

Unintended consequences of the 18-week referral to treatment standard in NHS England

Quinn, Laura; Bird, Paul; Remsing, Sandra; Reeves, Katharine; Lilford, Richard

DOI:

[10.1136/bmjqs-2023-015949](https://doi.org/10.1136/bmjqs-2023-015949)

License:

Creative Commons: Attribution-NonCommercial (CC BY-NC)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Quinn, L, Bird, P, Remsing, S, Reeves, K & Lilford, R 2023, 'Unintended consequences of the 18-week referral to treatment standard in NHS England: a threshold analysis', *BMJ Quality & Safety*.
<https://doi.org/10.1136/bmjqs-2023-015949>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.



OPEN ACCESS

Unintended consequences of the 18-week referral to treatment standard in NHS England: a threshold analysis

Laura Quinn ^{1,1}, Paul Bird,^{2,3} Sandra Remsing,⁴ Katharine Reeves,⁴ Richard Lilford ⁵

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2023-015949>).

¹University of Birmingham, Birmingham, UK

²Institute for Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

³West Midlands Academic Health Science Network, Birmingham, UK

⁴Health Data Science Team, Research Development & Innovation, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

⁵Institute of Applied Health Research, University of Birmingham, Birmingham, UK

Correspondence to

Laura Quinn, University of Birmingham, Birmingham, UK; l.quinn@bham.ac.uk

Received 19 January 2023

Accepted 11 April 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Quinn L, Bird P, Remsing S, *et al.* *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2023-015949

ABSTRACT

Objective In 2012, an '18-week referral to treatment standard' was introduced in England. Among people on the list of those waiting for hospital treatment at a point in time, the standard states that at least '92% of patients should have been waiting for less than 18 weeks'.

Targets can have unintended consequences, where patients are prioritised based on the target rather than clinical need. Such an effect will be evident as a spike in the number of hospital trusts at the target threshold, referred to as a threshold effect. This study examines for threshold effects across all non-specialist acute NHS England hospital trusts by financial year.

Methods A retrospective observational study of publicly available data examined waiting times for patients on the waiting list. We examined trust performance against the 92% target by financial year, from 2015/16 to 2021/22, using Cattaneo *et al*'s manipulation density test (test for discontinuity/spike in data around target threshold) for all patients and by type of treatment.

Results The proportion of NHS hospital trusts meeting the 92% target deteriorated over time. From 2015/16 to 2019/20, there was strong evidence of a threshold effect at the 92% target ($p < 0.001$). There was no evidence of a threshold effect in 2020/21 ($p = 0.063$) or 2021/22 ($p = 0.090$). Threshold effects were present across most types of treatment in 2016/17 and fewer types from 2017/18 onwards.

Conclusion We observed striking evidence of a threshold effect suggesting that while targets change behaviour, they do so in a selective way, focusing on the threshold rather than a pervasive improvement in practice. However, at the height of the pandemic, as almost no trusts could reach the target, the threshold effect disappeared.

INTRODUCTION

18-week referral to treatment standard

The constitution of the England National Health Service (NHS) affords patients 'the right to access certain services commissioned by NHS bodies within maximum waiting times'.¹ The NHS sets multiple waiting time standards including a '4-hour standard for accident and emergency (A&E) services', a '62-day standard for cancer treatment' and an '18-week referral to treatment standard'

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ The NHS referral to treatment standard states that 92% of patients should be waiting less than 18 weeks for treatment since their referral.
- ⇒ The effect of the threshold target on behaviour is unknown, particularly concerning how the target might affect hospitals that lie close to the target rather than those that have already met the target or fall far short of doing so.

WHAT THIS STUDY ADDS

- ⇒ This study shows that there was a threshold effect at the 92% target, which disappeared when most hospital trusts failed to reach the target during the COVID-19 pandemic.
- ⇒ The threshold effect suggests that while targets change behaviour, they do so in a selective way, focusing on the threshold rather than a pervasive improvement in practice.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Our findings provide further information that targets incentivise a change in behaviour, which is selective, not systemic, and suggest caution in the use of targets in the management of waiting lists.
- ⇒ If policymakers implement targets they should examine for threshold effects so that the policy can be modified to promote systematic improvement rather than effects specifically at the threshold.

across all conditions.² Sanctions for consistent failure to meet these monthly targets have varied over time. There has been a constant non-financial sanction in the form of increased performance management from regulators, potentially

leading to the down-grading of an organisation's Care Quality Commission rating.³ Initially, there was a financial sanction for any organisation failing the 92% threshold target. This financial penalty was up to 2.5% of healthcare income.⁴ These financial penalties were formally removed in 2017/18 when it was felt that, due to pressures in the delivery of the target, the fines were detracting from the ability of organisations to address backlogs in care. Some organisations experiencing significant failure had already been excused from the fines in 2016/17 provided they were meeting locally agreed trajectories for improvement.⁵

In this article, we study the 18-week referral to treatment standard introduced in 2012. The standard is based on examining the waiting list of patients waiting for treatment at a set date at the end of each month. Ideally, there should be no patients on the list that have been waiting for more than 18 weeks since referral. In which case, 100% of patients would have been waiting for 18 weeks or less. Since there will always be good reasons why some patients cannot be treated within 18 weeks, NHS England stipulates a more realistic threshold. The standard stipulates that no less than 92% of patients on the list of those waiting for treatment should have been waiting for 18 weeks or less.⁴ For hospital trusts that are not close to the threshold target, either too far away from reaching the target or surpassing the target, it is unlikely that the target will have an effect on hospital behaviour. For hospital trusts that are close to the threshold effect, the risk with such a threshold is that it could incentivise hospitals to select patients who are at risk of breaching the target over patients who have been waiting shorter times or who have already exceeded an 18-week wait.⁶ If so, this will show up as a spike in the number of hospital trusts where the target is only just met—that is, there will be a spike in hospital trusts recording 92% or 93% of patients who were waiting less than 18 weeks for treatment. Such a spike can be interpreted as a sign that the hospital was motivated by meeting this target, rather than by clinical need.

Threshold effects

One of the problems with targets is that they may direct attention to activity around the target rather than promoting a general improvement in performance. This behaviour will create a pattern in the data—a spike known as discontinuity—which will appear around the target threshold. There are examples of discontinuity in many areas other than in healthcare. An example includes a subsidy programme for grain production in China,⁷ where there was a large spike in farms producing just over 200 000 tons of grain required to trigger a subsidy. Another example is a programme in the European Union, where costs for tenders for public procurement spiked just below the target, which triggered a more detailed review.⁸ There is a large body of literature on performance targets and their effects in

healthcare but very few studies where threshold targets have been examined statistically. The NHS 4-hour A&E service standard⁹ and the Quality and Outcome Framework requirement to measure blood pressure in primary care¹⁰ both recorded spikes at their respective performance thresholds. However, they did not use formal statistical methods to check for discontinuity at the threshold targets such as those developed by McCrary¹¹ and Cattaneo *et al.*¹² An exception is a study examining NHS staff influenza vaccination rates¹³ that used McCrary's test to track the effects of the target as it was changed by the government from year to year.

In this article, we use Cattaneo *et al.*'s test to examine behaviour around the 92% target. Across all non-specialist acute hospitals in the English NHS, we hypothesised that there would be a discontinuity with a peak of hospitals just meeting the target of having no less than 92% of patients waiting less than 18 weeks for treatment since referral. A formal explanation of the methods we will use to check for a threshold effect is given below.

The type of sanction for failing to meet the threshold target changed across the study period. In 2017/18, the sanctions changed from financial and non-financial to purely non-financial. This change could lead to a change in behaviour around the target threshold. Furthermore, the COVID-19 pandemic also started in 2020/21 making it extremely difficult to meet the target. We therefore also examined for an attenuation of the threshold effect at the 92% target over time, in particular with reference to the period before and after the change in the financial sanction and after the COVID-19 pandemic started.

METHODS

Study design

This study is a retrospective observational study of referral to treatment times in all non-specialist acute NHS England hospital trusts from January 2016 to September 2021. Data are available for hospitals (including non-specialist acute, specialist and independent) on the NHS England website (<https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/>) and revisions are published periodically (usually every 6 months) in line with the revisions policy.¹⁴ Data were extracted for non-specialist acute NHS hospital trusts only. Data could not be included if a hospital trust made no submission in a particular month. Data before 2016 were not extracted as the format did not match later years. Results have been reported in line with the Reporting of studies Conducted using Observations Routinely-collected Data (RECORD), an extension of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) standards.

Data collection

The data are a series of point prevalence estimates ('snap-shots') of waiting times. At the end of each

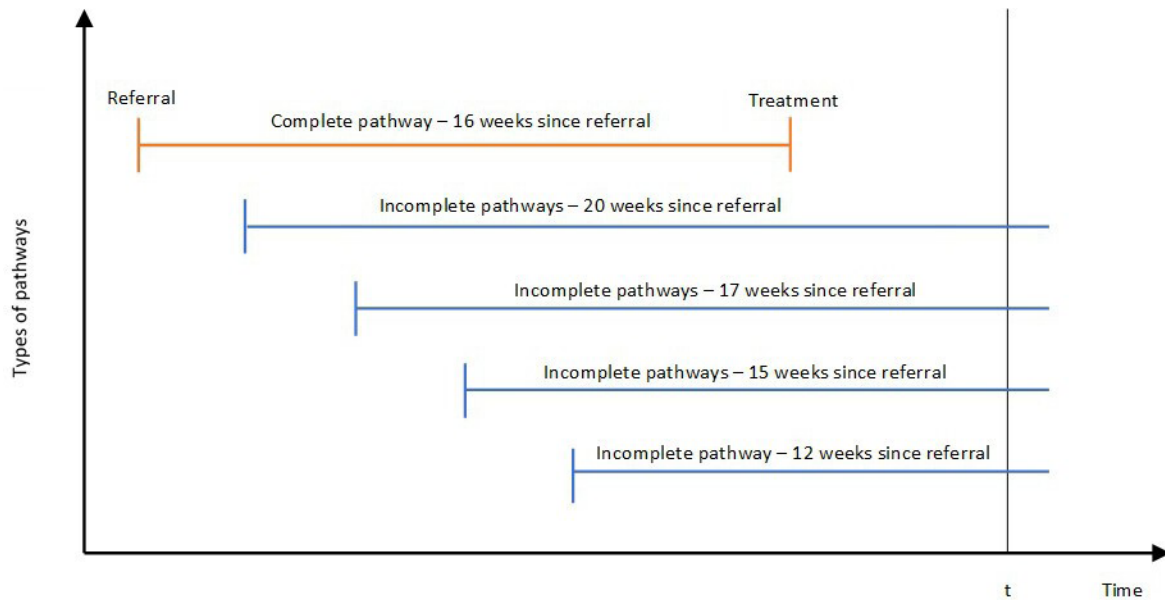


Figure 1 Example of pathways in NHS waiting times. This is a snapshot of waiting times for 1 month in one hospital trust; time *t* refers to the end of the month. Each line refers to a hypothetical patient pathway. The blue lines refer to incomplete pathways used for the main analysis. At the end of the month, three of the four patients have been waiting less than 18 weeks and one patient has been waiting more than 18 weeks. The orange line refers to a completed pathway for one patient who waited 16 weeks to start treatment after their referral (not counted in the main analysis).

month, each NHS hospital trust collects data on waiting times for all patients who are still on the waiting list and submits the data to NHS England.¹⁰ A patient pathway refers to the route or path that a patient will take if they are referred for treatment after visiting a referring doctor. The patient pathway can end if a clinician decides no treatment is necessary, the patient decides they do not want to be treated, or the treatment starts. Treatments can refer to being admitted to a hospital for treatment or operation, starting medication, fitting a medical device, agreeing to monitor the condition (to decide if further treatment is needed) or receiving advice from a clinician on how to manage a condition. Incomplete pathways refer to the waiting times for all patients that are still waiting to start treatment at the end of each month. Completed pathways refer to the waiting times (the time waited since referral) for all patients who have started their treatment during the month. A graphical description of the pathways is given in [figure 1](#).

For this study, we focus on incomplete pathways as these are used to measure the performance of hospital trusts. Thus, at least 92% of people with incomplete pathways should have been waiting for less than 18 weeks for the target to be met. The denominator is all patients still on the waiting list and the numerator is patients who have been waiting less than 18 weeks. The percentage of patients with incomplete pathways who had been waiting for less than 18 weeks for treatment since referral was calculated for each hospital trust in each month of the study period. There were 144 trusts (although the number of trusts varied by month due to organisational changes or missing data).

Statistical analysis

Data are published on the NHS website for each hospital trust on a monthly basis. These data were combined over the study period. The raw data give the number of patients with incomplete pathways who have been waiting for different numbers of weeks, from 1 to 52. First, we examine the average length of wait for patients with incomplete pathways. Second, we calculate the percentage of patients with incomplete pathways each month who have been waiting for less than 18 weeks for treatment since referral. Third, we averaged the monthly data over calendar years to test for a change in any threshold effect from one financial year to another.

We performed Cattaneo *et al's* manipulation test to check for evidence of discontinuity in the percentage of patients waiting less than 18 weeks at the target threshold of 92%. If there is no systematic manipulation at the target threshold, then the density should be continuous near the chosen target threshold and there will be no discontinuity (see [figure 2](#) for example). If there is systematic manipulation, then there will be a spike at the target threshold showing evidence of discontinuity, which we refer to as a threshold effect. McCrary developed a test to check for data manipulation at a target threshold. This was done by pre-specifying the area around the target (bandwidth) to test for discontinuity. Cattaneo *et al's* manipulation does not require this pre-specification of bandwidths and uses local polynomial density estimators, which improves the precision of the test.¹² Cattaneo *et al's* test produces a figure with a histogram of densities around the specified target threshold, local polynomial density

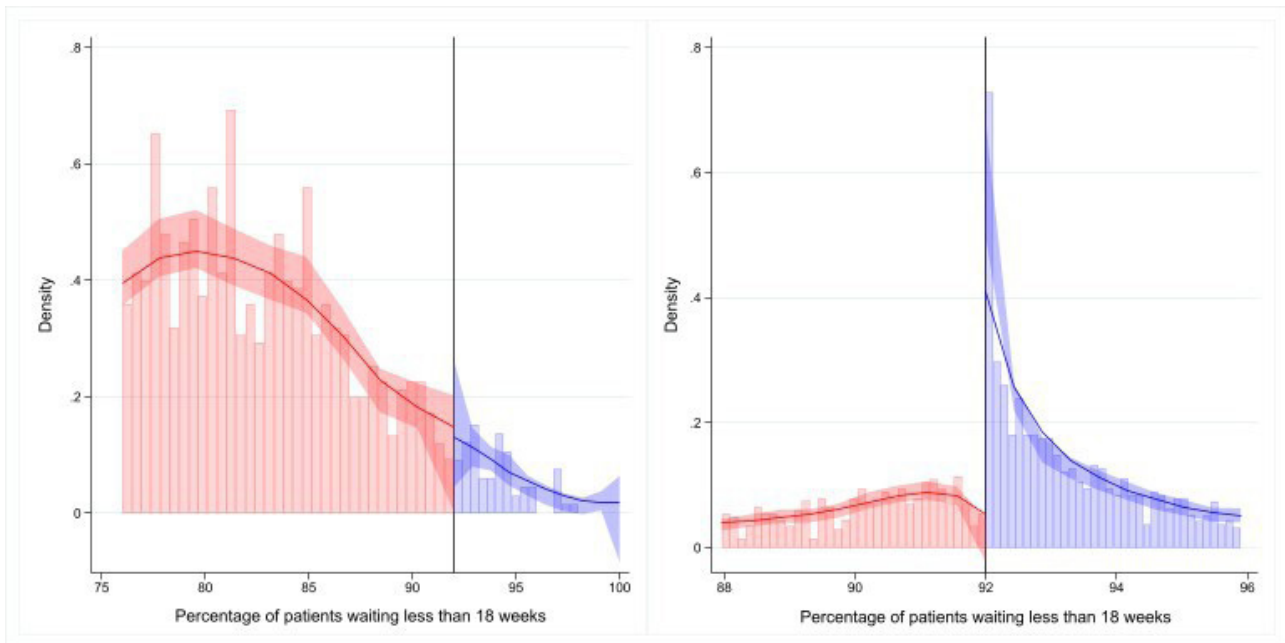


Figure 2 Hypothetical example showing when no evidence of discontinuity (left) and evidence of discontinuity (right). The figure includes a histogram of densities around the specified target threshold of 92%, the red and blue lines represent the local polynomial density estimate and the shaded area represents the 95% CI. Each bar refers to the number of hospitals with the percentage of patients waiting less than 18 weeks. The black line represents the target threshold, red bars represent trusts that did not meet the target and blue bars represent trusts that met the target. On the left side there is a continuous decrease in histogram showing no evidence of discontinuity. On the right side there is a spike in hospital trusts meeting the target, followed by a sharp decline, suggesting evidence of discontinuity.

estimates and 95% confidence intervals. The bandwidth estimates, t-statistic and p value are reported. Bandwidth estimates are a statistical measure of the width of the area around the target threshold where a threshold effect will appear if present. The t-statistic describes the level of evidence against the null hypothesis of no discontinuity.

These analyses were carried out in all patients and then repeated in the following treatment groups: cardiology; cardiothoracic; dermatology; ear, nose and throat; gastroenterology; general surgery; gynaecology; neurosurgery; ophthalmology; oral surgery; plastic surgery; trauma and orthopaedics; and urology.

RESULTS

Incomplete waiting times for all patients

The number of non-specialist acute hospital trusts ranged from 128 to 139 across financial years due to organisational changes (mergers/de-mergers) or trusts not submitting their data. Data are summarised by financial year. This resulted in two incomplete years, with 3 months of data for 2015/16 and 6 months of data for 2021/22. The average percentage of patients waiting for treatment each week since referral across months in hospital trusts are summarised by financial year in online supplemental file 1. On average each month, hospital trusts had 87% of their patients

Table 1 Percentage of patients waiting less and more than 18 weeks for treatment since referral. Averaged across hospital trusts and months by financial year

Financial year	Percentage of patients waiting			
	<18 weeks			>18 weeks
	Median (IQR)	Min	Max	Median (IQR)
2015/16	92 (91–94)	73	99	8 (6–9)
2016/17	92 (90–93)	72	100	8 (7–10)
2017/18	91 (87–93)	66	99	9 (7–13)
2018/19	89 (84–92)	68	99	11 (8–16)
2019/20	85 (81–91)	58	99	15 (9–19)
2020/21	64 (56–71)	28	99	36 (29–44)
2021/22	70 (63–76)	41	100	30 (24–37)
All	87 (77–92)	28	100	13 (8–23)

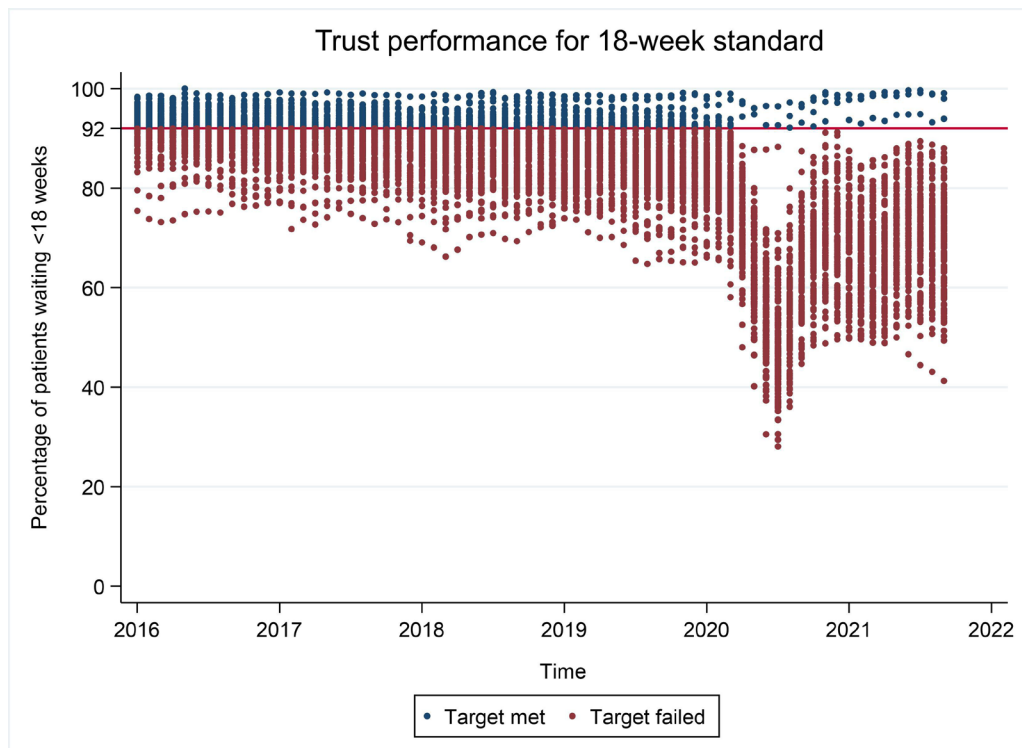


Figure 3 Percentage of patients waiting for treatment less than 18 weeks by trust each month from 2016 to 2021. Red line represents target, red points represent hospital trusts that failed to meet target and blue points represent hospital trusts that met the target.

waiting less than 18 weeks for treatment since referral (IQR 77–92%) (table 1). Across individual trusts, the proportion meeting the 92% standard decreased over time from a high of 92% (IQR 91–94%) in 2015/16 to a low of 64% (IQR 56% to 71%) in 2020/21. In figure 3, we show the percentage of patients who wait less than 18 weeks by trust, by month over the whole study period. The percentage of trusts where patients wait less than 18 weeks declines gradually from 2015/16 to 2019/20 after the removal of the financial sanction in 2016/2017, and then declines precipitously over the first COVID-19 epoch in 2020/21, before making a gradual recovery. When looking at the hospital trusts separately (see online supplemental file 2), we can see that the majority of hospital trusts had a slowly decreasing percentage of patients waiting less than 18 weeks for treatment until 2020/21 when there was a large decline.

Threshold effect

In figure 4, we examine for a threshold effect. Over the financial years from 2015/16 to 2019/20 there was a large spike in the number of trusts exactly meeting the 92% target threshold for the 18-week referral to treatment with a sharp drop after the target has been reached. In figure 4 we can see this spike reduces slightly in magnitude from 2016/17 to 2017/18 (when the financial sanction was withdrawn) followed by small reductions in 2018/19 and 2019/20. Nevertheless, striking threshold effects are in evidence despite the substantial decline in the proportion of hospitals

meeting the target. In 2020/21 and 2021/22 there is no visual evidence of a threshold effect as the majority of trusts did not meet the target threshold.

Effect of target threshold for all incomplete pathways

In figure 5 we can see that from 2015/16 to 2019/20 the number of hospitals failing to meet the 92% target is relatively stable across different percentages of patients as we read from left to right. However, when the percentage comes close to the target threshold, there is a reduction in the number of hospitals failing to meet the target. Then, at the 92% target, there is a high spike of hospitals reaching the target, which then rapidly tails off. This large spike in the density at the threshold is indicative of a threshold effect. Cattaneo *et al*'s manipulation test confirms that there is strong evidence of a threshold effect from 2015/16 to 2019/20 ($p < 0.001$) (table 2), even after the financial component of the sanction ceased in 2017/18. In 2020/21 and 2021/22, few trusts met the target threshold of 92% and no clear discontinuity is evident in the figure. The manipulation test confirms that there is no evidence of a threshold effect ($p > 0.05$).

Effect of target threshold for incomplete pathways by treatment group

Cattaneo *et al*'s manipulation test was performed for a set of treatment groups by financial year (p values are presented in online supplemental file 3). In 2015/16 there is evidence of a threshold effect in some of the treatment groups; however, we only have a few

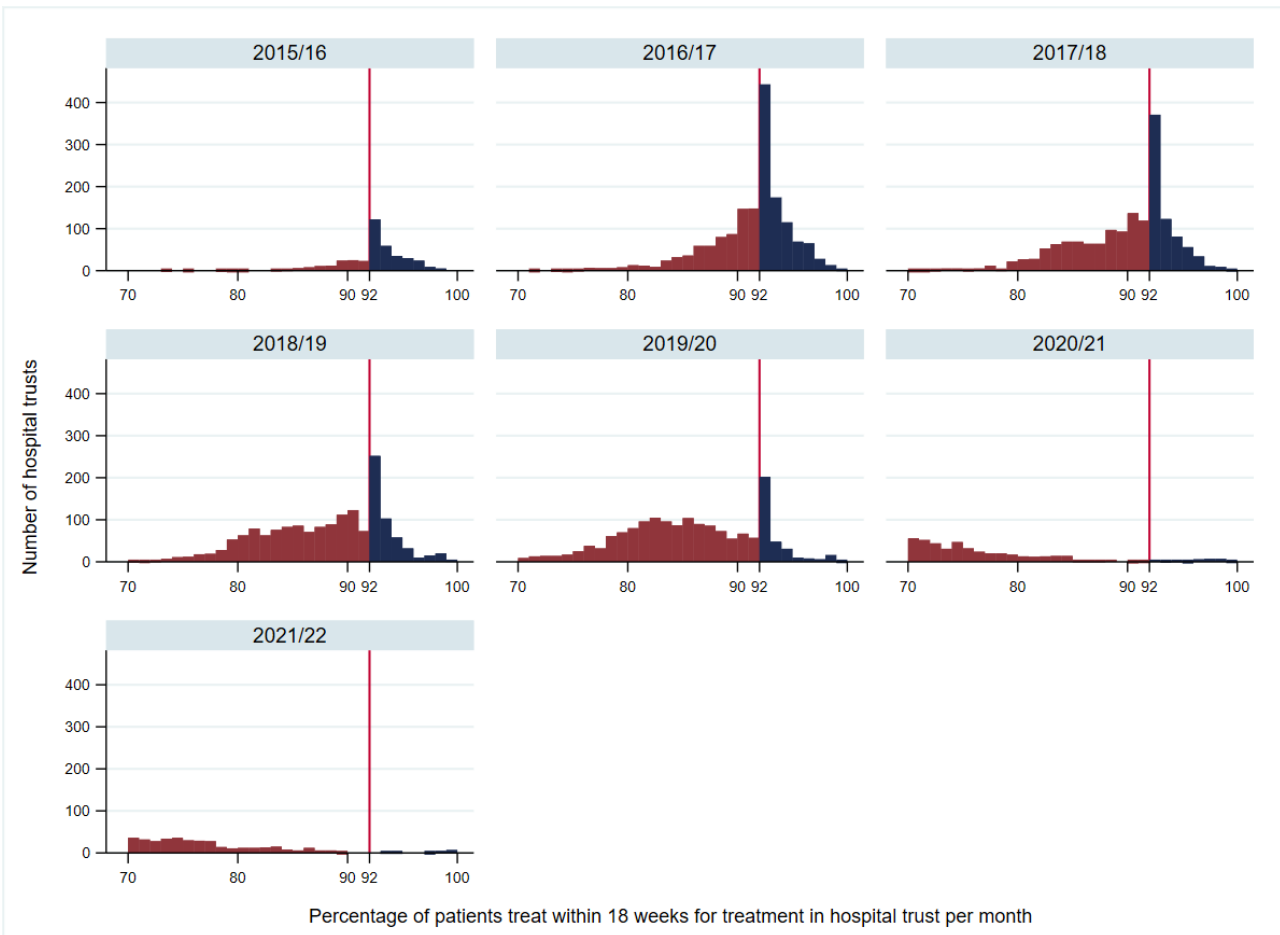


Figure 4 Bar graphs of percentage of patients waiting less than 18 weeks from 2016 to 2021. Each observation is the percentage of patients waiting less than 18 weeks in a hospital trust in 1 month. Red line represents target, red bars represent number of hospital trusts failing to meet target and blue bars represent number of hospital trusts meeting target.

months of data for this financial year. In 2016/17, there is evidence of a threshold effect at the target threshold in all treatment groups apart from gastroenterology, ophthalmology, plastic surgery and trauma. From 2017/2018 onwards there is a slight drop in the proportion of treatment groups where the threshold effect is still significant, but the numbers are too small to confirm this statistically.

DISCUSSION

Key results

The NHS has an ‘18-week referral to treatment standard’ stating that, at any one time, 92% of patients on the waiting list should have been waiting less than 18 weeks since referral for treatment. The percentage of trusts meeting the standard has decreased over time, especially during the COVID-19 pandemic in 2020/21 and 2021/22. From 2015/16 to 2019/2020 there is evidence of a threshold effect at the target threshold of 92%, although this phenomenon disappeared during the COVID-19 pandemic in 2020/21 and 2021/22.

From 2015/16 to 2019/20, there is strong evidence of a threshold effect at the target threshold, which is visible in [figure 5](#). There is stable density in the

percentage of patients waiting less than 18 weeks since referral for treatment before the target threshold, which then sharply increases at the 92% target, followed by a sharp decrease right after the target. This suggests some trusts treat the minimum number of patients waiting under 18 weeks to meet the target and that there is no incentive to treating patients waiting more than 18 weeks once the target is met. In 2020/21 and 2021/22, elective activity was severely curtailed by the COVID-19 pandemic to the extent that there was no possibility of a threshold effect (as seen in [figure 4](#)).

Causes of threshold effects

There are two types of behaviour by hospital trusts that may underlie threshold effects.¹³ The first behaviour concerns how hospital trusts perform in relation to achieving the target. Some hospital trusts near the 92% target will take action to clear the target, while those further from the target do not take such action (or are unsuccessful in doing so). We think the most plausible explanation is that those further from the target are discouraged, feeling further effort is futile, in line with expectancy theory.¹⁵ The effect on the system as a whole may be negative since targets are

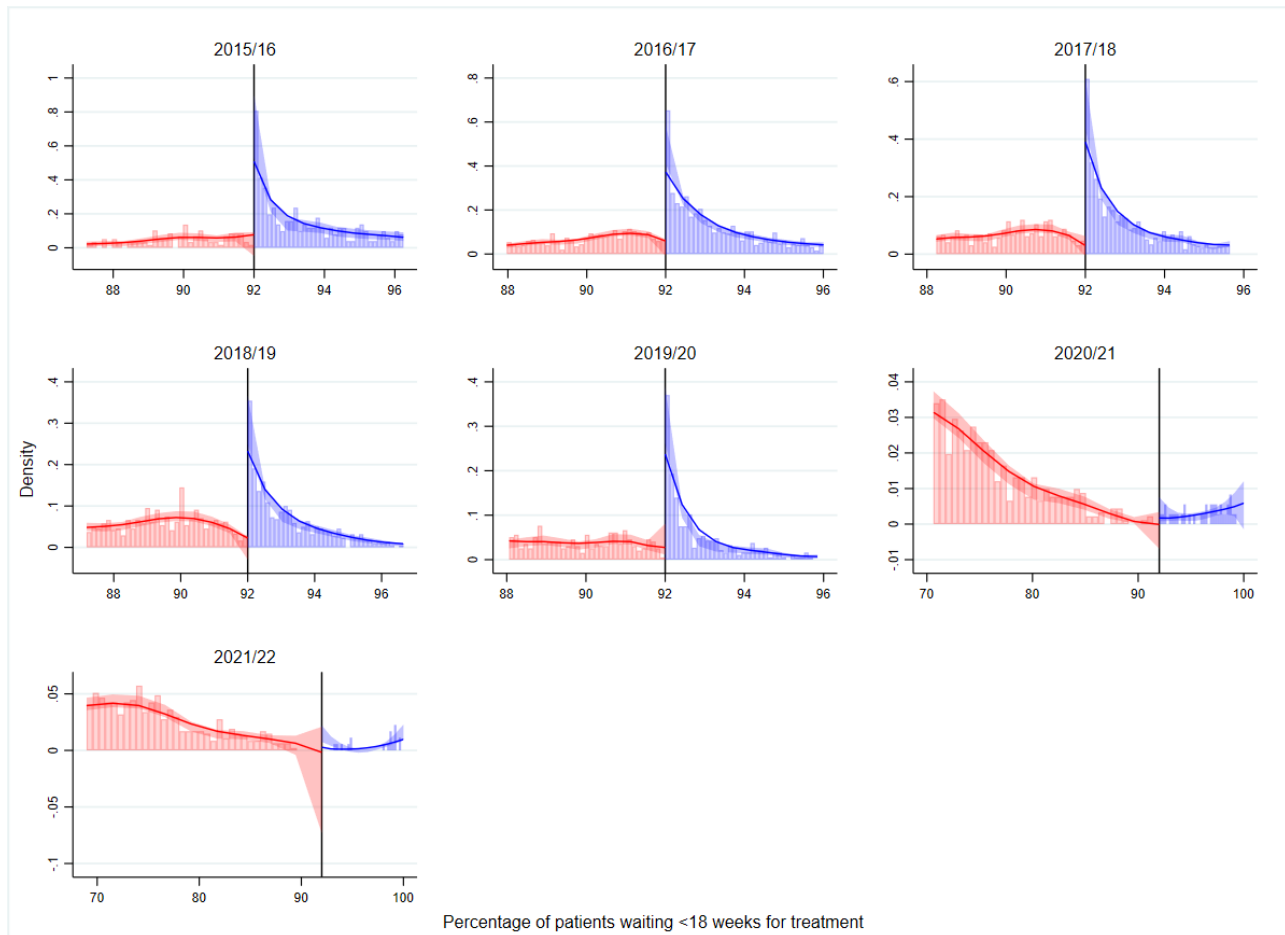


Figure 5 Cattaneo *et al*'s manipulation test at the 92% target for the 18-week referral to treatment standard by financial year. The figure includes a histogram of densities around the specified target threshold of 92%, the red and blue lines represent the local polynomial density estimate and the shaded areas represent the 95% CI. Each observation is the percentage of patients waiting less than 18 weeks in a hospital trust in 1 month. The black line represents the target threshold, red bars represent trusts that did not meet the target and blue bars represent trusts that met the target.

a form of extrinsic motivation and such motivation may supplant intrinsic motivations as in the famous case of financial rewards for blood donation.¹⁶ Our findings suggest that hospital trusts may choose whom to treat based on the target instead of clinical need, as the target provides no encouragement to treat patients who have only been waiting for a short time or to treat

patients who have already passed the 18-week wait. A recently published review suggested inappropriate or sub-optimal care, reduction in patient-centred care and exacerbation of inequalities are some of the possible unintended negative consequences of threshold targets for patients.¹⁷

The second behaviour refers to how hospital trusts report the data used for measuring performance in relation to the target. A hospital trust may be encouraged to be more accurate in reporting its data or it could lead to a hospital trust manipulating its data to meet the target.¹⁸ In this case, the target could be incentivising low rectitude behaviour. It is noteworthy that overall performance declined progressively over the pre-COVID-19 epoch. As it did so the threshold effect declined; spikes became slightly smaller and the threshold effect was apparent over a slightly smaller number of treatment categories. There are two possible non-exclusive reasons for the modest drop in threshold effects over the pre-COVID-19 epoch. First, pressure to reach the target may have declined somewhat as the proportion of peer hospitals failing to reach the target increased (a possibility consistent with

Table 2 Cattaneo *et al*'s manipulation test at the 92% target for the 18-week referral to treatment target by financial year

Financial year	Bandwidth estimate*		t-statistic	Robust p value
	Left cut-off	Right cut-off		
2015/16	1.59	1.42	6.32	<0.001
2016/17	1.34	1.34	9.22	<0.001
2017/18	1.35	1.22	9.53	<0.001
2018/19	1.60	1.55	8.66	<0.001
2019/20	1.31	1.28	6.77	<0.001
2020/21	7.13	7.13	1.86	0.063
2021/22	7.68	7.68	1.70	0.090

*Bandwidth estimates are a statistical measure of the width of the area around the target threshold where a threshold effect might occur. The bandwidth estimates increase as the number of hospitals near the target decreases (see figure 5).

expectancy theory mentioned above). Second, the financial sanction was dropped after 2016/17, leaving only the arguably attenuated bureaucratic and reputational sanction.

Comparison with the literature and importance of results

Target thresholds are used in many sectors including in healthcare systems; however, relatively few studies examine for a threshold effect. An example in healthcare is the uptake of influenza vaccination by NHS staff which showed there was a threshold effect and some trusts would just achieve the target, tracking the target as it changed over time.¹³ In the NHS, there are multiple waiting time targets used to measure hospital trust performance, such as the 4-hour standard for A&E services and the 62-day standard for cancer treatment as well as many other performance measures, which have not been examined for threshold effects. For the 18-week referral to treatment standard, there are possible negative unintended effects as outlined above. First, policymakers should be circumspect in their use of targets. Second, if targets are used, then policymakers should examine for threshold effects routinely. Third, targets should be carefully designed to mitigate threshold effects, say, by using multiple thresholds with different rewards/penalties.

Strengths and limitations

One strength of this study is that the data are publicly available and cover all non-specialist NHS acute hospital trusts in England. This study also used a recently developed manipulation test by Cattaneo *et al*, which requires little pre-specification, unlike the original McCrary test.

A disadvantage of this study is that pathways are only split by certain treatment groups and not by specific treatment types. As this was a purely quantitative study, the behavioural mechanisms and motivations that might explain the threshold effect could not be investigated. For example, we were unable to investigate the reasons for the persistence of the threshold effect even when the financial sanctions were removed. We speculate that hospitals were motivated by the desire to avoid unfavourable comparison with peer institutions. The COVID-19 pandemic has meant the majority of trusts have not met the target threshold since March 2020, obviating a threshold effect, but this provides an interesting case study in behaviour in the face of unattainable targets.

CONCLUSION

There was strong evidence of a threshold effect at the 92% target for the 18-week referral to treatment standard from 2015/16 to 2019/20. There was no evidence of a threshold effect in 2020/21 and 2021/22, likely due to most trusts failing to meet the target. While targets may improve hospital trust

performance in healthcare systems, our data suggest that this comes at a cost and that hospitals may target specific patients rather than implementing systematic change, which can help all patients and respect clinical judgement.

Twitter Richard Lilford @rjlilford

Contributors RL conceived and designed the study. LQ drafted the first version of the manuscript and completed the data analysis. RL, PB, SR and KR were involved in writing and reviewing the manuscript. LQ acts as guarantor.

Funding This study was funded by National Institute of Health Research Applied Research Collaboration West Midlands (grant number: NIHR200165).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The data used in this study are aggregated NHS monthly waiting times which are freely available to download from the following website: <https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/>.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Laura Quinn <http://orcid.org/0000-0001-9660-4631>

Richard Lilford <http://orcid.org/0000-0002-0634-984X>

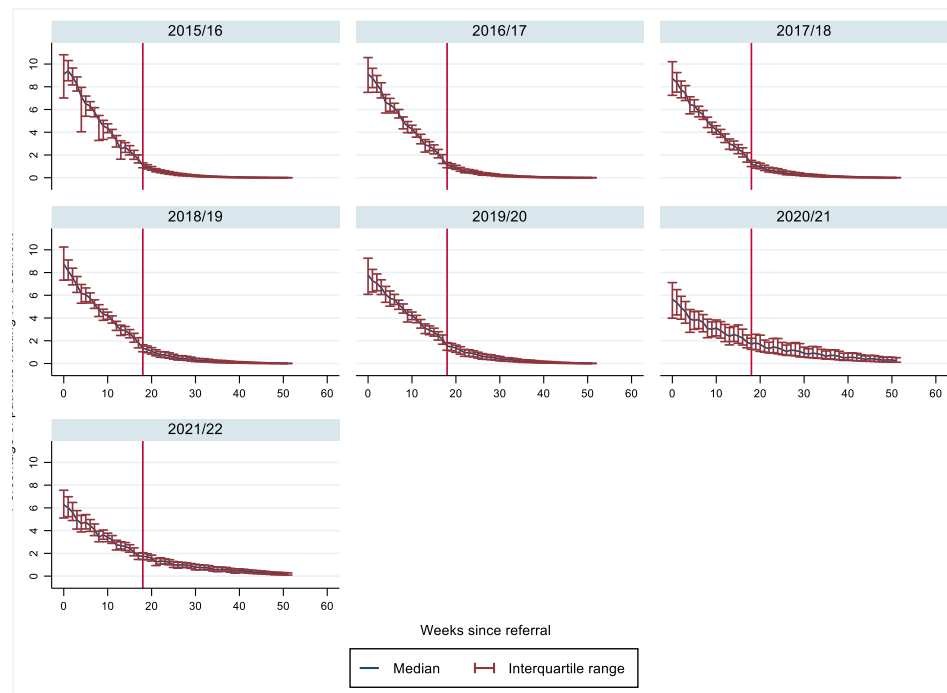
REFERENCES

- 1 England NHS. Referral to treatment. n.d. Available: <https://www.england.nhs.uk/rtt/>
- 2 Thorlby R, Gardner T, Turton C. *NHS performance and waiting times*. The Health Foundation, 2019.
- 3 UK Care Quality Commission. *Care Quality Commission - find and compare services*. 2022.
- 4 Insource. "18-weeks" penalties change again: this time it's good. 2013. Available: <https://www.insource.co.uk/18-weeks-penalties-change-again-this-time-its-good/>
- 5 NHS providers. Changes to performance fines. 2017. Available: <https://nhsproviders.org/media/3620/changes-to-performance-fines.pdf>
- 6 Iacobucci G. Hospital treatment targets have been simplified. *BMJ* 2015;350:h3090.

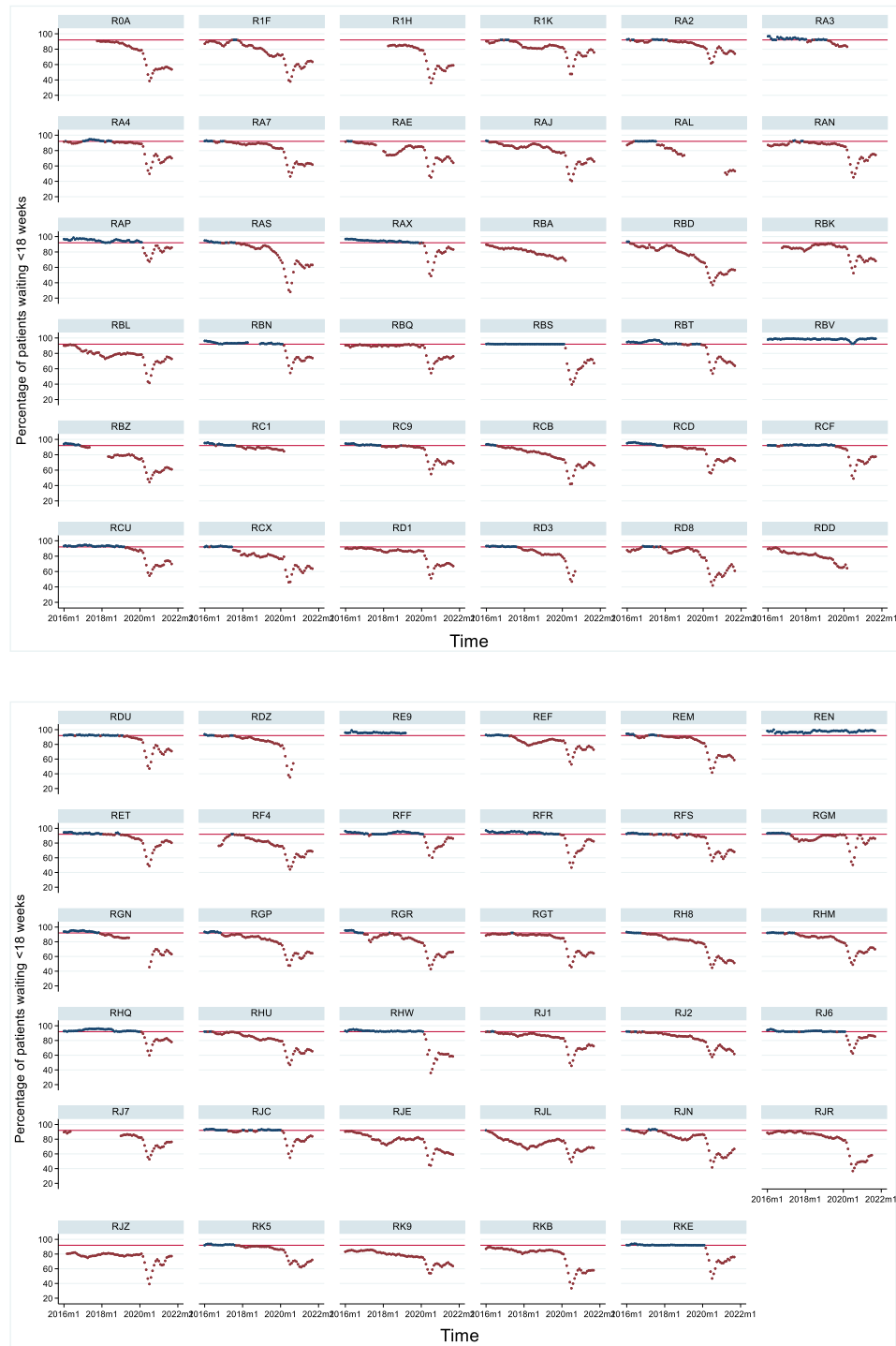
- 7 Zhang X, Yu X, You L. Does the Granary County Subsidy Program lead to manipulation of grain production data in China? *China Economic Review* 2020;62:101347.
- 8 Tas BKO. Bunching below thresholds to manipulate public procurement. *Empir Econ* 2023;64:303–19.
- 9 Gruber J, Hoeh T, Stoye G. *Saving lives by tying hands: the unexpected effects of constraining health care providers*. National Bureau of Economic Research, Inc, 2018.
- 10 Wilding A, Munford L, Guthrie B, *et al*. Family doctor responses to changes in target stringency under financial incentives. *J Health Econ* 2022;85:102651.
- 11 McCrary J. Manipulation of the running variable in the regression discontinuity design: a density test. *Journal of Econometrics* 2008;142:698–714.
- 12 Cattaneo MD, Jansson M, Ma X. Manipulation testing based on density discontinuity. *The Stata Journal* 2018;18:234–61.
- 13 Liaquat A, Gallier S, Reeves K, *et al*. Examining organisational responses to performance-based financial incentive systems: a case study using NHS staff influenza vaccination rates from 2012/2013 to 2019/2020. *BMJ Qual Saf* 2022;31:642–51.
- 14 England NHS. Consultant-led referral to treatment waiting times. 2022. Available: <https://www.england.nhs.uk/statistics/statistical-work-areas/rtt-waiting-times/>
- 15 Locke EA, Latham GP. *A theory of goal setting & task performance*. Prentice-Hall Inc, 1990.
- 16 Titmuss RM. *The gift relationship: from human blood to social policy*. London: Alien & Unwin, 1970.
- 17 Li X, Evans JM. Incentivizing performance in health care: a rapid review, typology and qualitative study of unintended consequences. *BMC Health Serv Res* 2022;22:690.
- 18 Mears A. Gaming and targets in the English NHS. *Universal Journal of Management* 2014;2:293–301.

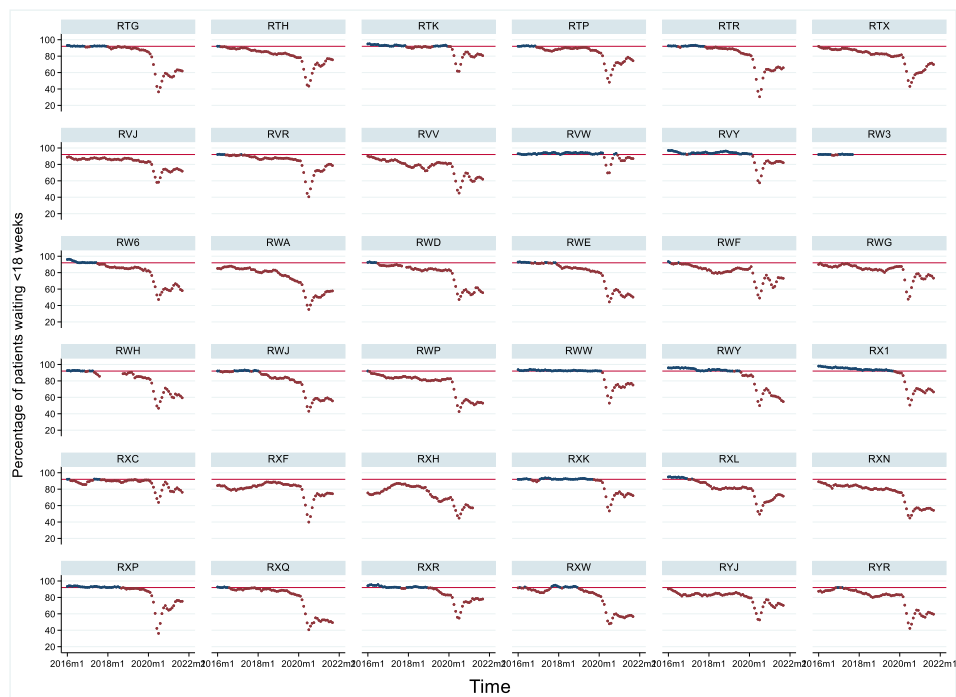
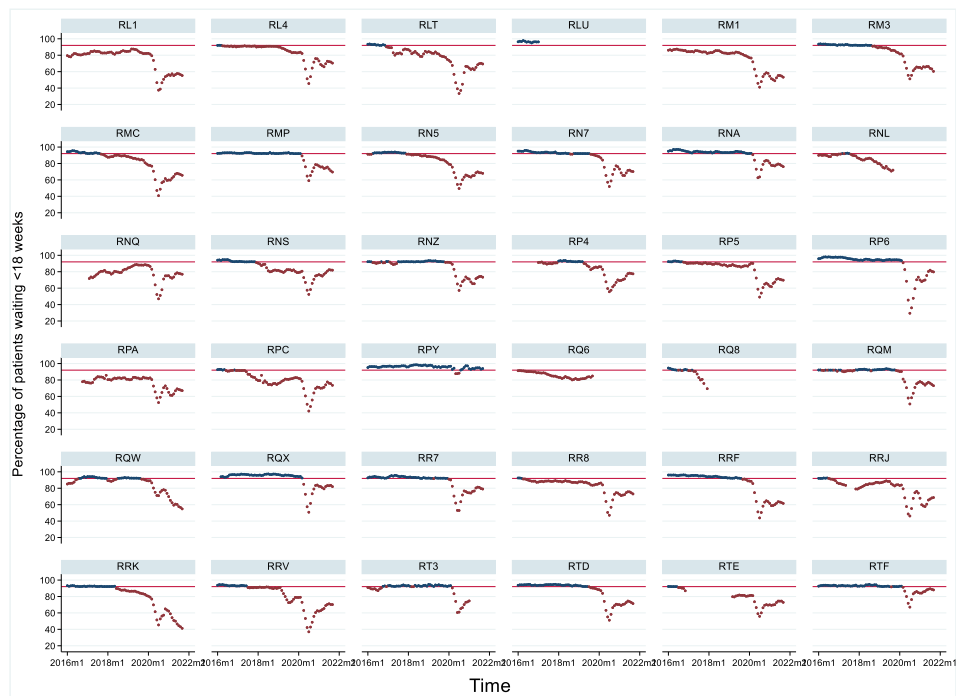
Supplementary Files

Supplementary File 1. Average percentage of patients waiting for treatment each week since referral across hospital trusts and months by financial year. Median and interquartile range given.



Supplementary File 2. Percentage of patients waiting less than 18 weeks for treatment by trust from 2016 to 2021. Red line represents target, red points represent when hospital trust failed to meet target and blue points represent when hospital trust met target.





Supplementary File 3. Cattaneo's et al manipulation test by financial year and treatment group

Treatment group	Robust p-value						
	2015/6	2016/7	2017/18	2018/19	2019/20	2020/21	2021/22
Cardiology	0.041	0.004	0.104	0.162	0.456	0.010	0.730
Cardiothoracic	<0.001	0.042	0.096	0.511	0.731	0.754	<0.001
Dermatology	0.731	0.020	0.923	0.966	0.336	0.911	0.331
ENT*	0.565	0.001	0.178	0.071	0.199	0.168	0.931
Gastroenterology	0.028	0.180	0.252	0.006	0.003	0.500	0.612
General surgery	0.194	<0.001	0.002	0.014	0.012	0.409	NR
Gynaecology	0.014	0.001	0.018	0.039	0.348	0.808	0.186
Neurosurgery	0.006	0.043	0.414	0.145	0.094	0.362	NR
Ophthalmology	0.437	0.086	0.936	0.916	0.654	0.323	0.396
Oral surgery	0.193	0.017	0.227	0.001	0.028	0.785	0.587
Plastic surgery	0.077	0.973	0.652	0.237	0.709	0.983	0.958
Trauma**	0.119	0.199	0.014	<0.001	0.002	NR	NR
Urology	0.027	<0.001	0.264	0.084	0.009	0.031	0.215

* Ear, Nose and Throat ** Trauma and orthopaedics.

NR – not reported due to low number of observations

Green shading refers to p-values <0.01, yellow shading refers to p-values <0.05