline.com/doi/abs/10.1300/J014v08n03_02.

Wellings, D., Appleby, J., Maguire, D., Hemmings, N., et al. (2020) *Public satisfaction with the NHS and social care in 2019* [Online]. Available at: https://www.kingsfund.org.uk/publications/public-satisfaction-nhs-social-care-2019.

Whiteley, P., Clarke, H., Sanders, D. and Stuart, M. (2016) Hunting the snark: A reply to "re-evaluating valence models of political choice"*. *Political Science Research and Methods*. [Online] 4 (1), 221240. Available at: doi:10.1017/psrm.2015.30.

Whiteley, P., Stewart, M.C., Sanders, D. and Clarke, H.D. (2005) The issue agenda and voting in 2005. *Parliamentary Affairs*. 58 (4), 802817.

Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis*. [Online]. Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.

Wilkin, S., Haller, B. and Norpoth, H. (1997) From argentina to zambia: A world-wide test of economic voting. *Electoral Studies*. [Online] 16 (3), 301–316. Available at:

doi:10.1016/S0261-3794(97)00028-0.

Williams, B. (1994) Patient satisfaction: A valid concept? *Social science & medicine*. [Online] 38 (4), 509516. Available at: http://www.sciencedirect.com/science/article/pii/027795369490247X.

Williams, B., Coyle, J. and Healy, D. (1998) The meaning of patient satisfaction: An explanation of high reported levels. *Social science & medicine*. [Online] 47 (9), 13511359. Available at: http://www.sciencedirect.com/science/article/pii/S0277953698002135.

Wolf, J., Niederhauser, V., Marshburn, D. and LaVela, S. (2014) Defining patient experience. *Patient Experience Journal*. [Online] 1 (1), 7–19. Available at: https://pxjournal.org/journal/vol1/iss1/3.

YouGov plc (2020) *YouGov Tracker Most Important Issues*

Zhao, Q. and Percival, D. (2016) Entropy Balancing is Doubly Robust. *Journal of Causal Inference*. [Online] 5 (1). Available at: doi:10.1515/jci-2016-0010.

Ziebland, S., Evans, J. and Toynbee, P. (2011) Exceptionally good? Positive experiences of NHS care and treatment surprises lymphoma patients: A qualitative interview study. *Health Expectations : An International Journal of Public Participation in Health Care and Health Policy*. [Online] 14 (1), 21–28. Available at: doi:10.1111/j.1369-7625.2010.00609.x.

Zur, R. (2017) When Valence Crushes: Explaining the Electoral Failure of the German FDP in the 2013 Election. *German Politics*. [Online] 26 (3), 380–397. Available at: doi:10.1080/09644008.2016.1257612.

10 Appendices

10.1 Appendix A1: Additional performance statistics and notes

10.1.1 Summary statistics

This section contains the full summary statistics for the British Election Study Internet Panel survey data used in the geographical variation models. Table 10.1 shows the categorical variables, while numeric variables are shown in Table 10.2. The outcome varible "changenhs" is shown in both formats here, but treated as numeric in the models.

Table 10.3 shows the summary statistics for the British Social Attitudes survey data used in the over-time models.

10.1.2 Note on the 2008 waiting time data

The 2008 median waiting time data in the modern metric appears suspiciously large, potentially as a result of changes in definitions or reporting criteria. I therefore try re-running the previous modern-only waiting times model but excluding 2008 data and find that the effect size would be much higher without 2008; the results are shown alongside the original results in Table 10.4.

Without 2008 data, the modern data gives a coefficient of 13% of a step on the Likert scale, and the difference between best and worst years reduces to 1.8 weeks; this gives a difference of 23% of a Likert scale step between the best and worst years. This adds weight to my suspicions

Table 10.1: BESIP summary statistics (categorical variables)

		N	%
partyid	Brexit Party	3629	2.2
	Conservative	56685	33.9
	Green Party	5562	3.3
	Labour	49633	29.7
	Liberal Democrats	14296	8.6
	None	35574	21.3
	Other	1625	1.0
	Plaid Cymru	1	0.0
region	East Midlands	15111	9.0
	Eastern	19039	11.4
	London	21502	12.9
	North East	8698	5.2
	North West	21179	12.7
	South East	28614	17.1
	South West	19308	11.6
	West Midlands	15858	9.5
	Yorkshire and Humber	17696	10.6
ethnicity	Other ethnic background	9973	6.0
	Other White background	6321	3.8
	White British	150711	90.2
changenhs	1	62819	37.6
	2	60281	36.1
	3	36233	21.7
	4	6738	4.0
	5	934	0.6
wave	10	20600	12.3
	11	21142	12.7
	12	24013	14.4
	14	21124	12.6
	15	21188	12.7
	16	27283	16.3
	17	24751	14.8
	18	6904	4.1
partyid2	Conservatives	56685	33.9
	Labour	49633	29.7
	Liberal Democrats	14296	8.6
	None or Don't Know	35574	21.3
	Other party	10817	6.5
month	Apr	18369	11.0
	Dec	15092	9.0
	Jun	18586	11.1
	Mar	21188	12.7
	May	56607	33.9
	Nov	37163	22.3
agecat	18-37	35042	21.0
	38-49	32960	19.7
	50-60	34405	20.6
	61-68	32871	19.7
	69-85	31727	19.0

Table 10.2: BESIP summary statistics (categorical variables)

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
age	68	0	52.7	16.2	18.0	55.0	85.0	
changenhs	5	0	1.9	0.9	1.0	2.0	5.0	

about the 2008 data, but to avoid cherrypicking, I use the 2008 data in the main analysis despite this concern.

10.2 Appendix A2: Additional analyses of the effect of recent NHS contact

10.2.1 Are non-partisans less engaged with the news?

In the BSA data, 40% of partisans report reading a morning newspaper 3 or more times per week, compared to 27% of non-partisans. The difference is partly explained by other demographic differences between partisan groups such as age, but non-partisans are still around 7.7% less likely to read a morning newspaper 3+ times a week than partisans after controlling for age, marital status, socio-economic group, region, ethnic minority status and year.

10.2.2 Full coefficients for overall satisfaction models

Figure 10.1 shows all coefficients and 95% confidence intervals for the overall satisfaction models (M1 and M1a).

Table 10.3: BSA summary statistics

		N	%
female	0	29157	44.0
	1	37088	56.0
agecat	17-32	13800	20.8
	33-42	12698	19.2
	43-54	13401	20.2
	55-67	12912	19.5
	68+	13434	20.3
marstat	Not married	12850	19.4
	Widowed	7541	11.4
	Married	32174	48.6
	Living as married	5335	8.1
	Separated or divorced	8297	12.5
seg	Professional/employers/managers	12503	18.9
	Intermediate non-manual	13151	19.9
	Junior non-manual	11578	17.5
	Supervisor/skilled manual	12039	18.2
	Semi-skilled manual	11158	16.8
	Other	328	0.5
	Unskilled Manual	3258	4.9
region	Greater London	8677	13.1
	North East	4272	6.4
	North West	9048	13.7
	West Midlands	6846	10.3
	Yorkshire and the Humber	7102	10.7
	East Midlands	6255	9.4
	East of England	5929	9.0
	South West	6981	10.5
	South East	11135	16.8
partyid_new	Conservative	20487	30.9
	Labour	23457	35.4
	Lib Dem/Alliance	6873	10.4
	None	8692	13.1
	Other	6736	10.2
eth_minority	0	59679	90.1
	1	5048	7.6

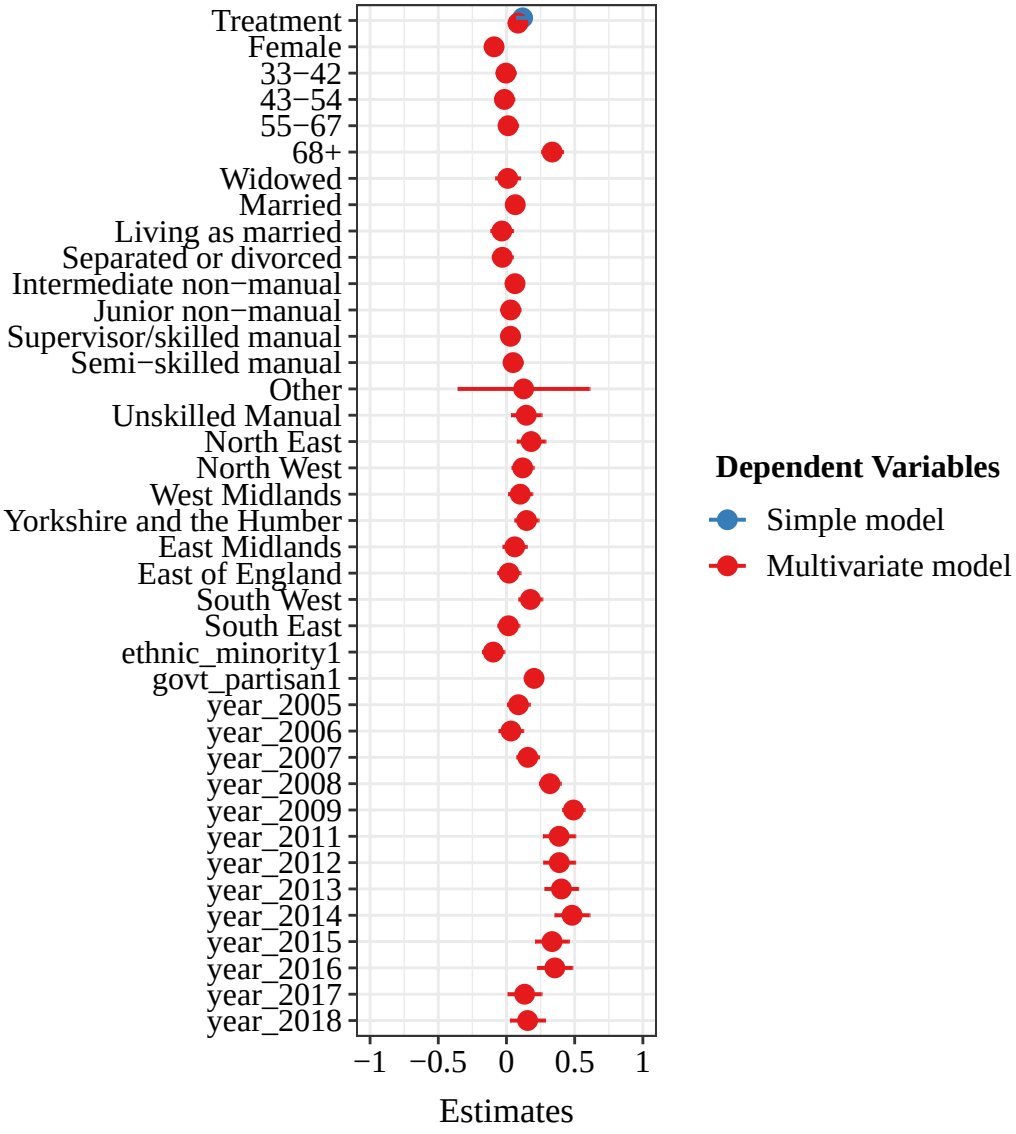

Figure 10.1: Overall satisfaction model (M1a) coefficients

Table 10.4: Comparison of modern waiting times with and without 2008 data

	Waiting times (full)	Waiting times (full, exc. 2008)
(Intercept)	3.798***	4.303***
	[3.693, 3.904]	[4.126, 4.481]
Median waiting time (weeks)	−0.053***	−0.130***
	[−0.066, −0.041]	[−0.155, −0.104]
Num.Obs.	24 224	21 427

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

	(1)
nonpartisan	−0.077***
	(0.006)
Num.Obs.	45 477

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

10.2.3 Hypothesis tests

The below graphs show the results of hypothesis tests that check whether differences in group-average treatment effect between different partisan groups are statistically significant.

10.2.4 Understanding Society partisanship model

To corroborate the findings of this model, I estimate a similar multivariate model using the Understanding Society data. I estimate the effect of NHS inpatient experience on evaluations of local medical facilities, mediated by partisanship, using a multivariate OLS model (M5), with interacted treatment terms for NHS inpatient status and partisanship, demographic controls, and year dummies. Regression output is shown in Table 10.5, and conditional contrasts (effects of treatment within individual partisan groups) are shown in Figure 10.3.

As was the case with the British Social Attitudes data, the results using Understanding Society data show that within both pro-government and pro-opposition partisan groups, those respondents with recent NHS inpatient experience gave higher evaluations of the NHS than

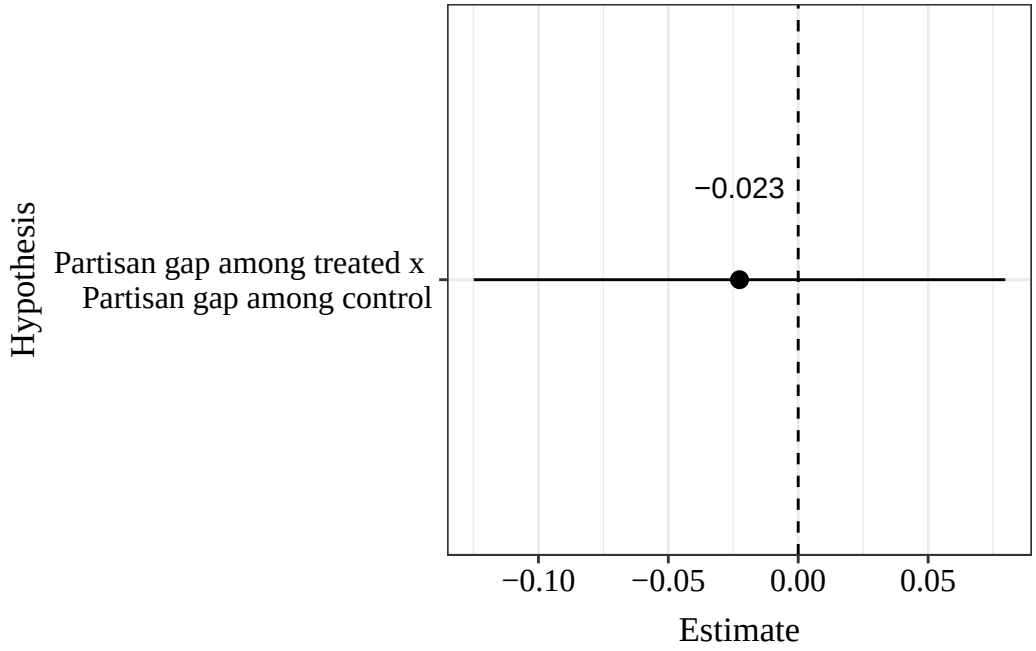

Figure 10.2: Hypothesis tests showing that the average treatment effect among Government partisans is different from that of opposition partisans

Table 10.5: Results of OLS regression on Understanding Society data (M5) showing the effect of recent inpatient stay by party identity group

	(1)
(Intercept)	2.522*** [95pc CI: 2.482, 2.561]
Recent inpatient	0.066* [95pc CI: 0.009, 0.122]
Opposition partisan	0.014 [95pc CI: −0.011, 0.040]
Non-partisan	−0.135 [95pc CI: −0.424, 0.153]
Female	0.055*** [95pc CI: 0.031, 0.080]
Age: 28-38	0.094*** [95pc CI: 0.051, 0.138]
Age: 39-49	0.120*** [95pc CI: 0.077, 0.164]
Age: 50-63	0.186*** [95pc CI: 0.145, 0.228]
Age: 64+	0.283*** [95pc CI: 0.242, 0.324]
Year: 2003	−0.054*** [95pc CI: −0.083, −0.025]
Year: 2008	0.097*** [95pc CI: 0.067, 0.126]
Recent inpatient x Opposition partisan	0.042 [95pc CI: −0.038, 0.121]
Recent inpatient x Non-partisan	0.226 [95pc CI: −0.510, 0.962]
Num.Obs.	12 665

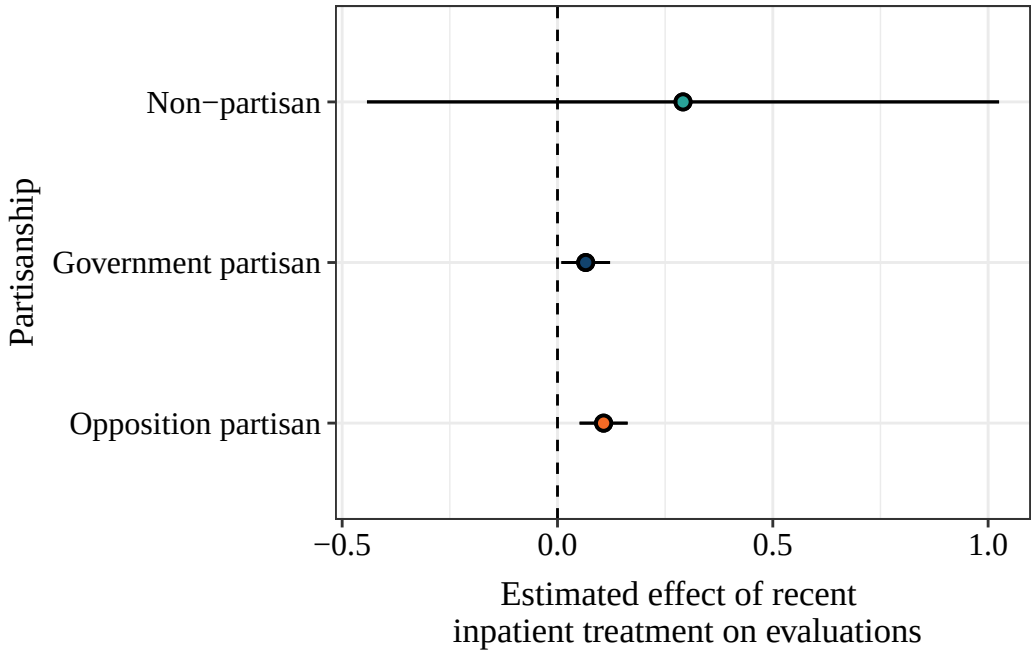

Figure 10.3: Conditional contrasts from OLS regression on Understanding Society data (M5) showing the effect of recent inpatient stay by party identity group

those without. The effect of NHS contact was again not statistically significant among non-partisans; however, this is far from surprising in the Understanding Society data, where only 27 respondents self-identified with no party, of which 4 were recent hospital inpatients. As a result, the 95% confidence intervals are very wide (almost -0.5 to +1) for non-partisans. The estimated conditional contrast between a typical government partisan in the "recent inpatient" condition compared to the control group was 7% of a step on the four-point Likert scale (95% CI: 1% - 12%) and for opposition partisans was 10% of a step (95% CI: 5% - 16%).

10.2.5 Logistic Regression

10.2.5.1 British Social Attitudes data

I first estimate a simple logistic regression model that tests how recent NHS experience (reporting having been an inpatient in the previous 12 months) affects satisfaction with the NHS. Table Table 10.6 shows the estimated coefficients and average marginal effects.

Table 10.6: Logistic regression results (effect of recent NHS contact)

	Logistic Regression (M1)	Average Marginal Effects
(Intercept)	0.172***	
	[0.132, 0.212]	
Recent Inpatient	0.262***	0.064***
	[0.182, 0.343]	[0.045, 0.083]
Num.Obs.	12 853	12 853

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

The results indicate that recent inpatients have a 6.4% higher probability of being satisfied with the NHS. The implied probabilities and 95% confidence intervals are shown on Figure 10.4.

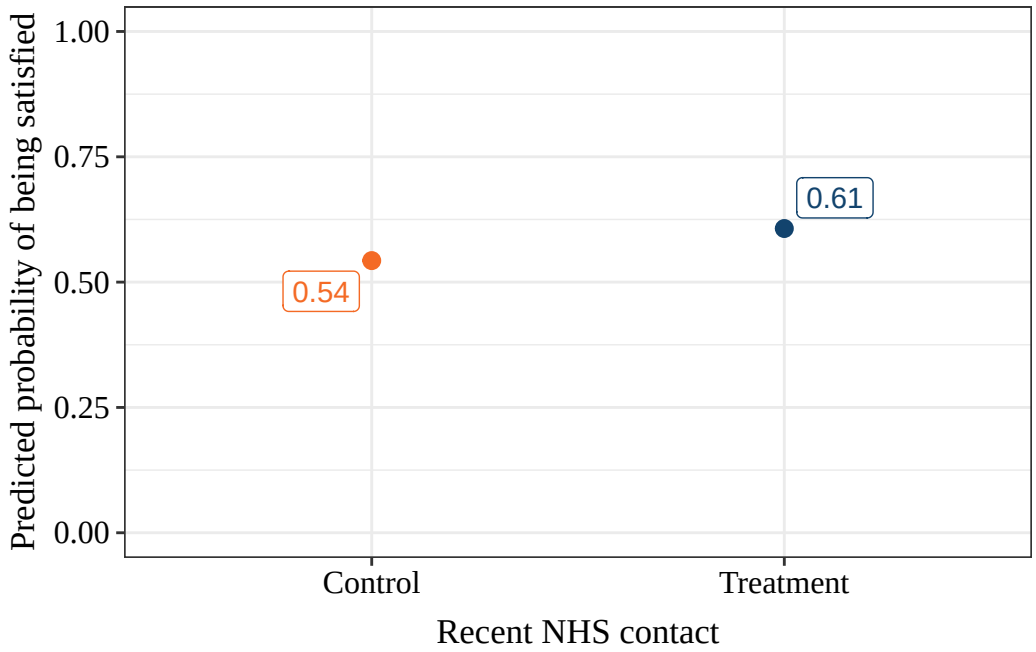

Figure 10.4: Predicted probability of being satisfied, from logistic model

Because both the outcome variable (satisfaction with the NHS) and the treatment (recent inpatient experience) could be influenced by other variables (e.g. age), I now estimate a multivariate logistic model including the demographic controls. Results are shown in Table 10.7.

The average marginal effect of being a recent inpatient (i.e. the increased probability of being

Table 10.7: Multivariate logistic regression results (effect of recent NHS contact)

	Multivariate logistic Regression (M1b)	Average Marginal Effects
(Intercept)	−0.533***	
	[−0.795, −0.272]	
Recent inpatient	0.221***	0.052***
	[0.136, 0.306]	[0.032, 0.072]
Num.Obs.	12 474	12 474
R2	0.033	0.033
R2 Adj.	0.028	0.028
AIC	16 645.0	16 645.0
BIC	16 942.2	16 942.2
RMSE	0.49	0.49
Std.Errors	IID	

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

satisfied with the NHS) falls slightly to 5.19% when demographic and political factors have been accounted for. Figure 10.5 shows the predicted probability of being satisfied with the NHS for those with recent inpatient experience and those without, after controlling for demographic differences.

10.2.5.2 Panel data

Since the Understanding Society outcome variable has only four categories, and therefore stretches the assumptions of OLS regression, I also carry out a multinomial logistic regression to check whether the estimated treatment effect is an artefact of forcing a linear structure on the 'satisfaction' variable. Political party, year, sex and age category are used as control variables.

As coefficients from multinomial logistic regression models are difficult to interpret, these were transformed into average predicted probabilities.

Figure 10.6 shows the average predicted effect of treatment on the probability of giving each response. For example, an estimate of 1% in the "Excellent" category would indicate that the average predicted probability of saying that local medical services were excellent was 1% higher

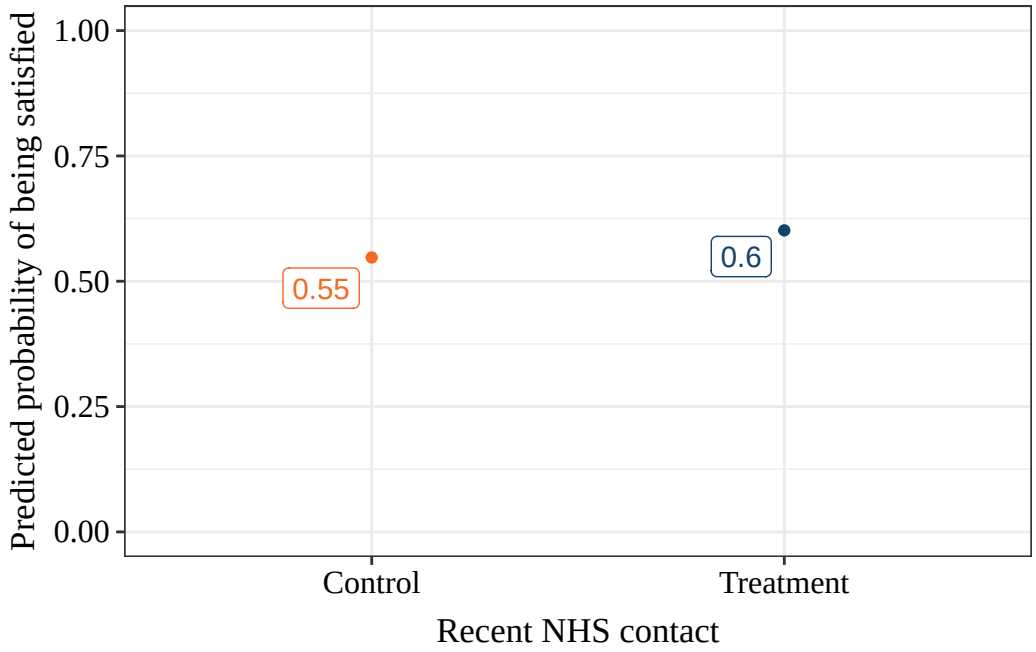

Figure 10.5: Predicted probability of being satisfied, from multivariate logistic model

Table 10.8: Average marginal effect of contact on probability of each level of satisfaction

Response	Estimated effect [95% CI]
Excellent	3.8 [3.4 - 4.3]
Fair	-3.8 [-6 - -1.6]
Poor	0.3 [0.2 - 0.5]
Very Good	-0.4 [-3.1 - 2.4]

among the recently treated group, compared to the control group. The underlying output is also shown in Table 10.8

As the table and graph show, having a recent hospital admission was associated with a higher probability of rating local medical services as "Excellent" (3.7% higher) or as "Poor" (0.4% higher). Recently treated respondents were less likely to rate local medical services as "Fair" (3.9% less likely). The difference for "Very Good" responses was not statistically significant at the p = 0.05 level.

On net, this means that recently treated respondents were more likely to give better ratings to local medical services, as shown in the main analysis. Interestingly, responses were more

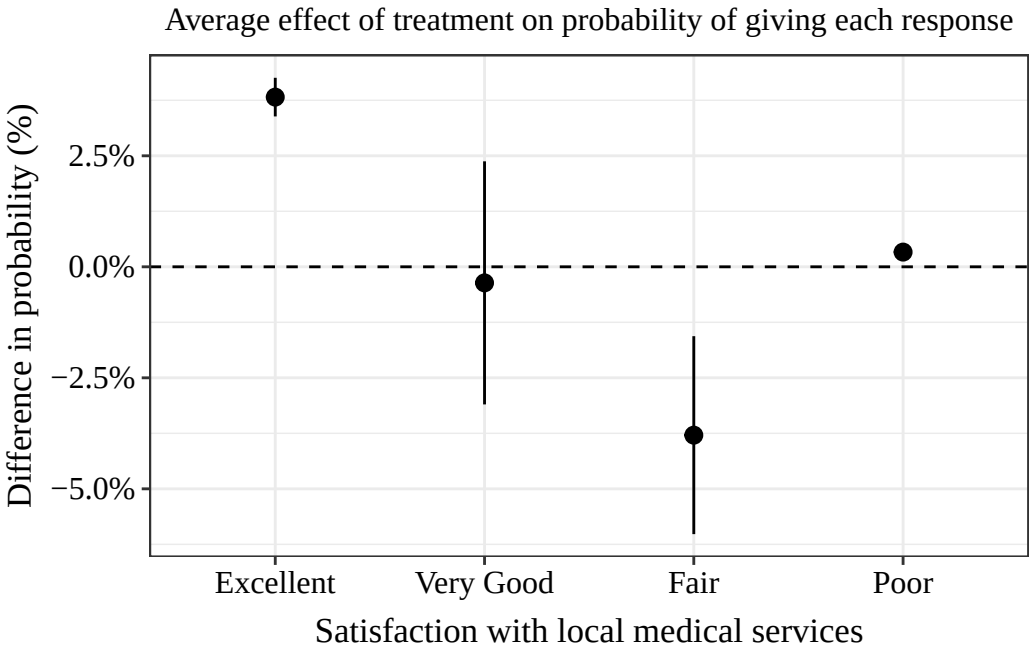

Figure 10.6: Effect of recent contact on satisfaction with local medical services

polarised in the recently treated group, with a higher average probability of giving either the best or worst rating; this is consistent with people updating their priors, but with a small number of people updating them downwards due to negative experiences.

10.2.5.3 Outpatient Analysis

As mentioned in the Research Design section, I now show key results from the models when re-analysed using recent outpatient status as the predictor.

Figure 10.7 shows sample size by year for the control and treatment groups. The control group is respondents who said that they had not been an NHS outpatient in the previous year, the treatment group is those who said that they had. People who did not express their level of satisfaction with the NHS are excluded.

I now estimate a simple bivariate regression of the effect of recent outpatient status on NHS satisfaction using the BSA data, as well as a multivariate regression that includes controls

ɔ answered both NHS satisfaction and inpatient status questions (exc. Don't Know)

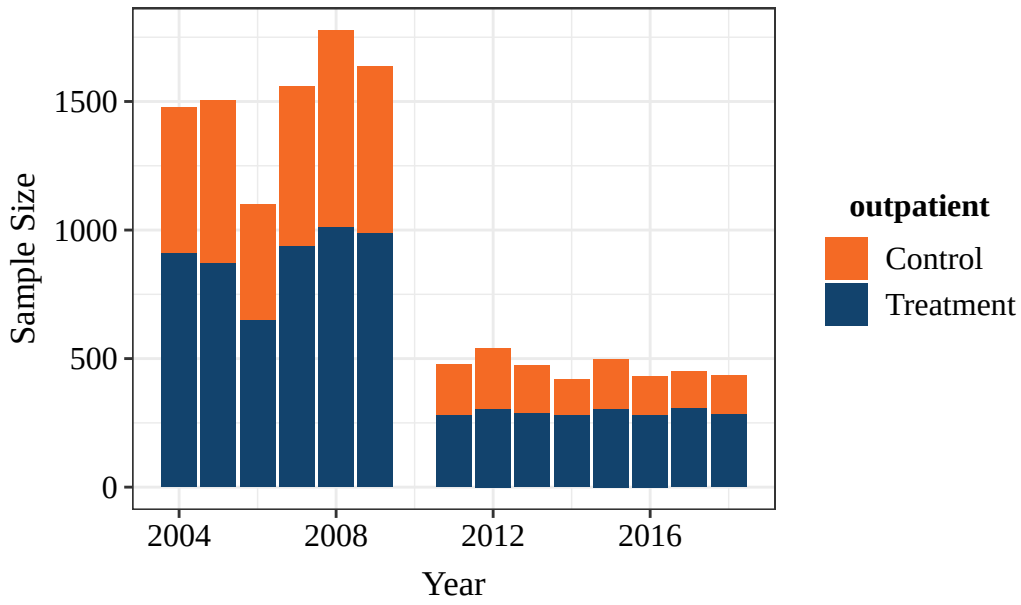

Figure 10.7: Sample sizes by group and year (BSA data)

for sex, age category, marital status, socio-economic group, region, ethnic minority status, government partisanship and year. Coefficients for the main effects are shown in Table 10.9 and Table 10.10, alongside the inpatient equivalents from the Results section.

I also re-estimate the 'partisanship' model using outpatient status as the predictor. Figure 10.8 compares the predicted satisfaction levels by partisanship in the outpatient and inpatient models. Because of differing response patterns, the modal respondent is not identical in each model (in the outpatient model the are aged 68+, rather than 55-67) - to make the baseline satisfaction equal, the control variables are set at the typical levels from the inpatient model.

As Figure 10.8 shows, differences between the two specifications are slight, indicating that information gained from recent outpatient experience has a similar impact across partisan groups to that gained from recent inpatient experience.

Table 10.9: Bivariate models (effect of outpatient contact)

	Bivariate (inpatient)	Bivariate (outpatient)
(Intercept)	3.312***	3.288***
	[3.288, 3.335]	[3.255, 3.320]
Recent inpatient	0.118***	
	[0.072, 0.165]	
Recent outpatient		0.112***
		[0.070, 0.154]
Num.Obs.	12 853	12 785
R2	0.002	0.002
R2 Adj.	0.002	0.002
AIC	40 589.8	40 546.2
BIC	40 612.1	40 568.5
Log.Lik.	−20 291.879	−20 270.076
RMSE	1.17	1.18

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10.10: Multivariate models (effect of outpatient contact)

	Multivariate (inpatient)	Multivariate (outpatient)
(Intercept)	2.857***	2.827***
	[2.745, 2.969]	[2.713, 2.941]
Recent inpatient	0.084***	
	[0.038, 0.131]	
Recent outpatient		0.077***
		[0.034, 0.119]
Num.Obs.	12 474	12 391
R2	0.051	0.053
R2 Adj.	0.048	0.050
AIC	38 838.7	38 737.3
BIC	39 143.3	39 041.8
Log.Lik.	−19 378.328	−19 327.672
RMSE	1.14	1.15

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

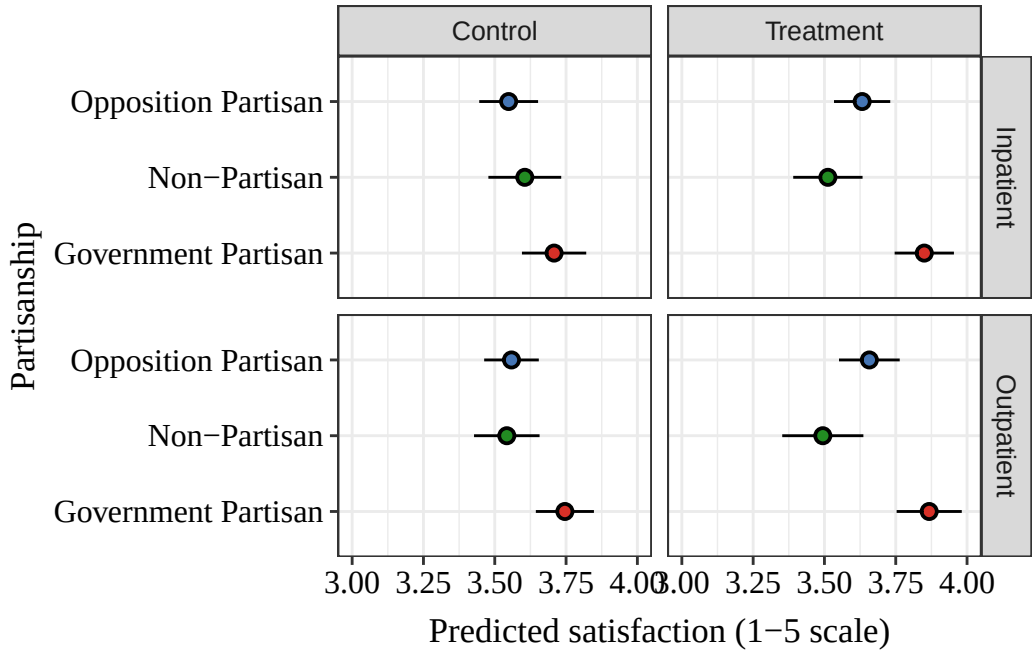

Figure 10.8: Predicted NHS satisfaction by partisanship and recent NHS outpatient status

10.3 Appendix A3: Additional retrenchment analyses

The figures in this section are Love plots showing the covariate balance for each of the four outcome variables. They demonstrate the effectiveness of the entropy balancing procedure in establishing covariate balance.

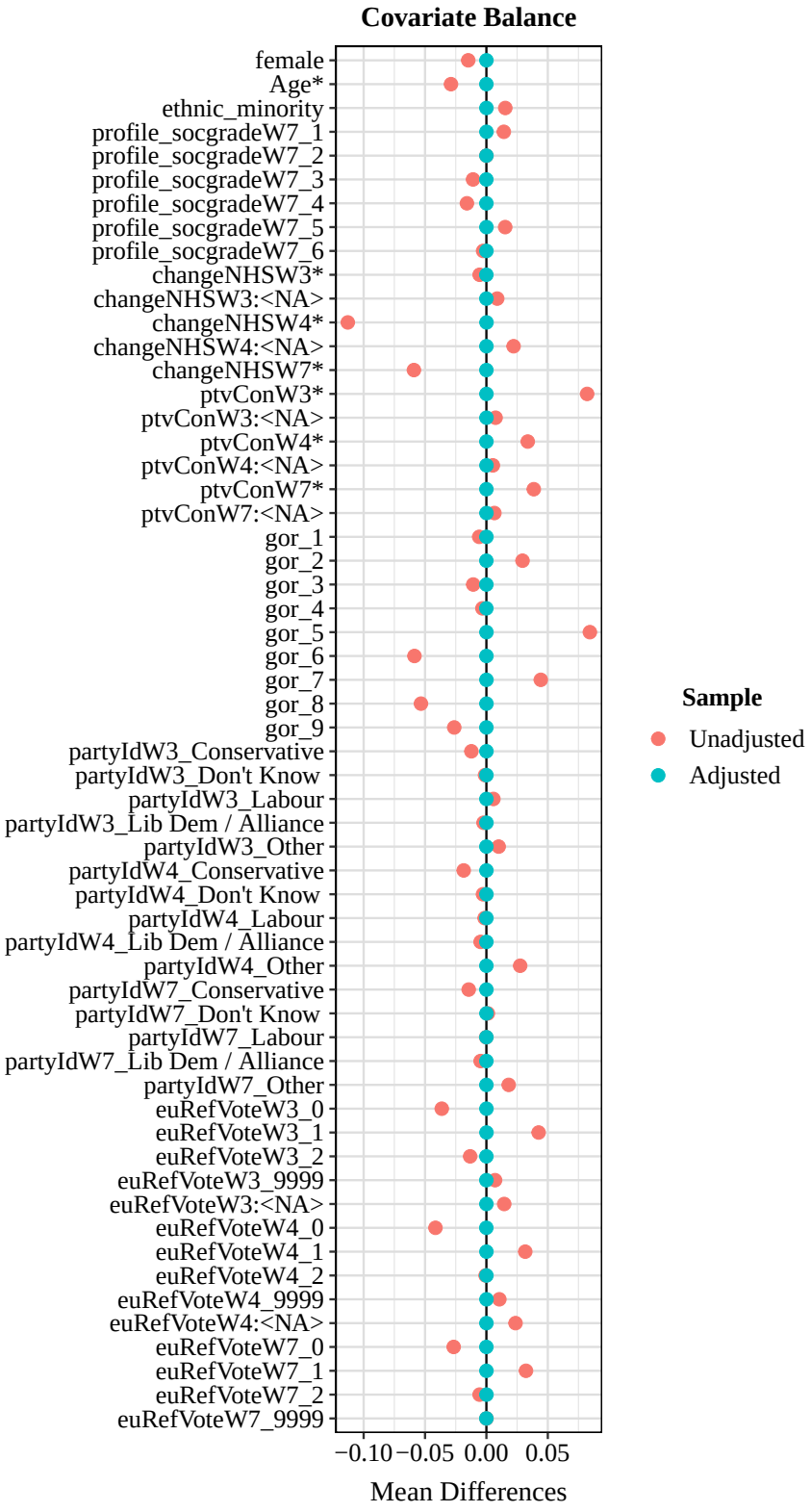

Figure 10.9: Love plot showing the covariate balance before and after entropy balancing, for the NHS change variable

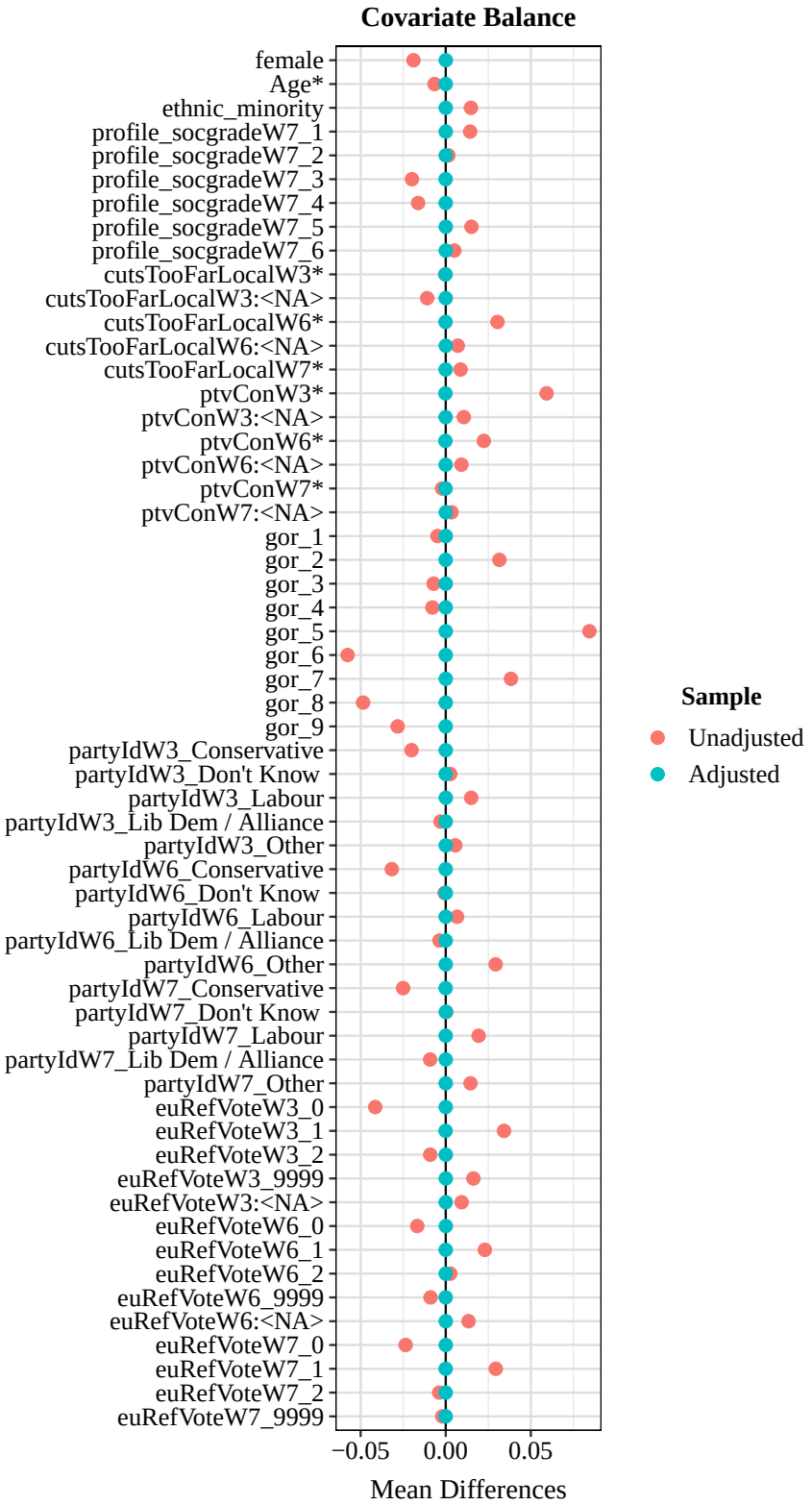

Figure 10.10: Love plot showing the covariate balance before and after entropy balancing, for the local cuts variable

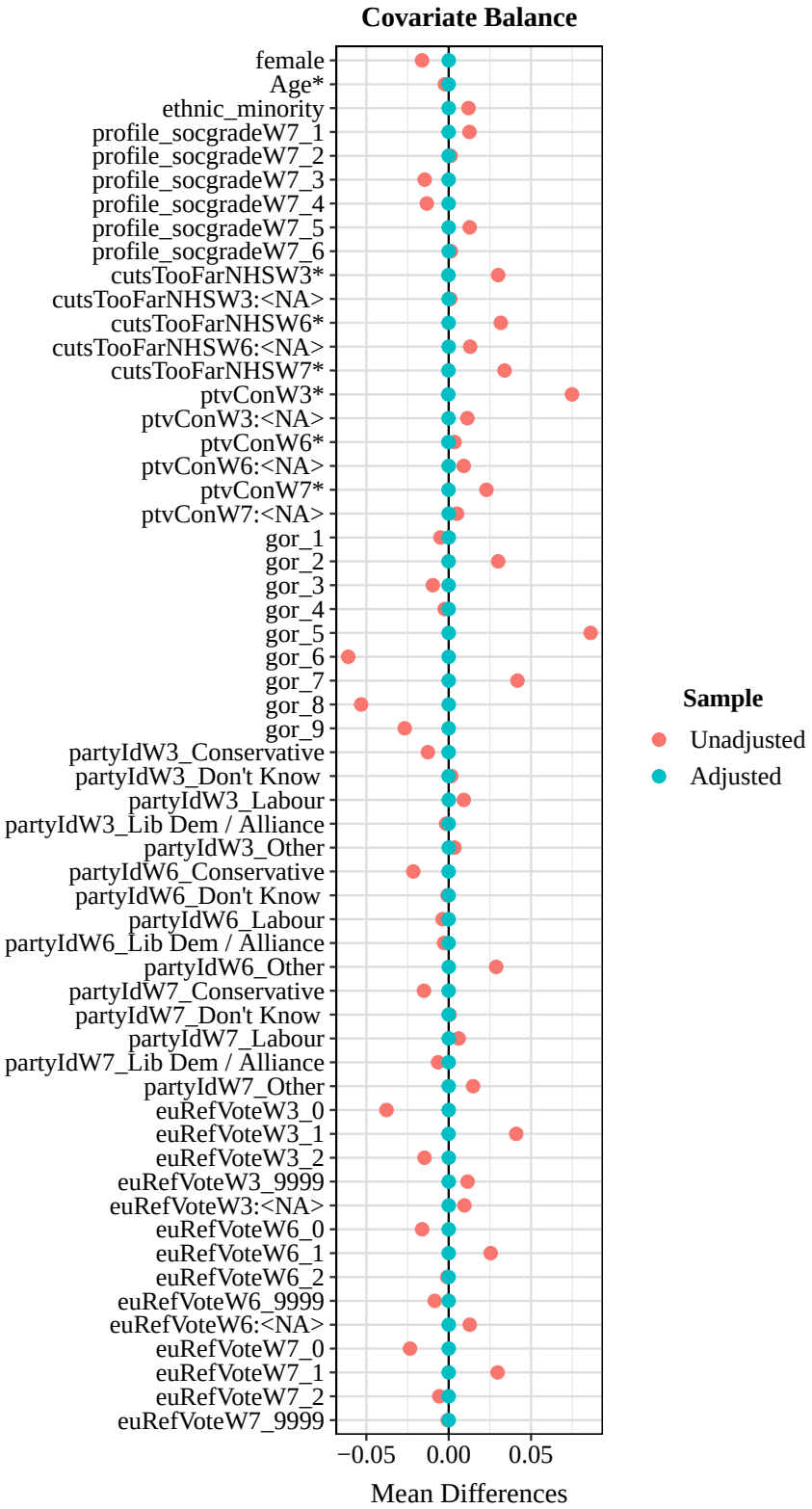

Figure 10.11: Love plot showing the covariate balance before and after entropy balancing, for the NHS cuts variable

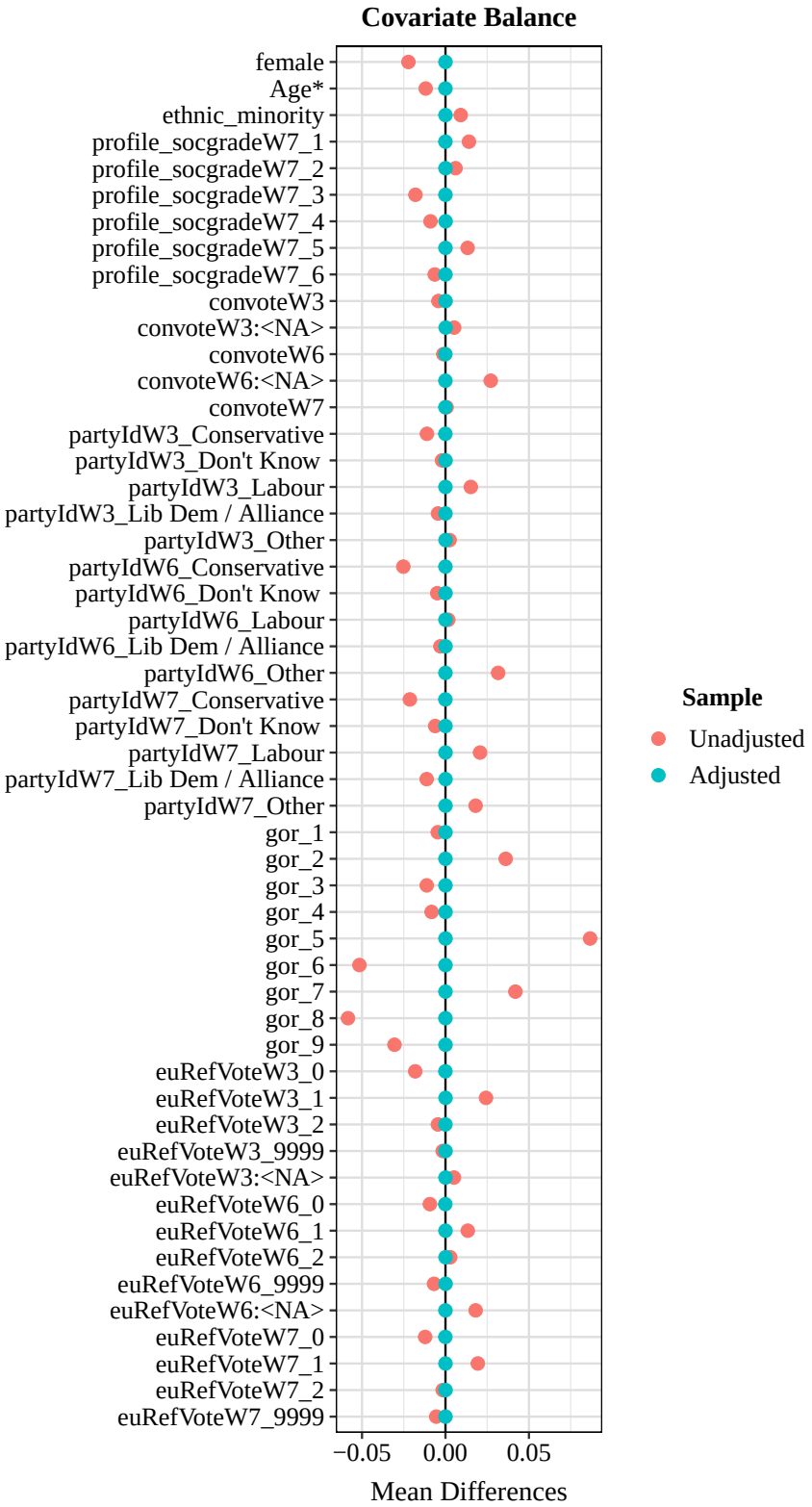

Figure 10.12: Love plot showing the covariate balance before and after entropy balancing, for the Conservative vote variable