CEPA
economics matters

Rating the rankings: How are the NHS acute league tables performing?

9 September 2026

WHITE PAPER

Important notice

This document was prepared by CEPA for the exclusive use of the recipient(s) named herein on the terms agreed in our contract with the recipient(s).

CEPA does not accept or assume any responsibility or liability in respect of the document to any readers of it (third parties), other than the recipient(s) named in the document. Should any third parties choose to rely on the document, then they do so at their own risk.

The information contained in this document has been compiled by CEPA and may include material from third parties which is believed to be reliable but has not been verified or audited by CEPA. No representation or warranty, express or implied, is given and no responsibility or liability is or will be accepted by or on behalf of CEPA or by any of its directors, members, employees, agents or any other person as to the accuracy, completeness or correctness of the material from third parties contained in this document and any such liability is expressly excluded.

The findings enclosed in this document may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties.

The opinions expressed in this document are valid only for the purpose stated herein and as of the date stated. No obligation is assumed to revise this document to reflect changes, events or conditions, which occur subsequent to the date hereof.

Contents

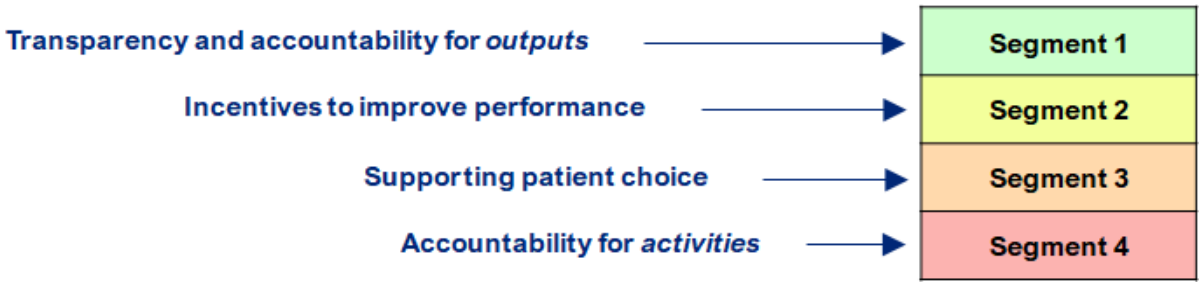
SUMMARY

> Good performance should be incentivised and rewarded. Poor performance should be held rigorously to account. And transparency and choice are essential, not nice to have. That's what lay behind our decision to publish new NHS league tables.
>
> **Wes Streeting**
> Previous Secretary of State for Health and Social Care
> 25th November 2025, address to the NHS Providers Conference[1]

Exactly one year ago, on 9th September 2025, the NHS league tables were (re-)introduced. The vision and purpose of the league tables can be summarised by a quote from the (then) Secretary of State, Wes Streeting, during a speech at the NHS Providers Conference. **In short, NHS league tables have been brought back to support transparency and accountability for performance, patient choice and incentives.**

One year on, league table rankings are a discussion point in trust board rooms, with patients and in national media. It is appropriate to consider whether the league tables are achieving the objectives that the government has set out. **This paper reviews the performance of NHS league tables for acute trusts after their first year.** Using the same format as the league tables themselves, our indicative ratings for each objective are as follows:

Figure S.1: Rating the NHS league tables for Acute Trusts: How well do they achieve their stated objectives?



Transparency and accountability

The league tables work well in generating transparency and accountability for a trust's outputs (e.g. waiting times), but they are much weaker in terms of accounting for the quality of a trust's activities (e.g. the performance of trust leadership, management and operations). Our rationale for this diverging assessment is based on statistical analysis: **Together, trust type and**

[1] Secretary of State's address to the NHS Providers Conference - GOV.UK

geography explain approximately 40% of the variability in NHS league table rankings.[2] Specialist trusts perform substantially better, while trusts with rural, coastal and more deprived catchment areas tend to perform worse. Financial performance reinforces these differences: 95% of small and medium acute trusts are in deficit, compared with 25% of specialist trusts.

Incentives to improve performance

Each trust's metrics are tailored to their circumstances, with a small set of common metrics across all trusts plus additional tailored metrics. **Our statistical analysis finds that these different metrics do not drive a trust's overall ranking in most cases**, i.e. even though trusts 'don't sit the same exam', these differences are not a material driver of trust rankings, so – on balance – we believe that tailored metrics are more likely to improve buy-in than to generate a sense of unfairness. Although the league tables do not account for structural differences between trusts (see previous point), performance management has been framed in a positive manner in a way that is likely to support improvements: *"Those at the bottom of the tables will receive more support"* and *"league tables will identify where urgent support is needed and allow high-performing areas to share best practices with others, taking the best of the NHS to the rest of the NHS"*.[3] [4] Therefore, the league tables appear to have been designed and communicated in a way that minimises any negative or perverse incentives.

Supporting patient choice

From a patient perspective, the financial metrics are less directly relevant than many of the other metrics. The significant role played by financial performance (e.g. the potential to drop into lower segments) ensures that the league tables cover the areas that are relevant to the government and NHS, but this dilutes the influence of patient-facing metrics within a trust's segment and ranking. Not all patients may understand the league tables sufficiently to look at specific relevant metrics (so may simply focus on aggregate ranking instead, which is a relatively high-level indicator), and other data tools are available to support patient choice. The variation in ranking over time for individual trusts – as well as the relative nature of the rankings – means that the league tables generate a relatively 'noisy' signal for patients.

To account for these various issues, we have rebuilt the league tables in several ways: (1) controlling for statistically significant structural factors; (2) putting trusts on a common set of metrics, i.e. 'sitting the same exam'; and (3) focusing on the metrics that are most immediately of interest to patients. These alternative views are available in our interactive dashboard.

---

[2] We found that acute trust subtype, rurality, coastality, and deprivation share explain R^2 = 43% of the variability in score. We also recorded the leave-one-out correlation and found out that these factors can predict more than a third of outcomes, at 35% on scores or 46% of rank.

[3] Secretary of State's address to the NHS Providers Conference, 2025. Available at: https://www.gov.uk/government/speeches/secretary-of-states-address-to-the-nhs-providers-conference

[4] DHSC, Landmark NHS league tables launched to drive up standards, 2025. https://www.gov.uk/government/news/landmark-nhs-league-tables-launched-to-drive-up-standards

Policy recommendations

Our policy recommendations are as follows:

- To **control for structural factors when assessing local performance** – to ensure that local leaders are evaluated fairly based on the factors that they can reasonably control.
- To **commit to tracking performance over time** – as this will help to highlight which trusts are making progress in the context of their local circumstances.
- To **create an adjusted set of metrics that are dedicated to patient choice** – as the current league table regime is better suited to supporting government oversight and accountability.
- To **undertake a review of individual metrics** – to consider whether more sophisticated metrics could add precision around oversight and accountability.
- To **consider releasing underlying data** – to support transparency, data validity and external analysis.

This paper

In this paper, **Section 1** provides an introduction regarding the theoretical rationale for league tables, along with their potential weaknesses when implemented in practice. Sections 2, 3 and 4 analyse, in turn, whether the league tables are meeting each of their main stated objectives, as follows:

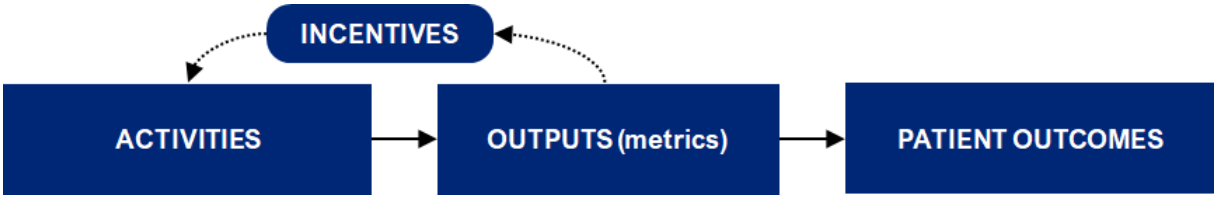
- **Section 2** assesses the role of league tables in improving transparency and accountability, drawing a distinction between outputs and activities.
- **Section 3** considers whether league tables are likely to generate positive incentives to improve performance.
- **Section 4** assesses whether league tables are likely to be effective as a tool to support patient choice.
- **Section 5** provides our recommendations in light of the analysis in this paper.

Finally, we include a series of detailed appendices which provide some statistical tables and add detail of our methodology not discussed in the main text.

1. INTRODUCTION TO THE NHS LEAGUE TABLES

A simple logic model provides the theoretical rationale for the establishment of league tables. Trust activities lead to outputs, which in turn lead to patient outcomes. Where incentives are in place, output performance assessment can also lead to positive changes in activities. So, output metrics serve as a leading indicator of patient outcomes, allow inferences to be drawn on the quality of a trust's activities, with rankings generating incentives (including financial and reputational) to improve performance.

Figure 1.1: Simplified logic model underpinning league tables



However, in practice, this logic model may not hold in its simple form. In particular, there are several issues which directly impact the ability of the NHS league tables to align with their stated purpose (see Summary section). Specifically:

- **In terms of transparency and accountability,** there are structural differences between trusts which limit our ability to infer the quality of activities from measured outputs. Therefore, output metrics help us to understand the 'end product' of one trust compared to other trusts, but this doesn't take into account the different structural circumstances faced by different trusts, such as size and geography.

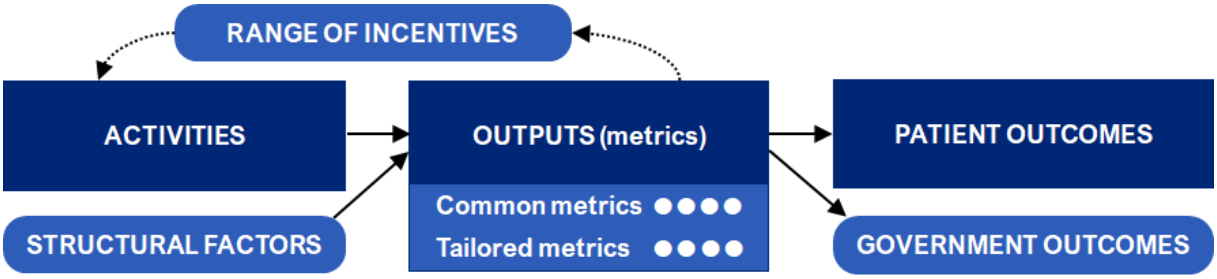
- **In terms of incentives to improve performance,** output metrics do not unequivocally drive the types of behaviours and activities that policymakers hope for or envisage. For example, output metrics can both positively encourage improved performance but also generate perverse incentives, e.g. to simply 'pass the exam'. The NHS league tables use a small set of common metrics across all trusts, plus additional tailored metrics for different trusts, and the impact of this varied set of metrics is discussed further below.

- **In terms of patient outcomes,** these are not identical to government outcomes (albeit there is considerable overlap). In the NHS league table, a financial override occurs for any trust in deficit,[5] which moves any trust in segments 1 or 2 down to segment 3. Therefore, in setting output metrics and a performance regime that addresses government outcomes, aggregate rankings will necessarily deviate from those associated with a purer focus on patient outcomes.

[5] 'In deficit' is defined as any trust with a planned deficit, financial support payments, or a negative variance versus their financial plan that exceeds any overall planned surplus. As of the Q4 2025/26 league tables, this is 71% of trusts.

The impact of these additional considerations on the simplified logic model is illustrated below.

Figure 1.2: Additional considerations relevant to NHS league tables



In the following sections we discuss the three main issues outlined above and conclude with our recommendations. A series of annexes contain additional detailed information.

2. TRANSPARENCY AND ACCOUNTABILITY

Key finding: League tables work well in generating transparency and accountability regarding a trust's outputs (e.g. waiting times). However, league tables are much weaker in terms of helping to assess the quality of a trust's activities (e.g. the performance of trust leadership, management and operations).

Approaches to performance assessment

Trust performance against its tailored set of metrics is impacted by a range of factors. Some are controllable – such as the quality of leadership and decision-making, culture and management, models of care, productivity and efficiency, and operational effectiveness. Some factors are harder to control – such as staff availability, estate configuration, and service mix. Finally, some factors are largely uncontrollable, including location (which impacts rurality for example), the size of the trust (at least in the short term) and the socio-economic characteristics of a trust's catchment area.

In this context – and with reference to the simplistic logic models above – performance assessments can be undertaken in different ways:

- First, a performance regime can focus on outputs, including through comparing trusts against each other (as the league tables do). This is a relatively patient-focused approach as trusts are evaluated purely in terms of what they deliver, e.g. waiting times. Such an approach may disadvantage certain trusts if there are uncontrollable, structural factors which impact some trusts more than others, e.g. a relatively dispersed population. Nonetheless, all trusts are assessed 'as if' equal: Outputs are the only concern and there are no special concessions for effort.

- Second, if it is understood that structural differences exist across trusts, and that these factors are materially beyond leadership control, alternative metrics can be sought – either by adjusting the outputs for these structural differences or by focusing on metrics which are more activity-based. Such an approach can provide a more robust assessment of whether the trust's leadership and staff are working effectively, and whether the trust is likely to be 'moving in the right direction'.

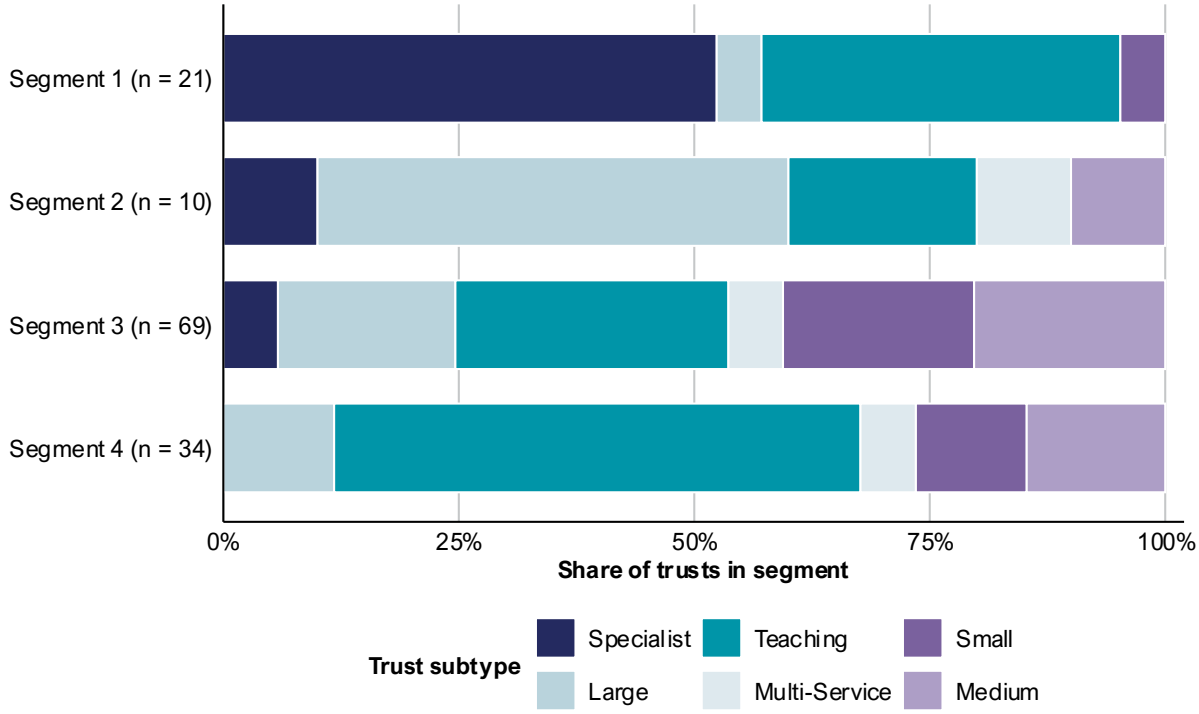
Structural factors

Considering potential structural differences, perhaps the most apparent observation from the NHS league tables is the high performance of specialist trusts, which perform well against many of the metrics. In the most recent rankings,[6] specialist trusts account for 8 of the top 10 spots, and just over half of the trusts in segment 1. In contrast, small and medium trusts generally perform poorly, with many found in segments 3 or 4, as shown below.

[6] NHS England, Acute trust league table – Q4 2025/26. June 2026. Available at: https://www.england.nhs.uk/long-read/acute-trust-league-table/

The high performance of specialist trusts is likely due to fewer emergency and/or A&E admissions (on average compared to other types of acute trust), and the potential for economies of scale arising from a specialist clinical focus.[7] The worse performance for small and medium acute trusts could be due to limited excess capacity or challenges in recruiting specialist staff. However, some caution should be applied to this finding: Previous research by Monitor (2013) *"found no clear evidence that smaller acute hospitals performed any worse clinically than larger counterparts"*.[8]

Figure 2.1: Proportion of different types of acute trusts in each segment within the league tables (Q4 2025/26)



Source: NHS England Oversight Framework acute league table (Q4 2025/26). Bars show subtype share within each oversight segment

Trusts in rural or coastal regions appear to perform worse. This could be due to longer ambulance response times and/or challenges with staff recruitment in rural areas, and *"high rates of preventable illness"* in coastal areas.[9] [10] [11] Higher local deprivation (which is associated

[7] Some specialist trusts operate specialist A&E's, such as children's only A&E, or Moorfields eye emergencies A&E.

[8] Monitor, 2013: Challenges facing small acute NHS hospitals. Available at: https://www.gov.uk/government/consultations/challenges-facing-small-acute-nhs-hospitals

[9] Palmer W, Appleby J and Spencer J, The costs of delivering health care in rural areas, Nuffield Trust, 2019. Available at: https://www.nuffieldtrust.org.uk/news-item/the-costs-of-delivering-health-care-in-rural-areas (accessed 2 September 2026).

[10] Harries M and Ushakova A, Modelling emergency response times for Out-of-Hospital Cardiac Arrest (OHCA) patients in rural areas of the North of England using routinely collected data, BMC Emergency Medicine, volume 25, 2025, article 8. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11724540/ (accessed 2 September 2026).

[11] Department of Health and Social Care, Chief Medical Officer's annual report 2021: health in coastal communities, 2021. Available at: https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2021-health-in-coastal-communities.

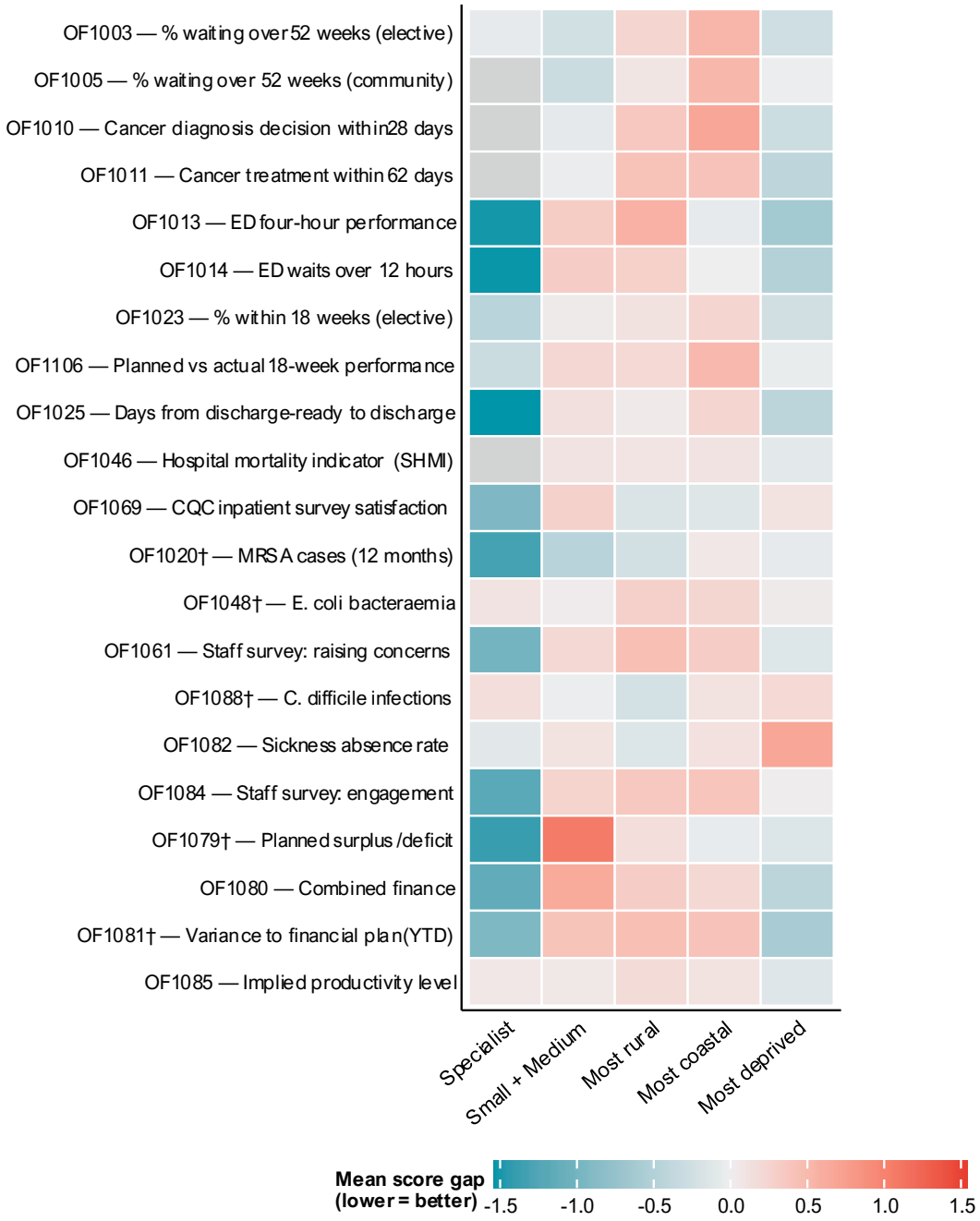
with greater morbidity and a later or more acute presentation of disease)[12] appears to be positively correlated with higher performance, but this is due to the relationship with rurality. After controlling for subtype, rurality, and coastality, a more deprived catchment area is weakly associated with worse performing trusts under our modelling.

The figure below provides a visual illustration of how the core metrics (underpinning the league tables) vary across different groups of trusts – with green indicating good performance (i.e. a lower metric score compared to the average trust) and red indicating worse performance:

[12] British Heart Foundation, Socioeconomic inequalities in heart and circulatory diseases in England, January 2025. Available at: https://www.bhf.org.uk/-/media/files/what-we-do/our-research/heart-statistics/health-inequalities/socioeconomic-inequalities-in-heart-and-circulatory-diseases-in-england.pdf.

Welch et al., The association between deprivation and hospital mortality for admissions to critical care units in England, Journal of Critical Care, volume 25, issue 3, 2010, pages 382–390. Available at: https://pubmed.ncbi.nlm.nih.gov/20074907/.

Figure 2.2: Core metrics by trust subgroup



Note: The groups Most Rural, Most Coastal, and Most Deprived are those in the highest third for each corresponding catchment. Note that these are each individual associations: The most deprived tercile is highly urban, and the relationship flips once controlling for rurality.

To test these relationships further, we undertook econometric analysis. As noted above, we found that trust subtype, rurality and catchment coastal share are correlated with trust scores, while deprivation is also correlated with worse performing trusts after adjusting for rurality. We also tested population catchment age, which was not significant after taking these other factors into account.

Together, **these structural factors predict approximately 40% of trust league table performance**.[13] That is to say, approximately 40% of the variability in trust performance is explained by structural factors that might be considered outside of the control of trust leadership and staff. Appendix A provides details of our econometric analysis that underpins this overall finding.

While there is still also substantial variation independent of these environmental factors, this finding suggests that league table positions are strongly associated with the operating environment. As such, while the league tables can tell the government which trusts are delivering the best performance in terms of metrics, on their own they would not be an accurate means with which to assess and performance-manage trust leadership, due to the materiality of these structural factors which are largely outside of their control.

Evidence of structural factors in financial performance

Structural factors – and specifically the type and location of a trust – are known issues which the NHS funding formula seeks to address through detailed analysis and adjustments. For example, through increased funding for trusts in rural areas,[14] or through additional funding to ICBs with remote trusts facing *"unavoidable smallness"*.[15]

However, despite these adjustments, the current approach does not appear sufficient to offset structural differences in financial pressures. The table below highlights the significant difference in financial performance across different types of acute trusts. Of the 39 small or medium trusts, 37 were in deficit, which may reflect a higher proportion of fixed costs. Large trusts perform better – with a 61% deficit rate. Specialist trusts have by far the strongest financial performance on average, with only 25% currently in financial deficit.

Table 2.1: Trust deficit status by type of acute trust

Acute subtype	Trusts	In deficit	Surplus	Deficit rate
Specialist	16	4	12	25%
Large	23	14	9	61%
Teaching	49	34	15	69%
Multi-Service	7	6	1	86%
Small	19	18	1	95%
Medium	20	19	1	95%

Average deficit rates

Small & medium 95%

Non-specialist 77%

All acute 71%

Note: NHS England classifies acute trust subtype using the trust type recorded in the Estates Returns Information Collection (ERIC). General acute providers are classified as small, medium or large; teaching trusts provide clinical education and training; specialist trusts predominantly undertake a single specialist function (for example oncology, orthopaedics or children's services); and multi-service trusts provide two or more functions, such as acute, community, mental health or learning-disability services.

[13] OLS R-squared is 0.426 (42.6%). Leave one out (LOO) squared Pearson correlation is 0.353 (35.3%).

[14] NHS England, Technical guide to allocation formulae and convergence for 2026/27 allocations, 2025. Available at: https://www.england.nhs.uk/wp-content/uploads/2025/12/allocations-2627-technical-guide-to-formulae-v2.pdf.

[15] NHS England, The unavoidable costs of being a small and remote hospital: technical report, 2024. Available at: https://www.england.nhs.uk/wp-content/uploads/2025/06/PRN01693_iii-the-unavoidable-costs-of-being-small-and-remote-technical-report.pdf.

The impact of adjusting for structural differences

The league tables measure trust performance as observed, including the effects of the different environments in which trusts operate. This is appropriate where the purpose is to identify where outcomes are poor and may require attention. However, it is less suitable for assessing how well a trust is performing relative to the circumstances it faces.

We have created a dashboard which adjusts for the structural factors discussed above. Adjustments are based on the coefficients derived from our econometric analysis (see Appendix A). For example, the full-sample model estimates that specialist trusts have an average score 0.5417 points lower than Large trusts, the reference category, conditional on the other covariates. We adjust for this by increasing every specialist trust's score to the average trust's characteristics. Similar adjustments are made for other structural factors based on the observed statistical associations.[16]

Having adjusted the scores for each trust, this provides an alternative 'structurally-adjusted' view of how trusts could be ranked. This is a necessarily crude measure and simply aims to get 'closer to the truth' of underlying management performance, by estimating adjusted trust scores as if all trusts faced comparable operating conditions. In effect, these adjustments control for differences in observed scores due to variations in trust type and geography.

Technical notes: The econometric modelling is linear, simplistic and is intended to provide indicative, order-of-magnitude adjustments. More complex modelling exercises – such as through interaction terms – could alter these adjustment parameters. Ranks are sensitive to small changes in underlying data and methodology. For example, Royal National Orthopaedic Hospital is 5th in the published table, shifting to 58th in the alternative 'structurally-adjusted' table before any financial override assumptions are applied, and rising back to 23rd if the financial override is included in the alternative table.

The tables below show the current Q4 2025/26 aggregate metrics and rankings, along with an adjusted view after controlling trust scores for structural factors. Specialist trusts become more distributed across the league table – with only one specialist trust (Papworth) in the 'structurally-adjusted top 10', while trusts in more rural, coastal, and deprived areas are adjusted upwards. See the dashboard on the CEPA website for further details, including estimated adjusted scores for individual trusts.

[16] The dashboard uses each trust's leave-one-out residual, re-centred on the overall mean score. This adjusts for all covariates jointly and is estimated without using the trust being adjusted.

Figure 2.3: The 'top ten': Published league table (left-hand side) versus 'structurally-adjusted' (right-hand side)

Published table 134 of 134

Rank ▲	Trust	Type	Average score	Segment
1	Royal Papworth Hospital NHS Foundation Trust	Acute - Specialist	1.22	1
2	The Christie NHS Foundation Trust	Acute - Specialist	1.26	1
3	The Royal Marsden NHS Foundation Trust	Acute - Specialist	1.27	1
4	Liverpool Heart and Chest Hospital NHS Foundation Trust	Acute - Specialist	1.45	1
5	Royal National Orthopaedic Hospital NHS Trust	Acute - Specialist	1.49	1
6	Moorfields Eye Hospital NHS Foundation Trust	Acute - Specialist	1.50	1
7	Northumbria Healthcare NHS Foundation Trust	Acute - Large	1.52	1
8	Chelsea and Westminster Hospital NHS Foundation Trust	Acute - Teaching	1.60	1
9	The Royal Orthopaedic Hospital NHS Foundation Trust	Acute - Specialist	1.61	1
10	The Clatterbridge Cancer Centre NHS Foundation Trust	Acute - Specialist	1.66	1

Structurally-adjusted table 134 of 134

Rank	Trust	Type	Adjusted average score ▲	Segment	Places moved
1	Northumbria Healthcare NHS Foundation Trust	Acute - Large	1.29	1	+6
2	Homerton Healthcare NHS Foundation Trust	Acute - Teaching	1.47	1	+31
3	Royal Papworth Hospital NHS Foundation Trust	Acute - Specialist	1.62	1	-2
4	West Hertfordshire Teaching Hospitals NHS Trust	Acute - Teaching	1.65	1	+28
5	Maidstone and Tunbridge Wells NHS Trust	Acute - Large	1.68	1	+30
6	Chelsea and Westminster Hospital NHS Foundation Trust	Acute - Teaching	1.71	1	+2
7	Imperial College Healthcare NHS Trust	Acute - Teaching	1.75	1	+4
8	Royal Berkshire NHS Foundation Trust	Acute - Large	1.77	1	+26
9	Royal Surrey NHS Foundation Trust	Acute - Medium	1.77	1	+29
10	Guy's and St Thomas' NHS Foundation Trust	Acute - Teaching	1.78	1	+2

3. GENERATING INCENTIVES TO IMPROVE PERFORMANCE

Key finding: Incentives under the current NHS league tables are likely to be reasonable. Although the league tables do not account for structural differences between trusts (see Section 2), each trust's metrics are tailored to their circumstances, and performance management has been framed in a positive manner by focusing on support for low-ranking trusts.

Incentivising improved performance across the NHS is a topic which has received much deliberation, including shifts in emphasis between collaboration and competition, and between local autonomy and national oversight.

Incentives can have a significant impact on trust behaviour – some of which may be positive and some of which may be geared towards 'passing the test'. For example, in 2009/10, following the introduction of the four-hour A&E waiting time target, aggregate provider data shows a *"rapid increase"* in the number of patients seen between 231 and 239 minutes (03:51 and 03:59 hrs), followed by a *"significant drop off"* in the number of A&E patients with a reported duration of 241 minutes or greater.[17]

While incentives are a complex area, it is possible to identify characteristics of the NHS league table regime that are likely to be supportive (and/or not supportive) of improved performance over time.

Tailored metrics

While NHS acute trusts share many common activities, they are also a heterogeneous group of organisations, ranging from specialist paediatrics to large teaching generalists, from small rural to large urban. As such, the NHS league tables do adjust the metrics applied to different types of trusts.

The chart below shows the proportion of trusts within each trust subtype for which a given individual metric is included. A dark colour indicates that a high proportion (up to 100%) of trusts within a given trust subtype (e.g. Acute – Large) have that metric included as part of that trust's aggregate score for the NHS league tables.

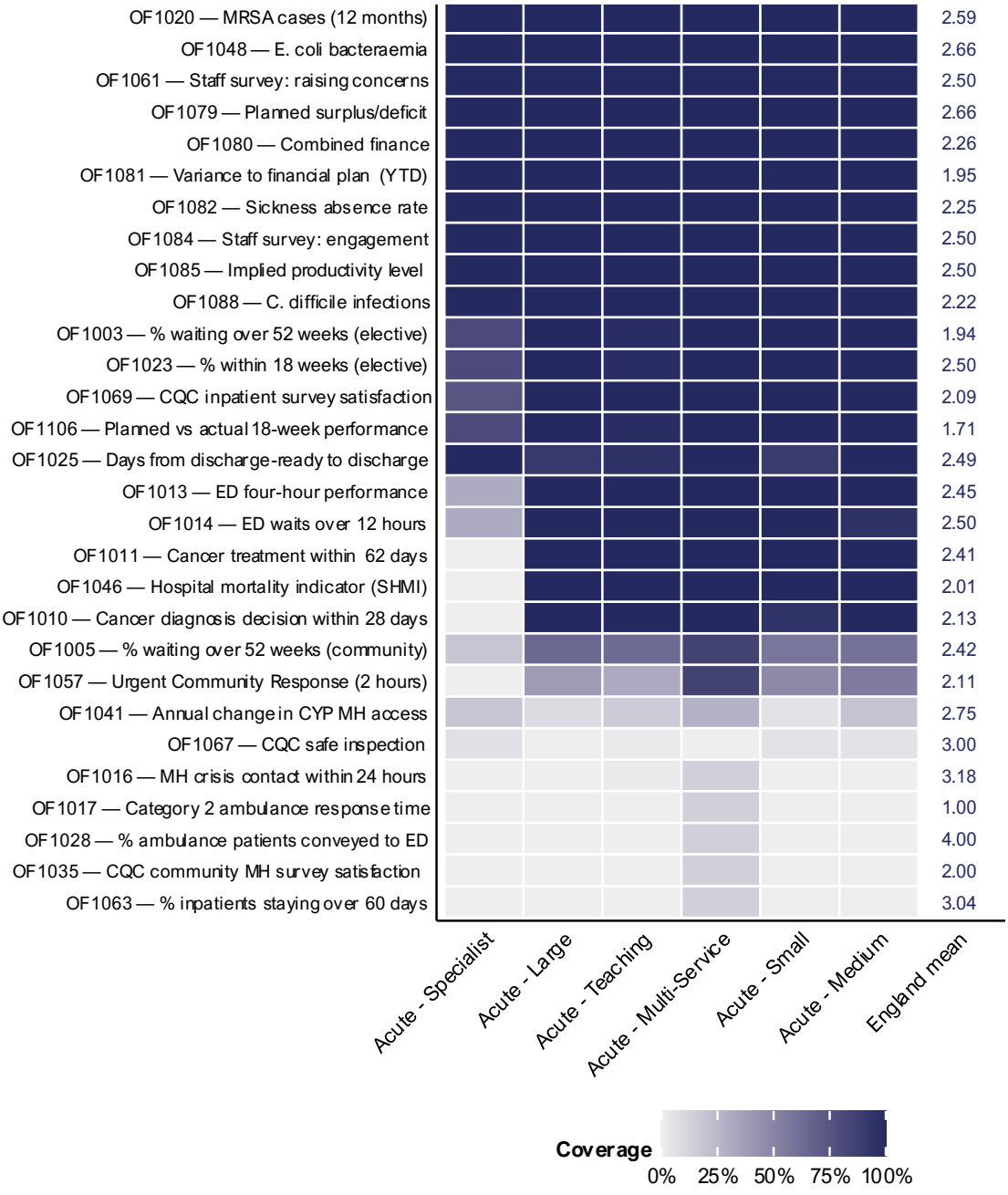
For example, it shows that:

- **Only 10 metrics are held by all trusts,** which are the three infection-related metrics, some staff survey scores, and financial performance.
- **Some metrics are only held by a single acute trust.** For example, the Isle of Wight NHS Trust is a combined ambulance and acute trust, leaving it as the only acute trust to

[17] NHS Accident and Emergency Attendances in England (Experimental Statistics) 2009-10, 2011. Available at: https://files.digital.nhs.uk/publicationimport/pub02xxx/pub02563/acci-emer-atte-eng-2009-2010-rep.pdf

also hold ambulance response time metrics. Similarly, Somerset trust is acute and mental health/community, and so it is the only trust scored on 60-day length mental health inpatients and CQC mental health survey scores.

Figure 3.1: Coverage of individual metrics across different types of acute trust, Q4 2025/26



Source: Phase 2 metric applicability. Tile colour = subtype coverage share. Numbers = England mean score among holders.

Impact of giving all trusts 'the same exam'

To consider the impact of comparing trusts using a consistent set of metrics, we retest all trusts on 'the same exam'. We identify 17 metrics that each apply to at least 90% of acute trusts. To construct a genuinely common instrument, we exclude all 16 specialist trusts and eight other

non-specialist trusts that lack at least one of these metrics. This leaves 110 trusts that are scored on the same 17-metric set. We compare their published rankings with rankings generated using only this common instrument, holding the population, aggregation and financial-override rules fixed (see Appendix B).

In practice, this adjustment makes a small difference for the majority of trusts. The ranks in the published NHS league table and the ranks for this adjusted 'same exam' scenario are highly correlated.[18] The mean change in rank is only 7.4 places, although 31% of trusts move 10 or more places, with a few trusts experiencing larger moves as a result of this truncated 'common metric' set.[19]

Moreover, we also tested whether the relatively high performance by specialist trusts in the NHS league tables is because they have metrics which are more amenable to better performance, e.g. whether they 'sit an easier exam'. Our analysis suggests that this is not the case (see Appendix B). Rescoring all 134 acute trusts on only the 10 metrics which are held by all acute trusts and holding the rest of the pipeline fixed, leaves the specialist advantage at 52 places against 54 on the published instrument.

While different trusts have a tailored set of metrics underpinning their league table ranking, our analysis suggests that differences in metrics across trusts are not the main driver of the league table rankings. As noted above, when we rank non-specialist trusts using a common set of metrics, the results remain very similar to the published table.

This is likely to have a positive impact on buy-in for the performance regime from NHS trusts and therefore to minimise perverse incentives:

- On one hand, the tailored metric set ensures that trusts are assessed appropriately for the activities that they have undertaken, which helps to mitigate criticism from the league tables attempting to implement a 'once size fits all' approach. Performance regimes that place too much weight on a small number of metrics can encourage providers to improve the measured outcome rather than performance itself. To avoid this, ideally all significant aspects of the performance should be covered by metrics to reduce the risk of perverse incentives.[20]

- On the other hand, the differences in metrics across trusts do not appear to be a material driver of the rankings. Therefore, the design of the league tables is also likely to minimise potential criticism around unfairness.

[18] The correlation is 0.946, where 1 would mean the ranks are identical.

[19] Under this scenario, several of the trusts that move the most places are those which have a relatively poorly performing metric for cancer or community waits, and this metric is removed in the 'same exam' scenario. For example, this includes Milton Keynes University Hospital (+41 places), Lewisham and Greenwich (+39), Gateshead Health (+30) and United Lincolnshire Teaching Hospital (+20).

[20] National Audit Office, The use of sanctions and rewards in the public sector, 2008. Available at: https://www.nao.org.uk/reports/the-use-of-sanctions-and-rewards-in-the-public-sector-4/.

Other incentives arising from the NHS league table approach

The NHS league tables are unlikely to perfectly incentivise the sort of behaviours that the government is seeking. Inevitably, pressure to improve against a set of reported metrics will lead to a greater focus on the metrics themselves, which may not perfectly align with the actions and activities that will optimise long-run performance.

The relative (comparative) nature of the rankings has also been questioned, as trusts only move up in the rankings if other trusts move down. First, it risks discouraging trusts who work to improve performance, only to see their league-table position remain unchanged if other trusts have done the same. This may be partially mitigated from 2026/27 as the league tables are set to take the 2025/26 scores as a baseline, in an effort to move away from a purely competitive relative rank structure. However, as many underlying metrics are still calculated using relative ranks, this issue is likely to persist across these metrics.

Second, it risks disincentivising collaborative behaviour, such as when a trust *"agrees to support other providers at the expense of its own short-term performance"*.[21] Gateshead Health FT CEO Sean Fenwick has specifically highlighted this issue, stating that the league table metrics do not recognise collaborative efforts between trusts, which could disincentivise ICB-level coordination.[22]

However, performance management has been framed in a positive manner in a way that is likely to support improvements. For example, previous Secretary of State Wes Streeting stated that the league tables *"will identify where urgent support is needed and allow high-performing areas to share best practices with others, taking the best of the NHS to the rest of the NHS"*.[23] This description does highlight accountability, but emphasises that there will be support available to trusts with low performance. The previous Secretary of State also specifically sought to alleviate concerns that the regime could return to 'naming and shaming' poor performers, by stating that "those at the bottom of the tables will receive more support".[24]

Moreover, some trusts have responded relatively positively to the re-introduction of NHS league tables. For example, improvements in league table performance at North Cumbria Integrated Care (NCIC) have been cited as a source of pride for the trust: *"They should be very proud of what they are achieving. This is not just about numbers, it is about the quality of care we provide for our patients, and that is what matters to us most"*.[25]

---

[21] NHS England, Assessing provider capability: guidance for NHS trust boards, 2025. Available at: https://www.england.nhs.uk/long-read/assessing-provider-capability-guidance-for-nhs-trust-boards/.

[22] HSJ, Recognise negative impact mutual aid can have on performance, pleads trust CEO. Available at: https://www.hsj.co.uk/policy-and-regulation/recognise-negative-impact-mutual-aid-can-have-on-performance-pleads-trust-ceo/8124180.article

[23] DHSC, Landmark NHS league tables launched to drive up standards, 2025. https://www.gov.uk/government/news/landmark-nhs-league-tables-launched-to-drive-up-standards

[24] Secretary of State's address to the NHS Providers Conference, 2025. Available at: https://www.gov.uk/government/speeches/secretary-of-states-address-to-the-nhs-providers-conference

[25] News and Star, 2026, NHS trust improves leaderboard ranking as waiting times fall. Available at: https://www.newsandstar.co.uk/news/26185418.nhs-trust-improves-leaderboard-ranking-waiting-times-fall/

Incentives issues for specific metrics

At the level of an individual metric, it is prudent to consider potential gaming risks and/or technical issues. Several issues with the current metrics are as follows:

- **Threshold effects.** For metrics such as 4-hour A&E waits, moving a patient from 4:01 to 3:59 improves metric performance, whereas seeing a patient within 5 hours (compared to 6 hours) does not. Threshold effects have been observed around the 4-hour mark, as noted above.

- **Endogenous targets disincentivise ambition in planning baselines.** Variance-to-plan metrics, such as the 'difference between planned and actual 18-week performance', disincentivises stretching and/or ambitious targets. This is partially mitigated by scoring absolute 18-week performance.

- **Metrics with low controllability.** Some metrics, such as discharge delay, are used to contextualise and evaluate acute trusts, despite social care capacity being a key driver.

- **Potential perverse incentives around staff welfare.** The metric for sickness absence disincentivises sick leave. While it is beneficial to incentivise productivity and reduce unnecessary sick leave, on the other hand, NHS data on presenteeism finds that over 50% of staff have reported working when not well enough.[26]

[26] NHS Employers, NHS Staff Survey 2025: health and wellbeing overview, 2026. Available at: https://www.nhsemployers.org/articles/nhs-staff-survey-2025-health-and-wellbeing-overview.

4. LEAGUE TABLES AS A TOOL FOR PATIENT CHOICE

Key finding: While many of the underlying metrics would be of interest to patients, the overall league table rankings are difficult for patients to interpret and impacted by financial metrics in a way that reflects government priorities more than patient preference. Given other sources of data are available to support patient choice, the NHS league table rankings do not provide the clearest possible signal to patients.

Previous Secretary of State Wes Streeting was clear that the NHS league tables are a tool to support 'choice'.[27] However, for several reasons, the league tables are not ideally suited to supporting patient choice.

Financial emphasis

"Restoring rigorous financial discipline" is identified specifically within the government's 10 Year Health Plan for England.[28] However, in the short term at least, financial discipline is unlikely to be a primary patient concern, and could even run counter to patient preference if it results in fewer staff and/or constraints on activity.

The rankings in the NHS league tables are significantly impacted by financial performance, as a trust in deficit is subject to a financial override which moves any trust in segments 1 or 2 down to segment 3. Financial metrics – and particularly the financial override – significantly reduces the ranking for several trusts in segment 1 or 2. This point has been made by Thea Stein (Nuffield Trust CEO), who argues that *"while financial balance is a reasonable indicator of efficiency and performance (with its own caveats), it's a poor proxy for the quality of care a patient will receive, and is irrelevant for patient choice"*.[29]

To address this, we have created an adjusted 'patient choice' league table in our dashboard. The dashboard strips out financial metrics (and NHS productivity metrics) to create a table focused on patient-facing metrics. These alternative 'patient facing' rankings are strongly correlated with the published NHS league tables (R-squared value of 0.93).[30] To note, this adjusted aggregate ranking would still provide only a relatively blunt signal for a patient seeking treatment; as discussed below, a patient would be best-served by considering a smaller number

[27] Secretary of State's address to the NHS Providers Conference, 2025. Available at: https://www.gov.uk/government/speeches/secretary-of-states-address-to-the-nhs-providers-conference

[28] UK Government, Fit for the Future, 10 Year Health Plan for England, 2025. Available at: https://assets.publishing.service.gov.uk/media/6888a0b1a11f859994409147/fit-for-the-future-10-year-health-plan-for-england.pdf

[29] Thea Stein, Nuffield Trust, 2026, Time to clear the tables: why NHS league rankings don't work. Available at: https://www.nuffieldtrust.org.uk/news-item/time-to-clear-the-tables-why-nhs-league-rankings-don-t-work

[30] Outside of the financial override (which has a significant impact on trust rankings), finance metrics only have a moderate weighting in the league tables, hence the high correlation between tables with and without them.

of specific metrics that relate most directly to their care needs. On the other hand, such an aggregate metric might still have value for an individual considering where to live.

Lack of clarity for patients

For several reasons, the NHS league table scores and rankings may not be fully clear for patients to interpret.

First, they present a headline measure (the aggregate league ranking for each trust) which is highly amenable to public communications, but which combines multiple different factors and metrics. Where a patient is seeking to choose between trusts for an appointment or procedure, they would be best served by looking at a smaller number of specific metrics, rather than the relatively blunt overall ranking.

In fact, this point is made on the NHS England website, which distinguishes between individual metrics and the overall (aggregate) rankings: *"[Individual Metric Ranking] league tables are primarily designed to give patients understanding of how performance varies across the country, to empower them to make more informed decisions about their care"*.[31] However, the broader public communications surrounding the NHS league tables rarely reach to this level of technical detail, and instead focus on the broader, aggregate rankings.

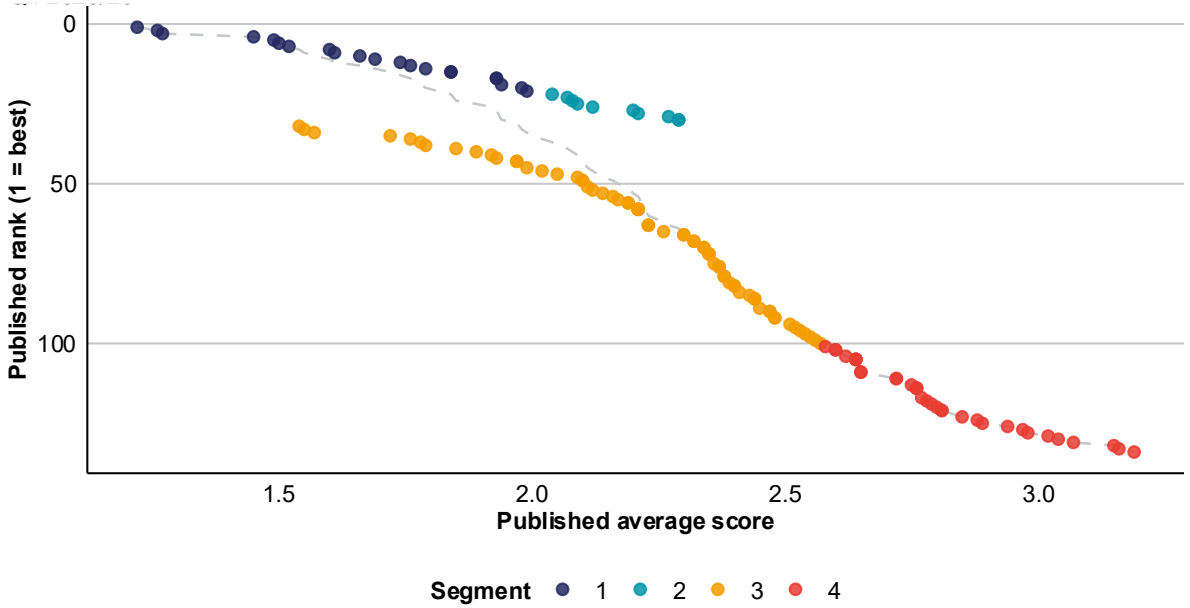
Second, as rankings are relative between trusts – and because rankings can be impacted significantly by financial performance – the meaning and significance of the rankings can be unclear to the public. For example, in December 2025 the CEO of University Hospitals Coventry and Warwickshire (Professor Andy Hardy) highlighted that the trust's relatively low league table position (132nd out of 134 trusts) could be perceived as the trust having unsafe levels of care, despite the trust receiving a 'good' rating from the Care Quality Commission.[32]

Third, the 'financial override' (for a trust in deficit) creates a discontinuity between a trust's aggregate score and their ranking. This is illustrated in the figure below, which shows the trust's average score (x-axis), their published rank (y-axis), and how the scores/ranking scatterplot would look if this financial override adjustment were not implemented (dotted line). While this negative adjustment to the ranking is understandable for an analyst or policymaker, it is less intuitive for an average member of the public.

[31] NHS England, NHS Oversight Framework: NHS trust performance league tables process and results, 2025. Available at: https://www.england.nhs.uk/long-read/nhs-oversight-framework-nhs-trust-performance-league-tables-process-and-results/.

[32] BBC, 2025, 'Table does not reflect hospital performance'. Available at: https://www.bbc.co.uk/news/articles/c4gkvr1dz01o

Figure 4.1: Published average score and ranking for each trust in published NHS league tables (Q4 2025/26)



Source: NHS England Oversight Framework acute league table (Q4 2025/26). Dashed line is rank by average score alone. Lower score is better.

High sensitivity

The NHS league table rankings are relatively sensitive to the underlying data (particularly the financial override adjustment), which generates two issues. First, trust rankings are relatively unstable from quarter to quarter, as highlighted by the Nuffield Trust's CEO Thea Stein, which means that the league tables generate a relatively 'noisy' signal.[33] Second, many of the underlying metrics are subject to substantial uncertainty, so NHS England sets out 'likely' intervals (ranges) for the scores and rankings. However, rankings are typically quoted in terms of their central estimate, which simplifies away (or masks) some of the underlying uncertainty.

Alternative option for a patient choice table

The NHS publishes a patient choice tool, 'My Planned Care', which compares wait times and CQC rankings by specialty within a given radius of any postcode. This is an excellent tool for patient choice, and its specificity by specialty allows patients to look for the best options for the specific care they seek. Adding some of the relevant league table data to this tool (beyond just the CQC rankings) could improve this tool further for patient choice.

[33] Thea Stein, Nuffield Trust, 2026, Time to clear the tables: why NHS league rankings don't work. Available at: https://www.nuffieldtrust.org.uk/news-item/time-to-clear-the-tables-why-nhs-league-rankings-don-t-work

5. POLICY RECOMMENDATIONS

We suggest the following improvements to the league tables, which could be done fairly quickly on the same underlying data:

Control for structural factors when assessing local performance

In terms of transparency and accountability, there are structural differences between trusts which limit our ability to infer the quality of activities from measured outputs. To reliably undertake an assessment of a trust's performance, it is necessary to understand and control for key structural differences – such as trust size and geography.

Commit to tracking performance over time

The league tables cannot determine if a trust improves over time. Currently, the competitive nature of the table means that one quarter of trusts will be in the bottom segment, although this is due to change in 2026/27. Creating a separate version of the table, in a non-competitive structure, would allow for effective tracking of the performance of NHS trusts over time. While the 2026/27 table does allow for growth in segment over time, some metrics still rely on a competitive structure preventing it from becoming a true improvement over time table.

Create an adjusted set of metrics that are dedicated to patient choice

The league tables are not designed as a patient choice tool, despite government and press advertisement of them as such. The NHS effectively already publishes a patient choice league table through the 'My Planned Care' tool,[34] which compares waits by specialty for all trusts within a given radius of a specific postcode, and includes first appointment wait, the 8 in 10 patient waits, and the CQC rating. Adding the league table metrics relevant to that specialty would produce a stronger patient choice tool.

Review individual metrics

Given the current NHS league tables appear to work best as a management and oversight tool – rather than to support patient choice – there could be benefit in developing metrics which are more sophisticated and/or complex. For example, A&E performance could measure a combination of average waits and long waits – to mitigate the risk of gaming around the four-hour threshold. Measuring staff sickness could also consider the risks of both absenteeism and presenteeism.

[34] NHS My Planned Care tool postcode search available at: https://www.myplannedcare.nhs.uk/find-my-hospital/postcode-search/

Consider releasing underlying data

Almost all the underlying data to reproduce the league tables from metrics is available, however not all raw metrics are reported, such as the difference between actual and planned 18-week performance and the underlying combined finance information. Given the majority of metric data is already publicly-released, providing the remainder would support transparency, data validity and external analysis.

Appendix A ECONOMETRIC ANALYSIS: STRUCTURAL FACTORS

Method Note: Deprived trust catchment is associated with being urban and is related to better trust performance, as in Figure 2.2 above. Once we control for rurality and coastality, the relationship flips and deprivation becomes weakly significant and negatively associated with trust performance. It is a modest conditional association, the weakest link of the three geography variables, with p = 0.059 under HC3 standard errors. For further reading on deprivation's effect on health outcomes, see Lusby et al (2025).

Our model is additive, so the impact of a trust with a high rural catchment share and a high coastal catchment share adds the effects together.

We model `published average score ~ catchment income deprivation + rural share + coastal share + trust subtype`

Table A.1: Statistical results for structure on trust score

Term	Coefficient	HC3 standard error	P-value	BH q-value
(Intercept)	1.7105***	0.1955	<0.0001	
Acute - Medium	0.2072**	0.1040	0.0484	0.0937
Acute - Multi-service	-0.0866	0.1369	0.5280	0.5280
Acute - Small	0.1579	0.1120	0.1613	0.1843
Acute - Specialist	-0.5417***	0.1211	<0.0001	0.0001
Acute - Teaching	0.1486	0.1026	0.1501	0.1843
Catchment income deprivation score	1.2255*	0.6420	0.0586	0.0937
Catchment rural share	0.9413***	0.2345	0.0001	0.0003
Catchment coastal share	0.5523***	0.1154	<0.0001	<0.0001

**significant at the 10% level, **significant at the 5% level, ***significant at the 1% level*

Recall that a larger score is a worse performing trust.

Trust subtype is modelled by an indicator variable with baseline Acute - Large

Catchment income deprivation score is a continuous variable that represents the proportion of people experiencing income deprivation in the catchment. For example, we find that, all other things being equal, a 10% increase in catchment area income deprivation is associated with a 0.12 higher trust score. That said, this coefficient is less precisely estimated than the others.

Rural and coastal shares are the proportions of the catchment population deemed rural or coastal. For example, a trust catchment increasing from 20% to 30% rural (a 10 percentage point increase) is associated with a 0.094 larger score.

A.1. STRUCTURAL ASSOCIATIONS WITH PUBLISHED AVERAGE SCORE

Further tables for Table A.1 above.

Table A.2: Further statistics on structural association model

Model statistic	Result
Trusts	134
In-sample R-squared, published average score	0.426
Adjusted R-squared	0.389
LOO squared Pearson correlation, average score	0.353
LOO Spearman correlation, published rank	0.461
Trusts predicted within 20 rank places, LOO	41.0%

The LOO figure of 0.353 is the squared Pearson correlation between held-out predictions and observed scores, which gives an out-of-sample measure of predictive association and reduces the optimism associated with the in-sample fit.

Catchment age was tested and does not add explanatory power after the preferred covariates: nested F = 0.64, p = 0.4248. In the age-added model its HC3 coefficient is -0.4261 (p = 0.3787).

The following variables/models were considered:

Table A.3: Table of Nested F-tests for models assessed

Adjustment block	Role	df	How tested	Joint F (p)	Sequential ΔR^2	Primary specification?
Trust subtype	Structural context / function	5	Nested F vs intercept; sequential R-squared (subtype first)	8.47 (p < 0.0001)	0.249	Yes
Catchment geography (income + rural + coastal)	Population and geographic constraint (joint block)	3	Nested F given subtype; sequential incremental R-squared after subtype	12.84 (p < 0.0001)	0.177	Yes
Income deprivation	Population context	1	Nested F dropping income from preferred; drop-one sequential R-squared	4.18 (p = 0.0430)	0.019	Yes
Rurality	Geographic constraint	1	Nested F dropping rural share from preferred	14.88 (p = 0.0002)	0.068	Yes
Coastality	Geographic constraint	1	Nested F dropping coastal share from preferred; drop-one sequential R-squared	19.08 (p < 0.0001)	0.088	Yes
Age (65+ share)	Population case mix	1	Nested F of preferred + age vs preferred; in-sample R-squared	0.64 (p = 0.4248)	0.003	No

We also considered using available beds as a measure of scale, and emergency admission share as a measure of case mix. These have significant overlap with subtype, and also are likely endogenous to trusts, and so were left out.

Note that Nested F-tests use the conventional OLS covariance structure, whereas coefficient-level inference in Table A.3 uses HC3 heteroskedasticity-robust standard errors.

Appendix B ADDITIONAL STATISTICS

B.1. SPECIALIST PERFORMANCE ON SPECIFIC METRICS

Overall, specialist trusts perform better on some of the metrics where it is harder to obtain high scores, as illustrated below. These are the hardest 6 metrics for non-specialists. We stop counting at 6, as the 7th hardest metric is 12-hour emergency department times (average score 2.56), which specialists do not sit.

Table B.1: Specialist outperformance on hard metrics

Metric	Specialist mean	Non-specialist mean	BH q
Planned surplus/deficit	1.44	2.82	<0.001
MRSA cases	1.43	2.75	<0.001
Discharge-ready delay	1.15	2.68	<0.001
E. coli bacteraemia	2.75	2.64	0.320
Staff engagement	1.45	2.64	<0.001
Staff raising concerns	1.61	2.62	<0.001

The metrics specialists are excused are on average easier for the trusts that sit them (2.34) than the ten metrics every acute trust holds (2.49). Specialists outperform on the hard metrics, as seen above. Therefore, focusing on the metrics only shared by all retains the specialist advantage.

B.2. METRIC COVERAGE

Table B.2: Metric coverage by number of trusts

Coverage across 134 acute trusts	Number of OF1 metrics
100%	10
At least 95%	15
At least 90%	17
At least 50%	21
Any acute trust	29

The 10 universal metrics are OF1079, OF1080, OF1081, OF1085, OF1020, OF1048, OF1088, OF1061, OF1082 and OF1084. Under the NHS aggregation recipe these become six effective slots: OF1079 and OF1081 are dropped, while the three infection metrics are replaced by one combined HCAI slot.

No Access to services or Effectiveness and experience metric applies to all 134 acute trusts. The 17-metric alternative therefore uses the at-least-90% set and ranks only the 110 non-specialist trusts that hold all 17 metrics. Eight other non-specialists and all 16 specialists are not ranked.

Appendix C METHODOLOGICAL NOTES

C.1. SCOPE AND SOURCES

This research used the 134 organisations in the acute-trust league table. Ambulance and mental health/community trusts are excluded due to the smaller sample size.

We used the following sources for structural covariates:
Trust catchments use the 2024 all-admissions weights in the OHID acute trust catchment workbook released in April 2026, available here: https://www.gov.uk/government/statistics/nhs-acute-hospital-trust-catchment-populations-april-2026

These weights allocate MSOA populations to trusts and are used to aggregate:

- IoD2025 Income Score (rate), using LSOA population denominators and an LSOA-to-MSOA bridge, available at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2025;
- the 2021 LSOA rural-urban classification, producing a population-weighted rural share, available at https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2021ruralurbanclassification; and
- ONS BUA22 coastal built-up areas, linked through output-area best-fit lookups, producing a population-weighted coastal share, available at https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/coastalbuiltupareasenglandandwales2022buaclassification.

C.2. ASSOCIATION AND CONTEXT ADJUSTMENT FOR 'STRUCTURALLY-ADJUSTED' TABLE

The preferred score model is:

```
published average score ~ catchment income deprivation + rural share
+ coastal share + trust subtype
```

It is estimated by OLS over all 134 acute trusts. The headline coefficient table uses HC3 heteroskedasticity-robust standard errors. Benjamini-Hochberg q-values adjust slope-term p-values within each model and standard-error family.

Predictive performance is assessed with leave-one-out prediction. For rank, predicted scores are converted to ranks and compared with published rank using Spearman correlation. For average score, the reported LOO 'R-squared' is the squared Pearson correlation of predictions and outcomes.

The context-adjusted score is the leave-one-out residual plus the mean published average score. Ranking this quantity describes how far each trust lies above or below the score expected from the fitted associations. It does not identify causal effects, estimate performance under a literal relocation or subtype change, or measure management quality.

Two context-adjusted copies are produced: one without OF5004 financial override and one that applies OF5004 after residual quartiles are formed, which are available via the toggle in the dashboard.

C.3. 'SAME-EXAM' TABLE

The core instrument contains the 17 OF1 metrics scored for at least 90% of acute trusts. Specialists are excluded because their service model means they do not hold the instrument; eight non-specialist trusts lacking at least one core metric are also excluded. The resulting ranked population is 110.

For both the baseline and alternative:

1. Start from published, non-zero OF1 scores.
2. Restrict to the same 110 trusts.
3. Apply the NHS OF5002 aggregation recipe.
4. Form national quartiles within the 110 trusts.
5. Apply OF5004 financial override.
6. Rank by adjusted segment and average.

The baseline uses each trust's full applicable OF1 instrument; the alternative uses only the 17 core metrics. Therefore, movement isolates the metric-set change while holding the population, published score source, aggregation, override and ranking rule fixed. Positive places gained means the rank number falls and the trust moves up.

C.4. 'PATIENT-FACING' TABLE

The patient-facing metrics table uses all applicable published acute OF1 scores outside Finance and productivity. It removes OF1079, OF1080, OF1081 and OF1085, and excludes OF5004 in the primary view. It retains access, patient safety, effectiveness and experience, staff engagement and sickness absence metrics. OF1086 is non-acute and therefore does not enter.

The remaining scores use the NHS aggregation recipe, national quartiles and competition ranking. The dashboard also calculates a view with the deficit override to show its separate effect.

UK

Queens House

55-56 Lincoln's Inn Fields

London WC2A 3LJ

T. **+44 (0)20 7269 0210**

E. **info@cepa.co.uk**

www.cepa.co.uk

Australia

CEPA Australia

Level 3, 333 Kent St

Sydney NSW 2000

T. **+61 2 9006 1308**

E. **info@cepa.net.au**

www.cepa.net.au