TheLabourPrefect

ANNUAL REVIEW . INAUGURAL EDITION

What the NHS workforce data says when you test it properly.

An evidence-led review of workforce pressure across NHS trusts in England, built entirely from published data

The Labour Prefect flagship annual review. Descriptive diagnostics of conditions observed, not forecasts. No trust is identified, as a matter of policy. Every threshold in this review was fixed in writing before any figure was computed.

PUBLIC DATA ONLY 210 TRUSTS 79 PRE-SPECIFIED TESTS

EXECUTIVE SUMMARY

We ran seventy-nine tests on our own doctrine. Twelve passed. Fifty failed.

Most workforce analysis reports what worked. This reports the whole battery. A result you cannot see the failures behind is advertising. The register is under NDA.

12 of 62

Scored tests that passed, from a register of 79.

9.1x

Turnover gap between the highest and lowest grade in England.

41 of 167

Trust-years in the largest single group among the held-out trusts, which is 24.6%. On development, 79 of 366, or 21.6%. Per organisation, 25.6% by modal year and 22.4% by mean score.

WHAT THE DATA SHOWS

- **The workforce leaves at the edges.** Turnover by age is a U, not a slope. By grade the gap is nine-fold.
- **Pressure is a trait, not an event.** Rank persistence 0.72.
- **Escapes hold less often than they look.** Escapers fall back more often than trusts never in.

WHAT TO DO FIRST

THE CROSS-CUTTING FINDING

No single pressure is dominant for more than 26 per cent of organisations, however we count it. A remedy for the average struggling trust is a remedy for a trust that does not exist. The board question is not "how bad is it" but "which one is ours, this year".

IF YOU READ NOTHING ELSE

Nineteen measures have been raced against the simplest guess, that next year's agency spend equals this year's. None won.

Ask any supplier claiming to predict it what they beat, and by how much.

CONTENTS

What is in this review

Sections 06 and 07 are the limitations.

PREFACE

Why this company exists

The Labour Prefect was founded on a single observation: the NHS publishes enough data to see structural workforce failure coming, and almost nobody reads it that way.

What the published record actually shows

Workforce pressure in the NHS is usually reported as a set of separate problems: sickness here, agency spend there, turnover somewhere else. Each is tracked, each has an owner, and each is managed as though it were the whole picture. What the published record shows, when the measures are put side by side, is that they are largely independent of each other and that no single one of them describes a given organisation.

That is an uncomfortable finding for a sector that buys national remedies, and it is why this company designs as well as detects.

What we will not do

We do not forecast. Every measure here describes conditions observed in a stated period, and where we have tested whether these measures could predict a future value, they could not. We publish that finding alongside the ones that worked.

We do not publish a single blended workforce score. One was tested and it failed. Three separate readings is the honest product, and it is harder to sell than one number, which is rather the point.

About the founder

The Labour Prefect was founded by Prince Opara, who holds CIPD Level 5 and an MSc in Leadership and Human Resource Management, and works frontline clinical shifts as a healthcare assistant. The company is registered in England and Wales, number 17199412.

The NHS publishes enough data to see structural workforce failure coming, and almost nobody reads it that way.

The founding observation

WHAT WE WILL NOT DO

We do not forecast, and we do not publish a single blended workforce score.

Both were tested. Neither survived, and saying so is the product.

FOREWORD

Why we publish what failed

A consultancy that only publishes its successes is asking to be trusted. A consultancy that publishes its failures is giving you a way to check.

There is an old problem in evidence. If you test fifty ideas and publish the three that worked, the three look impressive and the evidence is worthless. Anyone testing fifty ideas will find three that appear to work by chance alone. The only defence is to say in advance what you are testing, what would count as a pass, and then stand behind the whole register, failures included.

That is what this review does. In August 2026 The Labour Prefect wrote down seventy tests across fourteen constructs and nine more on volatility and concentration, seventy-nine in all, fixed the pass mark for each one in writing, and then ran them. Twelve passed. Fifty failed. The rest could not be answered on the data available and are recorded as unanswered rather than quietly dropped. This review reports every one of them in summary and names the failures it turns on. The register itself, test by test with the numbers each one produced, is a separate document and is available under NDA to anyone who wants to check it line by line.

Two of the failures were our own published claims

The register was not aimed only outward. It tested things this company had already said. Two of those did not survive, and both corrections are in this review rather than in a footnote somewhere. We would rather be the ones who found them.

The uncomfortable part of publishing a register is that the failures are more numerous than the passes and always will be. That is what testing looks like when the pass mark is set before the answer is known. The twelve that passed are worth more for the fifty that did not.

79

Tests pre-specified before computation. 62 produced a pass or a fail, 2 were underpowered, 15 could not be answered on public data.

> The pass mark is written before the test runs. Once written, it does not move. That single rule is what separates this from an opinion.
>
> From the test register

THE STANDARD WE HOLD OURSELVES TO

Every threshold in this review was fixed in writing before any figure was computed.

The register is dated and filed. Anyone can ask to see what we said we would test.

01

What the watch list corresponds to outside this company

A workforce measure is only worth anything if it lines up with something somebody else has judged independently. This is the first evidence that it does.

The fair question about any workforce measure is whether it corresponds to anything outside the spreadsheet it was built in. We tested it against two independent judgements: the Care Quality Commission's provider rating, and the NHS England Oversight Framework segment.

CQC RATING	TRUSTS	AVERAGE WARNING SIGNS CARRIED
Outstanding	21	0.37
Good	86	0.52
Requires improvement	75	0.72
Inadequate	2	2.12

Provider-level rating, latest inspection, against the mean number of warning signs carried across published years. 184 trusts.

The trusts on the watch list sit roughly **half a CQC rating band**, and **two thirds of an oversight segment**, worse than those not on it. That separation is stronger than any simpler rule we tested, and within acute trusts alone it clears our threshold on one of the three outcomes.

The rule requires two or more warning signs in **two or more published years**. The band below says why that repetition matters.

$z = 3.9$

Separation between watch-list trusts and the rest on the CQC well-led rating. Our pre-specified threshold was z = 3.3.

The persistence requirement is not a stylistic preference. It is where the signal is.

Why the rule needs two years

THE FINDING INSIDE THE FINDING

Two warning signs in a single year is one of the weakest rules we tested. The same two signs across two years is the strongest.

That is why the watch list has always required repetition, and it is now tested rather than assumed.

01B

What that evidence does not establish

The result above is strong enough to survive being stated honestly, so it is.

What this does not establish

It is not a forecast, and it is not a prediction. Every comparison above is between organisations at the same time. Nothing here says a trust on the watch list will be rated worse in future, and we do not claim it.

Three of these tests clear the threshold we set ourselves in advance, and only just. Our standing bar is p below 0.001. The watch-list comparison clears it against all three outcomes. Twelve comparisons were run in total, and correcting for that number would raise the bar to a level that only one of the three clears. We report it as strong rather than settled.

Some groups are small. Within acute trusts alone the watch list flags seven to ten organisations depending on the outcome. The direction is consistent across every cut we tried; the individual numbers are fragile.

A percentile is not exact, and we can say by how much. Resampling the peer group two thousand times, holding a trust's own value fixed, moves its percentile by about **seven points either way**. So a reading at the 95th is honestly somewhere around the 88th to the 100th. In a small peer group it is coarser still: among the ten ambulance services one place in the ranking is worth ten percentile points, so the gap between the 80th and the 90th is a single organisation moving. Treat a percentile as a band, not a position.

Two things we tried did not work, and are reported because they did not. The NHS England oversight file carries no variation within a trust across six years, so no before-and-after comparison against it is possible at all. NHS Vacancy Statistics are published at region and sector level with no organisation column, so they cannot be joined to a trust and were abandoned.

Reported as strong rather than settled.

The standard we hold ourselves to

THE TEST WE WOULD APPLY TO ANYBODY ELSE

Ask which comparisons were run, not only which ones were published.

Twelve were run here. All twelve are in the working note behind this review.

02

Four Findings

Four things the published data supports, stated with the limits that apply to each. Nothing here is a forecast, and nothing here required data from any trust.

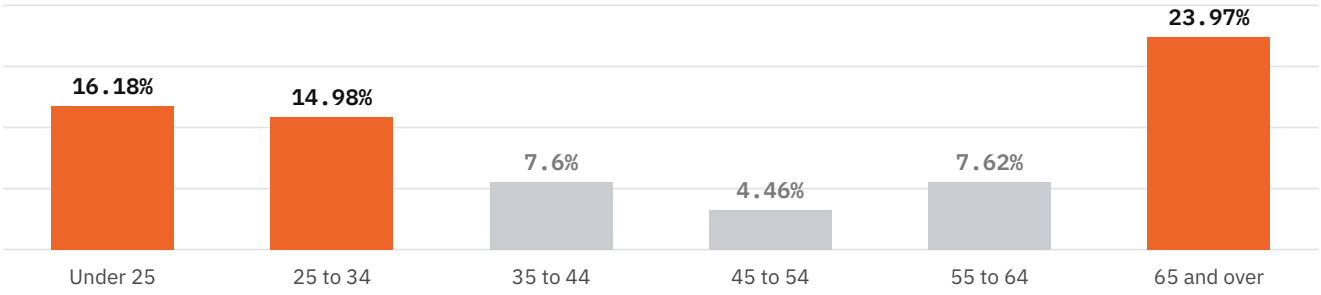
02.1

The workforce leaves at the edges, not the middle

Retention is usually discussed as one number. The national data says it is two quite different problems at opposite ends of a career, with a calm stretch in between that the single number hides.

Think of a career in the NHS as a long bridge. Most conversations about retention treat the whole bridge as equally rickety. The data says otherwise: the bridge is solid in the middle and gives way at both ends.

ANNUAL TURNOVER BY AGE BAND, ALL STAFF, ENGLAND, YEAR TO APRIL 2026.



NHS England Digital. Rates are the publisher's full time equivalent rates. The counts here are headcount and divide to a higher figure: 12,349 of 49,020 is 25.2 per cent against 24.0 on FTE. In people, 12,349 of 49,020 aged 65 and over against 16,637 of 351,316 aged 45 to 54. VALIDATED, tests 10.1 and 10.2. England-wide only: no published file carries age or tenure at trust level, so no trust can be told where its own cliff sits.

The highest band runs at **5.4 times** the lowest. Split by grade the gap is wider still, **9.1 times**, with Foundation Year 2 doctors at 39.7 per cent on an FTE basis, where 3,021 people left a workforce of 7,578, against consultants at 4.4 per cent, where 3,180 left a workforce of 65,318.

Why this matters for a workforce plan. An initiative aimed at everybody is aimed largely at the middle of the bridge, where most of the workforce sits and leaving is least common. Aimed at the two ends, it is aimed where the leaving happens. That is an argument about targeting from the turnover shape; we hold no data on what retention programmes cost.

The limit, stated plainly. These are England-wide figures. We can describe the shape of the cliff nationally. We cannot yet tell an individual trust where its own cliff sits, because no published file carries age or tenure by organisation. Any supplier who tells you otherwise from public data should be asked which file.

WHAT A BOARD SHOULD TAKE FROM THIS

Ask for your leaver rate split by age band and by grade, not as one number.

You already hold this. Your own establishment data carries tenure and age at exit, which the published national files do not.

02.2

Pressure is a trait, not an event

Trusts under structural pressure this year are, overwhelmingly, the trusts that were under it last year. That is bad news for the trust and useful news for anyone trying to decide where to look.

Workforce pressure is usually reported as weather: this quarter is bad, last quarter was better. Tested across the panel, it behaves much more like climate. Where a trust ranks against its peers one year is a strong guide to where it will rank the next.

Three separate measures were tested for this and all three held. Whether a trust is flagged persists at a rank correlation of **0.72** across 429 observations. The gap between staff signalling they will leave and staff actually leaving persists at 0.57. Most strikingly, the tendency of a trust *not to notice* its own distress persists at **0.80**, which is the most stable thing in the entire register.

The uncomfortable implication

If pressure were weather, a bad year would justify waiting. It is not weather. A trust sitting high on these measures three years running is not having a run of bad luck; it is describing a structural condition that has stabilised. Waiting is a decision, and on this evidence it is usually the wrong one.

What this does not say. It does not say a trust cannot change. It says change has not been happening on its own. That is a different and more actionable statement.

0.72

Rank persistence of the fragility flag, n 429. VALIDATED.

0.80

Persistence of the blindness gap, n 358. The most stable measure in the register.

> Trusts under pressure stay under pressure. That is the finding, and it is not a comfortable one.
>
> Test 6.2

WHAT A BOARD SHOULD TAKE FROM THIS

If a pressure has been elevated for three years, stop describing it as a difficult period.

Across this panel, pressures elevated for several years seldom resolved without a material change in conditions.

02.3

Escapes hold less often than they look

Most trusts that get out of the top quartile of agency dependence stay out. They fall back more often than trusts that were never there, and that difference is the finding.

This is the one finding in the review that supports a claim the sector already makes, and it earned its place by beating a properly chosen comparison rather than by looking dramatic.

Take every trust that was in the heaviest quartile of agency dependence and then got out. Two years later, **77.8 per cent** are still out. That figure is the average of three separate two-year windows rather than one pooled count, and here is what it is the average of: **8 of 8** organisations held in the first window, then **4 of 6** and **4 of 6**. Twenty escaping organisations in total, each appearing once. That is a small base, and it is the reason to treat this as strong rather than settled. Against it: how many trusts that were *never* in the top quartile are still out two years later. The answer is **91.1 per cent**: 79 of 87, then 83 of 88 and 74 of 84, across the same three windows. Both figures are equally weighted averages of their three windows, and both are in Appendix B.

The gap is **13.3 percentage points**, and that gap is the finding. Getting out is real. Staying out is measurably harder for an organisation that has been there than for one that has not.

The analogy, because the number alone understates it

It is the difference between someone who has quit smoking and someone who never smoked. Both are non-smokers today. Only one of them is at elevated risk of being a smoker again in two years. Treating the two as the same person is how a recovery plan gets stood down eighteen months too early.

The limit. This is an association observed over a defined window. It does not identify what causes the return, and it does not predict which individual trust will fall back.

13.3pp

Gap between escapers and trusts never in the top quartile, two years on. VALIDATED, test 1.2.

20

Escaping organisations behind the 77.8%, across three two-year windows. Each appears once.

Getting out is real. Staying out is harder for an organisation that has been there.

Test 1.2

WHAT A BOARD SHOULD TAKE FROM THIS

Do not stand down an agency reduction programme at the point the numbers first look good.

On this evidence a first exit should not be treated as proof of sustained improvement.

02.4

There is no typical struggling trust

This is the finding with the widest consequences, and it is the reason a sector-wide workforce remedy so often disappoints the organisations it was designed for.

Across the panel, each trust in each year was sorted by which of the measured pressures sat highest for it. If workforce failure had one dominant shape, one pressure would account for most of them. It does not. **The largest single group is 21.6 per cent, which is 79 trust-years out of 366.** The other five pressures take 72, 71, 66, 47 and 31 between them.

The unit is a trust-year, not a trust. Those 366 observations come from 125 organisations, most appearing in three years. Earlier editions of this review did not make that distinction, so the count was rerun with one row per organisation: the largest group is 25.6 per cent by an organisation's most frequent dominant pressure, 22.4 per cent by its average scores. **Roughly three in four organisations have a different dominant pressure from the largest group, on either reading.** Design one intervention for the average struggling trust and you have designed it for a trust that does not exist.

The awkward part. Only 65 of those 125 organisations, 52 per cent, carry the same dominant pressure in every year they appear. For the rest it changes. So the question this review tells a board to ask is a question about a year, not a permanent trait. Ask it again next year. These per-organisation figures are descriptive checks on the unit and were not part of the pre-specified test.

WHAT A BOARD SHOULD TAKE FROM THIS

The dominant pressure here is a question about this year.

Half of these organisations carry a different one from year to year, so it is a question to ask again rather than a label to fix in place.

02.4A

Why this is not obvious from a dashboard

Each measure looks like a sector-wide problem on its own. What a dashboard cannot show is that they are not the same trusts.

A dashboard shows each measure separately, and each one looks like a sector-wide problem because in aggregate it is. What a dashboard cannot show is that the trust at the top of one measure is usually not the trust at the top of the next. Across the panel these measures are largely independent of one another. 2 of the 3 pairs sit within 0.10 of zero. The strongest relationship, Normalised Fragility against Departure Lag, is +0.22 across 203 trusts, which is weak but not absent. Because this is Spearman's rank correlation, its square is not read here as the proportion of variation explained in the original measures.

One qualification, and it matters. The measures are near-independent at a point in time, but across a one-year gap they do relate to each other, mean absolute lagged correlation 0.205. Notably the links run in the opposite direction to the sequence this company previously published. That correction is in section 03.

What follows for a board. The useful question is not how your trust compares on a national average. It is which of these pressures is yours, because on this evidence the answer differs from your neighbour more often than it matches.

WHAT A BOARD SHOULD TAKE FROM THIS

Ask which pressure is dominant here, before asking how bad it is.

A ranking tells you where you sit. It does not tell you what you are dealing with, and the second question is the one that changes what you should do.

02.4B

The one finding here that was checked outside its own data

Everything here comes from looking at data, and looking enough times means some of what turns up is real and some is an accident of which organisations you happened to look at. Nothing inside them tells you which is which.

41 of 167

Trust-years in the largest single group among the held-out organisations, from 57 of them, which is 24.6%. On the panel the measure was built from it is 79 of 366, from 125 organisations, or 21.6%.

0.943

Agreement between the two groups on the full ordering of the six pressures.

A result you cannot see the failures behind is advertising. A result never checked outside its own data is a hypothesis.

On method

WHY THIS ONE AND NOT ANOTHER

A held-out group can be spent once, so it was spent on the finding the company rests on.

If no single pressure dominates, a sector-wide remedy is aimed at a trust that does not exist. That is the argument this document makes, so it is the one that had to be tested somewhere it could fail.

So before any of this analysis began, the 519 organisations in the ESR panel were split at random by fixed seed, by organisation and not by year: 364 to development, 155 set aside, none in both. Of the 519, 238 are trusts; the split put 167 of them in development and 71 in the held-out group. After complete-case filtering, that leaves 366 trust-year observations from 125 development organisations and 167 from 57 held-out ones. Every other figure here was produced without the 155. Two nearby numbers are not the 519: the 210 quoted elsewhere is the filtered trust panel behind the percentile findings, and the 519 trust-financial-year pairs behind the agency finding are observations from 178 trusts.

It held

The largest single group among the held-out organisations is **24.6 per cent**, 41 trust-years out of 167, contributed by 57 organisations. The same pressure ranks first in both groups, and the whole ordering of the six agrees at a rank correlation of 0.943.

Why 57 and not 155. Of the 155 organisations held out, 84 are not trusts, mostly Integrated Care Boards, and are outside this comparison. That leaves the 71 held-out trusts, of which 57 carry all six measures in at least one year. Earlier editions of this review printed 71 as the count behind the result and explained the shortfall by survey coverage. Both were wrong, and this is the correction.

Neither figure should be quoted as though it were the other. Together they say one thing: **no single pressure dominates the sample.** The largest group takes 21.6 per cent of development trust-years and 24.6 per cent of held-out trust-years. With one row per organisation it takes 25.6 per cent by modal year and 22.4 per cent by mean score.

Three limits, stated with the result. This replicates a shape rather than testing significance, and was described that way in writing before the run. The pre-specified unit is the trust-year, so the per-organisation figures in section 02.4 are a later check on that unit and not part of what was confirmed here. And a held-out group is spent when it is opened: the frozen analysis remains rerunnable, but these organisations are no longer unseen and will not be used to validate a revised rule.

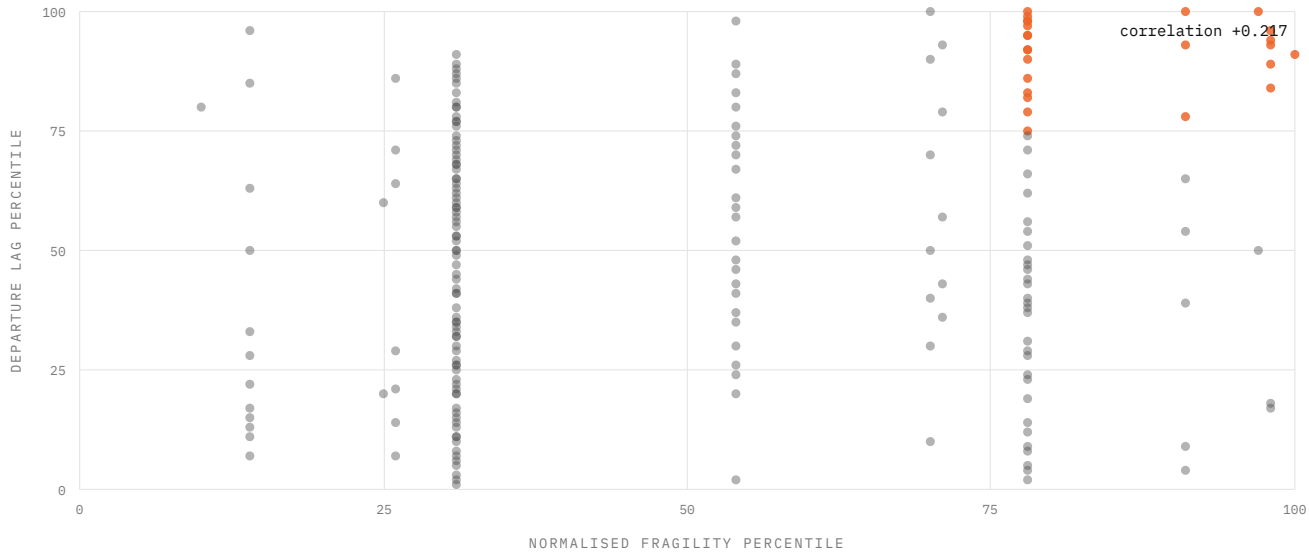
02.5

What near-independence actually looks like

The claim on the previous page is that these measures do not move together. A reader is entitled to see the cloud rather than take a coefficient on trust.

Each dot is one trust. If the two measures moved together the dots would form a diagonal band. They do not. The dots in ember are the trusts elevated on both, and there are few of them.

NORMALISED FRAGILITY AGAINST DEPARTURE LAG, 203 TRUSTS, PERCENTILE WITHIN PEER GROUP.



Spearman rank correlation +0.217. This is the STRONGEST of the three pairs; the other two sit closer to zero. Computed at build time from the current panel, not quoted.

Being precise about this, because we got it wrong once. Earlier versions of this review described the measures as having a mean pairwise correlation of 0.004. Re-checked against the current panel that figure does not reproduce. The accurate statement is the one above: two of the three pairs sit near zero, and the strongest is +0.22. Weak, not absent. Because this is a rank correlation, we do not square it into a proportion of variation explained in the original measures.

WHY THIS MATTERS MORE THAN THE COEFFICIENT

A trust at the top of one measure is usually not at the top of the next.

That is what makes a single blended workforce score misleading, and it is why this company does not publish one.

03

What We Tested And Rejected

The spine of this review. Every construct that failed its own pre-specified threshold, including two claims this company had already published.

03.1

The assumptions that did not survive

Each of these was written down as a testable claim with a pass mark, before any figure was computed. Each one failed that mark.

That workforce failure runs in a sequence. This company previously described six pressures as stages a trust moves through in order. Four of the five consecutive links were testable. **All four failed**, and the loop does not close. What survives is six named mechanisms, separately defensible, that a trust can carry in any combination. What does not survive is the order. We now present a map, not a sequence.

That higher agency spend signals future deterioration. Nineteen predictors have been raced against the simplest possible guess, that next year's agency spend equals this year's. None won. Equivalence testing then bounded **eighteen of the nineteen**: none of those adds more than **1.9 percentage points**, eight of the eighteen add less than one, and the median is 1.2. The nineteenth, tested on 84 observations, was inconclusive rather than ruled out.

That workforce pressures move together. They do not, at least not at the same moment, and the figures are on the previous page. A trust can look settled on one measure and be badly exposed on another, which is exactly why a single blended workforce score is not published by this company and will not be.

That the pattern gets worse in a straight line. Short-staffing does accelerate exhaustion rather than tracking it evenly, and that held. The same shape did not hold for turnover, and no threshold effect was found. One of three passed.

0

Constructs upgraded to validated this year.

50

Tests that failed their own pre-specified threshold.

We report all of this because the discipline is the product. A firm that publishes only what worked is asking to be trusted rather than checked.

Why this section exists

THE RULE THAT MAKES THIS SECTION MEANINGFUL

The pass mark for every test was fixed in writing before the test was run.

Without that rule, a register of failures proves nothing: anyone can move the line after seeing where the ball landed.

03.2

Three of our own published claims, corrected

A register that only tests other people's assumptions is not a register. These three were ours, and all three need correcting in public.

One: an association we described as movement was not movement

This company published a finding that trusts with a wider departure gap were more likely to enter the regulator's highest oversight category the following year. Re-checking the underlying file for this review found that **no trust in it changes category in any year**. Across five year-on-year transitions covering more than 170 trusts each, the number that moved was zero.

The most likely explanation is ordinary: one snapshot per trust appears to have been copied across every year column when the file was assembled. The association we found may well be real, but it cannot be what we said it was. It compares fragile trusts with badly-rated trusts, which is a statement that they tend to be the same trusts. It is not a statement that one leads to the other.

The claim is withdrawn pending a properly year-varying file from the published quarterly source. Three further tests that depended on the same file are recorded as unanswered rather than failed.

Two: a null result that rested on a broken column

We previously reported that trusts cutting agency spend did not simply substitute bank staff instead, and reported it as a null: we looked, and found nothing. Re-checking found that **143 organisations report zero bank spend in every single year**, which is not credible as real spending. It is absence recorded as zero.

Treating those as missing rather than as zero leaves too few organisations to answer the question at all. So the honest description is not "we tested it and found nothing". It is **"we cannot yet test it"**. That is a smaller claim, and it is the accurate one.

WHAT BOTH OF THESE HAVE IN COMMON

Neither was a mistake of judgement. Both were defects in a source file that looked fine until somebody opened it.

That is the argument for re-checking the file rather than the summary, and it is why the third correction overleaf exists.

03.2

The third correction, and how it was caught

This one was found while doing something else, which is usually how.

Three: the turnover rates in section 02.1 were half their true size

The retention cliff is the newest finding in this review, and until this edition every rate in it was **exactly half** what it should have been. The under-25 rate was stated as 8.09 per cent; it is 16.18. Foundation Year 2 doctors were stated at 19.86 per cent; they are 39.71. Consultants were stated at 2.18; they are 4.37.

The cause was arithmetic, not judgement. The publisher supplies an opening and a closing workforce for each period, and our code added the two together where it should have averaged them, then summed a national total alongside its own sub-totals. **The ratios were never affected.** The finding itself, that turnover is 5.4 times higher at one end of a career than in the middle and 9.1 times higher at one grade than at another, stands exactly as it was tested, because both tests were scored on the ratio. Only the rates printed beside it were wrong, and they are corrected throughout this edition.

How it was caught, which is the part worth having

It surfaced while converting every percentage in this review into a headcount. The published 8.09 per cent for the under-25s implies an all-staff turnover rate of about 5 per cent. The same feed puts it at **10.11 per cent**, and this company has been saying "about ten per cent of staff leave" in public for months. The contradiction was sitting inside our own document and nobody saw it for twenty-six days.

A percentage on its own hides that. A percentage carrying the number of people it describes does not, which is why every share of a population in this edition now travels with its headcount, and why the rule is enforced by the software that builds these documents rather than by anyone remembering it.

Why we are publishing all three. Together they make our evidence base look thinner than it did last month. We would still rather be the ones who found them, and a reader is entitled to know which of a supplier's claims have survived re-checking and which have not.

THE QUESTION WORTH ASKING ANY WORKFORCE SUPPLIER

Which of your published findings have you re-tested, and which of them did not survive?

Every analytics supplier has some. The distinction that matters is whether they found them first, and whether they told you.

04

What was observed after agency share fell

The common response to an elevated agency position is to cut it. Across the panel, a lower agency share and higher absence appeared together, with the absence measured in the six months after the agency year ended.

Across the panel, each £100 of **implied annual agency-spend reduction at a fixed pay bill** was associated with this much pay-bill-equivalent absence over the **six months following the financial year end**:

SCOPE	PER £100 OF ANNUAL AGENCY REDUCTION	OBSERVATIONS	CONFIDENCE
Whole panel	**£2.70**	519	Clears the threshold, p = 0.0003
Below this line the splits are **exploratory** robustness checks of the same model, not four separate studies. None of them clears the threshold.			
Acute	£2.77	311	p = 0.005
Mental Health	£1.68	136	p = 0.088
Community	£5.01	42	p = 0.081
Ambulance	£7.17	30	**p = 0.542, not reliable**

Estimated across 519 trust-financial-year pairs from 178 trusts. The illustration overleaf uses the whole-panel figure, which is the only one that clears the pre-specified threshold.

The illustration, and why these pounds are not a cost, overleaf.

145k

People who left the 210 panel organisations in the year to May 2026. 31k were nurses, about £381m to replace on the NHS Shared Business Services estimate of £12,000 a head, which is theirs and not ours. Not applied to the other 113k.

£2.70

Per £100 of implied annual agency reduction

519

Trust-financial-year pairs

0.0003

p value on the whole panel

04B

Why that figure is not a cost, and not a saving

The pounds put a value on time lost to absence. They are not money leaving the organisation, and they cannot be netted against a saving.

At scale. On a £100m annual pay bill, a two point reduction in agency share is roughly £2m. Against that, the panel puts about **£54,000** of pay-bill-equivalent absence **over the following six months**.

One assumption, stated. The model measures agency as a share of total staff cost, so a fall in that share is not by itself cash saved: it can happen with a growing pay bill. Converting it to pounds holds the pay bill fixed, which is what "implied" means in the table opposite.

Why that is not a cost, and not a saving. Staff are generally paid while off sick, so their salary already sits in the pay bill. The figure values the lost time at pay-bill rates; it is not cash leaving the organisation, and how much cash is involved depends on how the absence is covered. It is also not a net saving, because a pay-bill-equivalent valuation cannot be subtracted from a cash saving.

This is not a forecast, and it does not establish cause. It describes what showed up across the panel in the six-month outcome window used for the headline estimate. A twelve-month window was also fitted. It points the same way, does not meet our threshold of p below 0.001, and is reported as a sensitivity in the Technical Assurance Pack rather than here. The analysis does not show whether redesigning the work first changes the relationship. That would need testing inside the organisation.

£54k

Pay-bill-equivalent absence over six months on a £100m annual pay bill, against a two percentage point reduction in agency share.

A valuation of lost time is not a cash saving, and cannot be netted against one.

Why this is not a saving

THE QUESTION A FINANCE DIRECTOR SHOULD ASK

Has anybody looked at absence alongside the reduction in agency share?

Agency expenditure is a cash measure. Absence is a loss-of-time measure. They are observed over different windows, so agency expenditure alone is not the whole picture.

05

One thing any board can test against its own data in 30 days

Everything here comes from published data. The fastest way to find out whether it describes your organisation is to check it against data only you hold.

Three tests, one per measure, that need nothing from us. If the pattern is not there, the measure is not describing you, and we would want to know.

Normalised Fragility

Take the three warning signs and ask your own systems how long each has been elevated. If Normalised Fragility describes this organisation, you will find that at least one has been above its peer median for three consecutive years and that no paper went to a committee about it in the last two. The test is not whether the figure is high. It is whether anybody escalated it.

Departure Lag

Exploratory, and it tests something different from the measure above. Compare last year's leavers with a matched group who stayed: did sickness, bank or overtime uptake move in the six months before leaving? Fix what counts as a change first. A difference supports the withdrawal explanation; finding none does not invalidate a rate comparison made at organisation level.

Pressure Displacement

Find the last time this organisation reduced agency share by at least two percentage points of the service pay bill. Compare average monthly sickness absence during the following six months with the preceding twelve. Examine the same services across both periods. If the pattern appears, assess how the absence was covered and how it was valued. Exploratory; it does not establish cause or cash cost.

We would rather be corrected than believed.

Why this page exists

THE COMMITMENT BEHIND THIS PAGE

If your own data contradicts this review, we will say so in the next edition.

Two published claims have already been withdrawn that way.

06

If we are wrong, this is how

We have told you what failed. It is only fair to say what could still be wrong with what passed.

Twelve findings survived a pre-specified threshold. That is a real bar and it is not an infallible one. Four things could still overturn any of them, and naming them is cheaper than being caught by them.

One. A published source could be wrong, and one already was

This review corrects two of our own claims that rested on defective source files: a regulator file with no year-to-year movement, and a spending column recorded as zero for 143 organisations. Both looked fine until they were checked directly. Others may be waiting, and the only defence is to keep checking the file rather than the summary.

Two. Association is not mechanism

Every finding here describes a pattern in observed data. None demonstrates that one thing causes another, and where the register tested a causal claim it generally failed. A reader who converts these findings into causes is going further than the evidence.

Three. The panel is English NHS trusts over four years

It is not primary care, not social care, not the devolved nations, and not a longer window. A finding that holds here may not hold elsewhere, and we make no claim that it does.

Four. Twelve passes out of sixty-two scored is still twelve tests

The threshold was set at p below 0.001 precisely because running seventy-nine tests at a conventional level would produce about four passes by luck. That guards against coincidence; it does not eliminate it. Each passed finding produced evidence inconsistent with its stated null at the pre-specified threshold. That is not the same as a probability that the finding is true, and we do not claim one. We will re-run the register next year.

Why print this at all. Because a report that lists only what it found is asking to be believed, and a report that lists what would change its mind is offering to be checked. We would rather be checked.

THE QUESTION WORTH ASKING OF ANY WORKFORCE ANALYSIS, INCLUDING THIS ONE

What would have to be true for this conclusion to be wrong?

If the answer is nothing, it is not a finding. It is a belief with a chart.

07

Method, limits, and what a board should do next

Short by design. A method section that dominates a report is usually compensating for something.

What this review is built from

Published NHS sources only: the NHS Staff Survey, provider annual accounts, and NHS England workforce statistics derived from the Electronic Staff Record. A panel of 210 trusts across four data years, 2021 to 2024, with monthly workforce data from April 2018 where the analysis required it. **No data was requested from any trust**, no data-sharing agreement was needed, and no individual is identifiable from any figure here.

The three rules

Thresholds were fixed in writing before computation. The significance threshold was set at p below 0.001 rather than the conventional 0.05, because running seventy-nine tests at the conventional level would be expected to produce about four passes by luck alone. And every test is reported, including the ones that could not be answered.

What this review is not

It is not a forecast. Each measure here is validated as a description of conditions observed in the period stated. Nothing in it predicts what any trust's figures will be next year, and where we tested whether they could, they could not.

Three questions for a board in 2027

One. Which of these pressures is dominant in this organisation? Not how we compare on average, but which one is ours.

Two. How long has it been elevated? If the answer is three years or more, it is a structural condition and not a difficult period.

Three. What did we stop doing, and did anything actually improve? The strongest available move is usually removing an action that is making a mechanism worse, and it costs nothing.

210
Trusts in the panel

79
Pre-specified tests

0
Data requests to any trust

APPENDIX A

Data sources and how each measure is built

Included following structured adversarial review of the draft. A company that says it shows its working has to show its working.

The three measures, and exactly what goes into each

MEASURE	INPUT	PUBLISHED SOURCE	YEARS
Normalised Fragility	Count of distress markers: sickness absence, agency share of the pay bill, and staff intention to leave	NHS sickness absence returns; provider annual accounts; NHS Staff Survey	2021 to 2024
Departure Lag	Gap between the share of staff intending to leave and the share who actually left	NHS Staff Survey, against the realised leaver rate	2021 to 2023
Pressure Displacement	Agency share of the total staff bill, against sickness absence	Provider annual accounts; NHS sickness absence returns	2023

How a value becomes a percentile, and a percentile a band

Each measure is ranked **within the organisation's own peer group**, never against the whole of England. The percentile is the share of peers scoring at or below. At or above the 90th bands as severe, at or above the 75th as elevated, below that as within range. **These thresholds were fixed in writing before any figure was computed** and have not moved since.

Continued overleaf.

THE RULE THAT MAKES THE REST OF THIS DOCUMENT MEANINGFUL

Every threshold was fixed in writing before any figure was computed.

Without that rule a register of results proves nothing, because anyone can move the line after seeing where the ball landed.

APPENDIX A

Data sources and measures, continued

The peer groups

Acute, Mental Health, Community and Ambulance, taken from the provider classification in the published source rather than assigned by us. No figure in this document ranks an ambulance service against an acute trust.

Missing data

Where a measure is not published for an organisation it is left blank. It is never imputed, never set to zero and never ranked, because an absent value is not a low one.

The watch-list rule, which is separate from the percentiles

An organisation joins the watch list when it carried two or more distress markers in two or more published years. This is an absolute rule, not a peer comparison, and it requires persistence: one bad year does not qualify.

What we cannot see

The published record carries no tenure, no age profile, no pay band, no vacancy detail and no staff-group split at organisation level. Six planned tests were blocked outright for exactly this reason, and fifteen in all could not be scored on public data. Both are recorded as unanswered rather than as failures.

About the imagery

The cover artwork and any interior scenes are **generated illustrations, not photographs of real people or premises**, and they carry no evidential weight. No individual is depicted and no organisation's building is shown.

One image is a real photograph: the portrait of the founder on the preface page. Every other image is generated. We separate the two because every claim here is sourced, and the pictures should not be the one thing a reader takes on trust.

210
Trusts in the panel

4
Published years

20 of 22
Registered findings printed in Appendix B

APPENDIX B

The evidence register, one of four

The 20 findings this review rests on, each with the published source behind it. This covers the headline figures and not every number in the document.

Verification. Every headline figure here is covered by a dated internal register recording its public source, reporting period, sample, calculation and expected result. It held 22 checks when last run, on **15 September 2026**, and all 22 matched their expected outputs. The methods are recomputation from the published source, comparison against the stored output of the analysis that produced the figure, independent recalculation and comparison against a frozen held-out record. A sanitised methodology and evidence schedule is available to an appointed independent reviewer under controlled access.

FINDING	PUBLISHED SOURCE AND PERIOD	SAMPLE AND METHOD	WHAT IT ESTABLISHES, AND WHAT IT DOES NOT
Watch-list separation on the CQC Well-led rating, z 3.91, against a bar of 3.3 fixed in advance	CQC provider ratings; NHS England Oversight Framework segments Ratings at latest inspection, April 2026 extract; warning signs 2021 to 2024	184 trusts Difference in mean outcome between flagged and unflagged trusts, as an approximate z	Clears the bar of 3.3 fixed in advance, on all three outcomes. Twelve comparisons were run, so strong rather than settled. Cross-sectional: it says nothing about a future rating.
Mean warning signs by CQC provider rating across 184 trusts: 0.37, 0.52, 0.72 and 2.12	CQC provider ratings Latest inspection, April 2026 extract	184 trusts Mean number of warning signs carried, by rating band	Descriptive. The Inadequate band is two trusts, so its mean is fragile.
Turnover by age band, England, year to April 2026: 16.18, 14.98, 7.60, 4.46, 7.62 and 23.97 per cent, each with its leavers	NHS England Digital, NHS Workforce Statistics, turnover by grade and age band. HCHS staff in trusts and core organisations in England, excluding primary care Rolling twelve months to April 2026	1,541,372 staff across six age bands Leavers over the mean of the opening and closing workforce. Rates are the publisher's full-time-equivalent rates; counts are headcount	Validated, tests 10.1 and 10.2. England-wide only: no published file carries age at organisation level, so no trust can be told where its own cliff sits.

WHAT A READER IS ENTITLED TO

The source, the period, the sample and the limit, for every finding.

Where a sample is counted in trust-years rather than trusts, this register says so, because the two are not the same number.

APPENDIX B

The evidence register, two of four

FINDING	PUBLISHED SOURCE AND PERIOD	SAMPLE AND METHOD	WHAT IT ESTABLISHES, AND WHAT IT DOES NOT
5.4 times by age and 9.1 times by grade, Foundation Year 2 at 39.71 per cent against consultants at 4.37	As the row above Rolling twelve months to April 2026	Six age bands; every published grade Highest band rate divided by lowest, computed from unrounded rates	Validated. The ratio was the pre-specified gate and was unaffected by the rate defect corrected in August 2026.
All-staff turnover of 10.11 per cent, the figure that exposed the halved rates	As the row above Rolling twelve months to April 2026	1,541,372 staff Leavers over mean workforce, headcount	Descriptive. Published because it is the figure that exposed a halving defect in our own earlier rates.
Rank persistence: the fragility flag 0.72 on 429, the blindness gap 0.80 on 358, the departure lag gap 0.57 on 283	NHS Staff Survey; NHS England workforce statistics; provider annual accounts 2021 to 2024	429 consecutive-year pairs from 146 organisations; 358 pairs from 125; 283 pairs from 144 Spearman rank correlation between a measure and the same measure a year later	Validated against a threshold of 0.50 fixed in advance. Descriptive persistence, not a forecast.
Escapes from the top agency quartile: 77.8 per cent hold against a base rate of 91.1, a 13.3 point gap	Provider annual accounts 2020 to 2024, in three overlapping two-year windows	20 escaping organisations; 259 comparison organisation-windows Share still outside the heaviest agency quartile two years on, averaged equally across the three windows	Validated against a gap of 10 points fixed in advance. The escaping base is small, which is the main reason to treat it as strong rather than settled.
The counts behind 77.8 per cent: 8 of 8, then 4 of 6 and 4 of 6, twenty escaping organisations in total	Provider annual accounts 2020 to 2024	8 of 8, then 4 of 6 and 4 of 6 The per-window numerators and denominators behind the averaged rate	Published so a reader can judge the base rather than take the average on trust.
No pressure is dominant for more than 21.6 per cent of trust-years: 79 of 366, then 72, 71, 66, 47 and 31	NHS Staff Survey; NHS England workforce statistics; provider annual accounts 2021 to 2024	366 trust-years from 125 organisations Each measure standardised, then the highest per trust-year tabulated	Validated. The unit is the trust-year. The same count per organisation is two rows below.

APPENDIX B

The evidence register, three of four

FINDING	PUBLISHED SOURCE AND PERIOD	SAMPLE AND METHOD	WHAT IT ESTABLISHES, AND WHAT IT DOES NOT
Confirmed on held-out organisations at 24.6 per cent, 41 trust-years of 167, full ordering agreeing at 0.943	As the row above 2021 to 2024	167 trust-years from 57 organisations The same procedure applied to organisations held out of the analysis until it was finished	Confirmed against gates fixed in writing beforehand. Replicates a shape rather than testing significance. The frozen analysis remains rerunnable; these organisations are no longer unseen and will not be reused to validate a revised rule.
The same test with one row per organisation: 25.6 per cent largest by modal year, 22.4 per cent by mean score, on 125 organisations; 52.0 per cent keep the same dominant pressure in every year	As the row above 2021 to 2024	125 organisations The trust-year result re-counted one row per organisation, under two aggregations: most frequent dominant pressure, and highest mean standardised score	NOT pre-specified. A later check on the unit of analysis. The finding holds at 25.6 and 22.4 per cent, and the same check shows only 52 per cent of organisations keep the same dominant pressure in every year they appear.
The split behind that: 519 organisations, 364 to development and 155 held out, of which 238 are trusts, 167 and 71	NHS England workforce statistics, Electronic Staff Record Fixed before any analysis began	519 organisations, 364 development and 155 held out, of which 238 are trusts Random split by organisation, stratified by benchmark group, fixed seed	No organisation appears on both sides, and that is checked rather than assumed.
The strongest pair, Normalised Fragility against Departure Lag, at +0.217 across 203 trusts; mean absolute 0.114	NHS Staff Survey; NHS England workforce statistics; provider annual accounts 2021 to 2024	203 trusts on the strongest pair; 168 and 164 on the other two Spearman rank correlation between the three diagnostics, on percentiles within peer group	Descriptive. Weak but not absent. A rank correlation is not squared into a share of variation. Replaces a 0.004 figure that did not reproduce and is withdrawn.
The same measures across a one-year gap: mean absolute 0.205	As the row above 2021 to 2024	393 trust-years Mean absolute rank correlation between each pair across a one-year gap	Validated. The strongest links run backwards relative to the sequence this company previously published and has withdrawn.
Pressure Displacement: beta minus 0.054 at p 0.0003, on 519 trust-financial-year pairs from 178 trusts	Provider annual accounts; NHS sickness absence returns Financial years 2021 to 2023	519 trust-financial-year pairs from 178 trusts Regression of the six-month sickness response on the change in agency share, controlling for prior sickness and provider type	Validated against four gates fixed in advance. Association, not cause. On a panel extended to 2024 it fails its first gate, so this window remains the authoritative one.

APPENDIX B

The evidence register, four of four

FINDING	PUBLISHED SOURCE AND PERIOD	SAMPLE AND METHOD	WHAT IT ESTABLISHES, AND WHAT IT DOES NOT
Pay-bill-equivalent absence over six months per 100 pounds less agency spend: 2.70 whole panel, then 2.77, 1.68, 5.01 and 7.17	As the row above Financial years 2021 to 2023	519 trust-financial-year pairs; sector splits of 311, 136, 42 and 30 The coefficient expressed as pay-bill-equivalent pounds per 100 pounds less agency spend, over the six months it was measured across	Only the whole-panel figure clears the threshold. The sector rows are robustness splits of one model, not four studies, and the ambulance and community rows rest on very small samples.
About 54,000 pounds of pay-bill-equivalent absence over six months against a two point agency cut on a 100 million pound annual pay bill	As the row above Financial years 2021 to 2023	An illustration on a stated 100 million pound pay bill Arithmetic on the whole-panel figure: two per cent of 100 million is 2 million; 2 million divided by 100, multiplied by 2.70215, is 54,043	An illustration, not a measured outcome for any organisation. It does not establish cash cost, and it does not show when within the six-month outcome window the difference arose.
145,430 leavers across the panel, 31,755 of them nurses and health visitors, about 381 million pounds to replace	NHS England Digital, NHS Workforce Statistics turnover benchmarking; NHS Shared Business Services for the unit cost Rolling twelve months to May 2026	145,430 leavers across 204 of the 210 panel organisations Leaver headcount, with the 12,000 pound unit cost applied only to nurses and health visitors	The unit cost is NHS Shared Business Services' and not ours, and is never applied to a population it was not measured on.
Nineteen predictors raced against next year's agency spend and none beat the lazy guess; eighteen are bounded at 1.9 points or less	TLP equivalence analysis on NHS England workforce statistics and provider annual accounts Registered 2 August 2026; panel 2021 to 2024	19 predictors, 84 to 505 observations each Partial correlation against a one-year-lagged baseline, then a two one-sided test at 90 per cent confidence	Eighteen of the nineteen are bounded: none of those adds more than 1.9 percentage points to what last year's figure already explains, eight of the eighteen add less than one, and the median bound is 1.2. The nineteenth, an unknown classification category on 84 observations, is reported as inconclusive rather than ruled out.
The register, counted as cells: 76 emitted, of which 12 pass, 50 fail, 2 underpowered and 12 descriptive, after the 5 August re-run	The TLP pre-specified test battery, gates lodged in writing before computation Registered August 2026	76 emitted cells, from 79 pre-specified tests Counted from the machine-written result files, not from a prose summary of them	Cells and tests are different units. The four files emit 76 cells; the battery was 79 pre-specified tests. The 15 tests recorded as unscored are the 12 descriptive cells plus 3 tests that emitted no cell, and 12 passed plus 50 failed plus 2 underpowered plus 15 unscored is 79. The 15 could not be answered because the data required is not published at organisation level.

APPENDIX C

Terms of use, and the limits of this document

A public document should say what it is, what it is not, and what happens when it turns out to be wrong.

What this document is

An independent analysis produced from publicly available NHS data. It is published by The Labour Prefect Ltd for information. It is not commissioned by, endorsed by, or produced in cooperation with any NHS organisation named or described in it.

Permitted use

You may read, print and circulate this document within your own organisation, and quote from it with attribution to The Labour Prefect Ltd. You may not resell it, present it as your own work, or reproduce it in a commercial product without written permission.

What it is not

It is **not advice**. It is not legal, clinical, financial or safe-staffing advice, and it is not a substitute for professional judgement. Any decision taken after reading it remains the reader's own, under the reader's own governance.

It is **not a forecast**. Nothing in it predicts a future value for any organisation.

It is **not an assessment of any organisation's quality, safety or leadership**. The measures describe workforce structure only.

Accuracy, and what we do when we get it wrong

Every figure is computed from published sources at the time of writing and is believed accurate. Published sources are sometimes wrong, and so are we. Where we find an error in our own work we correct it in public, in the next edition, and say what it was.

The Labour Prefect Ltd accepts no liability for decisions taken in reliance on this document. If you believe a figure here is wrong, tell us and we will check it.

Data protection

No personal data was used to produce this document. Every source operates at organisation level and no individual is identifiable from any figure in it.

APPENDIX D

Acknowledgements and contributors

Nothing here was produced from nothing.

Data

This document would not exist without the organisations that publish NHS workforce data openly and consistently. **NHS England** and **NHS England Digital** for the NHS Staff Survey, the workforce statistics derived from the Electronic Staff Record, and the sickness absence returns. **Individual NHS providers** for annual accounts filed to a common standard, which is what makes a panel possible at all. All are used under the Open Government Licence.

Method

The pre-specification discipline used throughout, fixing every threshold in writing before computation, is taken from clinical trial practice rather than invented here. We claim no credit for the idea, only for applying it to workforce analysis, where it is uncommon.

Challenge

Drafts of this document underwent structured adversarial review using four commercial large language models: ChatGPT, Gemini, DeepSeek and Kimi, each directed at the methodology rather than the prose. That was methodological challenge testing, not academic peer review, and the distinction is stated plainly because the phrase "external review" is often used to imply the second. **No independent human audit has yet been completed**, and commissioning one is an open task rather than a finished one.

The exercise was still worth doing. It found real faults: claims that outran the evidence, a structure that repeated itself, too little method and too little visual evidence. It is not a substitute for a statistician or a workforce scientist reading the underlying data, and it should not be read as one.

Contributors

This edition was researched, analysed, written and produced by **Prince Opara**. Where future editions carry contributions from others, they will be named here individually.

Analytical tooling was built with AI assistance under human direction. Every figure was computed by the engine from published data. Some are recomputed as this document is built and the rest are transcribed from the machine-written result of the analysis that produced them, which the evidence register records figure by figure. Language models supported the analytical tooling, adversarial review of the draft, and editorial revision, all under human direction. Prince Opara reviewed the evidence, approved the final wording, and is responsible for every claim here.

APPENDIX E

Every technical term in this document, in plain English

Every technical term used in this document, in plain English. These are the same definitions published on our website, read from the same file, so the two cannot say different things.

MEDIAN

The middle value once every organisation is placed in order. With an even number of organisations it is normally the average of the two central values. Half score above it and half below.

PERCENTILE

Your place in the queue, as a number out of 100. At the 88th percentile, 88 of every 100 comparable trusts sit at or below you. Higher is worse on every measure we publish.

PERCENTAGE POINTS

The plain difference between two percentages. If 22 of every 100 staff say they intend to leave and 10 actually go, the gap is 12 percentage points. Calling that 12% would mean something different.

RHO

How closely two rankings move together, from -1 to +1. Zero means no monotonic rank association was detected, which is not the same as no relationship of any kind. Near +1 means a trust high on one is usually high on the other.

BETA

The size of a relationship, in the units being measured. Our beta of minus 0.054 means that for every 1 percentage point cut in agency staffing share, permanent-staff sickness rose by about 0.054 percentage points. The minus sign means one went up as the other came down.

P VALUE

How surprising the observed result would be under the stated null model. A p value of 0.0003 means a result at least this extreme would occur about 3 times in 10,000 if that model held. It is not the probability that the finding is a fluke. Our bar was fixed at p below 0.001 before running anything.

PERSISTENCE

Whether a trust keeps its position year to year. High persistence means the ranking is sticky, so this is a repeatedly observed relative position rather than a bad year. It does not establish what produces it.

CORRELATION

Whether two things tend to move together. It does not mean one causes the other. Our three measures are close to uncorrelated, which is why we report them separately instead of averaging them into a single score.

QUARTILE

One of four equal groups. The heaviest agency quartile is the quarter of trusts spending most on agency staff.

DECILE

One of ten equal groups. The most exposed decile is the tenth of trusts sitting furthest from stability on a measure.

TRUST-YEAR

One trust in one year, counted as a single observation. 519 trust-years means 519 combinations of a trust and a financial year, drawn from 178 trusts. It is not 519 different organisations.

EFFECT SIZE

How big a difference is, separately from whether it is real. A result can be statistically solid and still too small to act on. One of our measures failed on exactly that: the direction was right, the size was not.

BASE RATE

What happens normally, used as the comparison. 91.1% of trusts that were never heavy agency users are still out two years later. That is the base rate the 77.8% escape figure has to be judged against, and the gap between them is the finding.

UNDERPOWERED

There was not enough data to give the test a fair chance of finding an effect, so it counts as neither a pass nor a fail. Two of our seventy-nine were underpowered, and we report them separately rather than counting them as failures.

CONFIDENCE LABEL

Our own three-step grading. Reliable cleared the bar set in advance. Indicative points the same way on weaker evidence. Estimate only means the sample is too small to lean on. Figures also carry a plain note such as "holds"; not a fourth grade.

PRE-SPECIFIED

The pass mark was written down and dated before the specified analyses were run, so nobody can move the goalposts once the answer is known. We say pre-specified rather than pre-registered: the protocol is internal and dated, not lodged with an independent registry.

The Labour Prefect

What we do

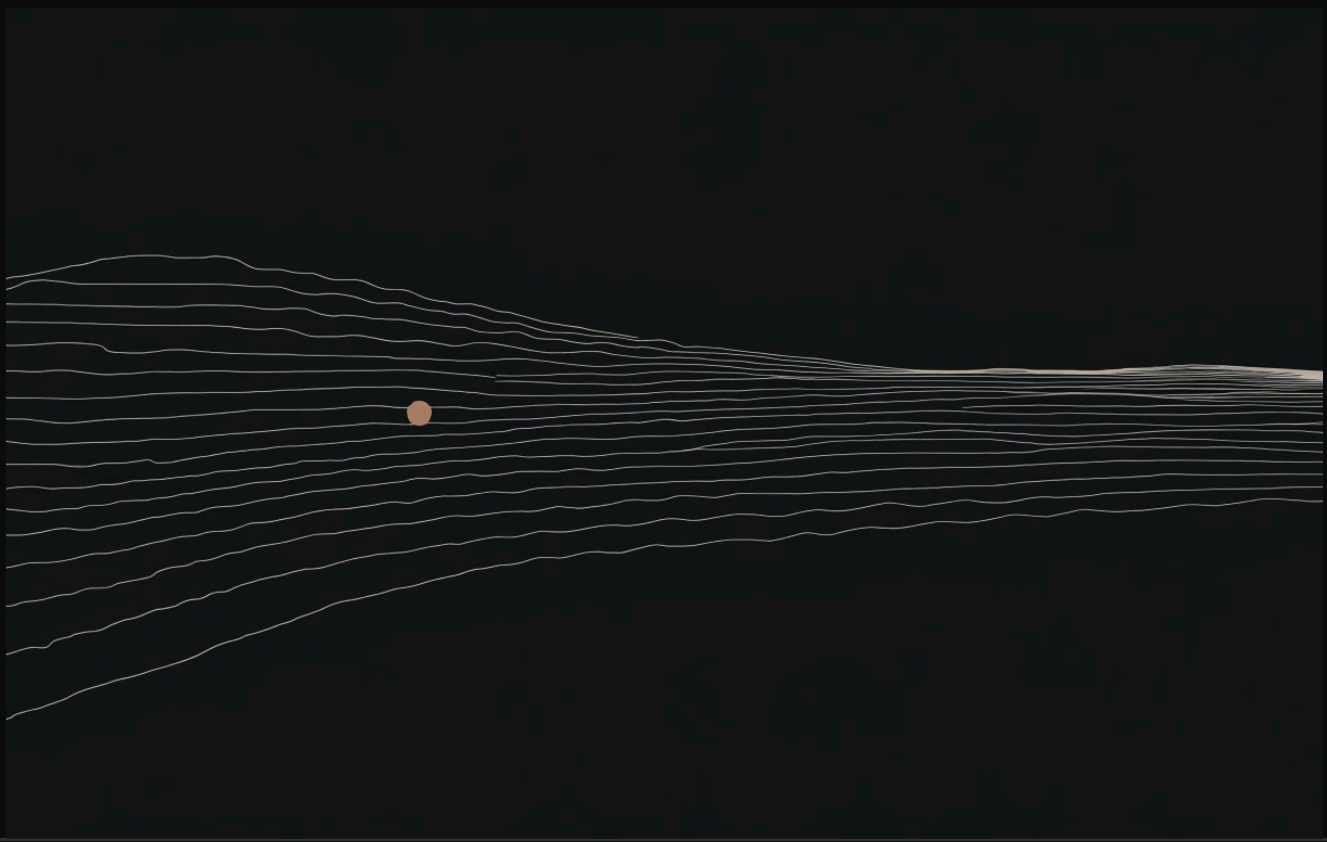
The Labour Prefect is a workforce intelligence architecture company. We detect structural staffing failure in published NHS data before it surfaces in standard dashboards, and we design the structural interventions that resolve it.

What this review is

An annual, public, evidence-led review of workforce pressure across NHS trusts in England. Built entirely from published sources. No trust is identified, as a matter of policy, and no organisation paid for its inclusion.

How to go further

A Workforce Structural Review applies this method to one organisation and answers which of these pressures is yours. Diagonal then tracks whether it moves.



The Labour Prefect Ltd . Registered in England and Wales, company 17199412 . Registered office: 71-75 Shelton Street, Covent Garden, London, WC2H 9JQ