White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/242796/

Version: Published Version

Article:
Dhali, A., Mandal, S., Aithal, G. et al. (2026) Training experience across UK medicine: deanery differences versus site variation—General Medical Council National Training Survey 2025 analysis. Frontiers in Medicine, 13. 1833862. ISSN: 2296-858X

https://doi.org/10.3389/fmed.2026.1833862

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

frontiers in Medicine

TYPE Original Research
PUBLISHED 24 June 2026
DOI 10.3389/fmed.2026.1833862

Check for updates

OPEN ACCESS

EDITED BY
Hollis Lai,
University of Alberta, Canada

REVIEWED BY
Alena Petrakova,
Palacký University Olomouc, Czechia
Dominic Atraszkiewicz,
Leeds Teaching Hospitals NHS Trust,
United Kingdom

*CORRESPONDENCE
Saikat Mandal
✉ Saikat.Mandal@nottingham.ac.uk
Arkadeep Dhali
✉ arkadipdhali@gmail.com

†These authors share senior authorship

RECEIVED 18 March 2026
REVISED 19 May 2026
ACCEPTED 28 May 2026
PUBLISHED 24 June 2026

CITATION
Dhali A, Mandal S, Aithal G and Sanders DS (2026) Training experience across UK medicine: deanery differences versus site variation—General Medical Council National Training Survey 2025 analysis.
Front. Med. 13:1833862.
doi: 10.3389/fmed.2026.1833862

COPYRIGHT
© 2026 Dhali, Mandal, Aithal and Sanders. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Training experience across UK medicine: deanery differences versus site variation—General Medical Council National Training Survey 2025 analysis

Arkadeep Dhali[1,2]*, Saikat Mandal[3,4]*, Guruprasad Aithal[3,4†] and David S. Sanders[1,2†]

[1]Academic Unit of Gastroenterology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom, [2]School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom, [3]School of Medicine, University of Nottingham, Nottingham, United Kingdom, [4]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, United Kingdom

Background: The General Medical Council's (GMC) National Training Survey (NTS) is widely used to benchmark postgraduate training experience in the UK. However, for actionable interpretation, it is necessary to distinguish deanery-level patterns from variation that is more locally specific to individual sites and specialities.

Methods: We analysed the 2025 GMC NTS post-speciality-by-site extract for medicine, benchmarked against all medical speciality posts. Each row represented an aggregated site × speciality × indicator cell. The primary metric was a gap score, defined as the site mean minus the national mean. Analyses followed a prespecified descriptive framework: deanery × indicator profiling, indicator-level comparisons with effect sizes and false discovery rate correction, intraclass correlation coefficients (ICCs), site-level correlation and volatility analyses, a suppression audit, outcome-category comparison, and principal component analysis (PCA)/cluster analysis.

Results: The NTS data included 26,190 scored observations across 16 deaneries, 31 medical specialities, and 18 indicators. The mean deanery gap ranged from +3.23 in the North East to −3.13 in the East of England. The North East scored above the national average in 17 of 18 indicators, with the largest positive gaps in regional teaching (+7.34), rota design (+4.55), and reporting systems (+4.44). The East of England showed the largest negative gap in study leave (−10.17). After correcting for multiple testing, 17 of 18 indicators differed across deaneries; local teaching was the only indicator that did not differ. The largest deanery-level effect sizes were noted for regional teaching, clinical supervision out of hours, teamwork, and reporting systems. However, ICCs were modest overall, with the highest ICC observed for regional teaching (0.0973), indicating that a majority of the variance remained outside deanery-level clustering. Adequate experience correlated strongly with overall satisfaction (Spearman ρ = 0.848). Study leave showed the greatest site-to-site variation [standard deviation (SD) = 17.82]. Suppression for cells with fewer than three responses affected 21,060 of 47,442 rows (44.4%). Official NTS outcome categories also differed across deaneries (χ^2 = 597.9, df = 60, $p < 0.001$).

Conclusion: Training experience in UK medicine varies across deaneries and between individual sites. The findings support a dual improvement

approach—deanery-wide attention to domains with stronger regional patterning, such as regional teaching and out-of-hours clinical supervision, alongside site-level review of highly variable domains such as study leave and rota design. PCA clustering and the resilience index should be interpreted as hypothesis-generating. Interpretation is also limited by the use of aggregated survey data and substantial suppression due to small cell counts.

KEYWORDS

clinical learning environment, deanery, National Training Survey, postgraduate medical education, quality, training, variation

Introduction

Postgraduate medical education in the UK is delivered through rotational clinical posts across NHS trusts/boards and training sites. A deanery or equivalent postgraduate training body provides regional organisation, quality management, and oversight, whereas the individual site is the hospital or clinical location where day-to-day clinical work, supervision, teaching, and workplace learning are delivered. This structure indicates that poor or strong training experiences may arise from regional policy and governance, local site implementation, speciality-specific pressures, or a combination of these levels.

Postgraduate medical education is delivered predominantly through day-to-day clinical work, where the clinical learning environment shapes competence acquisition, professional identity, motivation, and retention (1). Supervision remains a core mechanism through which trainees gain safe autonomy, receive feedback, and translate clinical exposure into learning (2). Practical barriers, including time pressure, workload, competing service demands, and variable organisational support, can erode supervision quality and consistency (3).

The quality of trainees' working and learning conditions has implications for patient safety and workforce sustainability. Systematic reviews report associations between clinician wellbeing, burnout, and safety-related outcomes, including perceived safety incidents and poorer quality-of-care indicators (4, 5). Burnout is common during postgraduate training across specialities and countries (6, 7). Physician burnout has also been associated with reduced career engagement, which supports the need for organisational responses rather than a narrow focus on individual resilience (8). Training quality and working conditions, including job autonomy and structured support, have been associated with junior doctors' intentions to leave clinical practice (9). UK graduate follow-up surveys similarly describe variation in early-career doctors' perceived support and training quality (10).

More recent UK evidence reinforces why National Training Survey indicators should be interpreted as system-level signals. Trainee perceptions of supervision, rota design, teamwork, and satisfaction have been correlated with excess mortality at the NHS organisational level (11), and a single-centre rota redesign report suggests that the rota structure can be modified to improve equity and staffing processes (12). Survey work among British doctors has also highlighted rota issues, training pressures, leave, and work–life balance as barriers to home-life satisfaction (13).

The GMC NTS is an established UK-wide survey used to monitor and benchmark postgraduate training experience (14, 15). The 2025 GMC report describes the survey as a large national dataset completed by approximately 71,000 doctors and highlights continuing concerns about workload, burnout, and escalation culture (16). The GMC education data tool provides trainee indicator outputs at multiple aggregation levels, including post-speciality by site, deanery/National Health Service England (NHSE) local office, trust/board, and site (17). Previous studies have examined speciality-specific variation and quality improvement applications using NTS-derived data; however, it is necessary to distinguish deanery-level patterns from site-driven variation when interpreting medical speciality results (18–21).

The research question for this study is as follows: Within UK medical speciality posts, which aspects of 2025 NTS variation are more consistent with deanery-level patterns, and which appear predominantly site- or speciality-driven? The specific objectives were to (1) describe deanery-level and speciality-level gap profiles; (2) quantify indicator-specific deanery differences using effect sizes; (3) estimate the proportion of variance attributable to deanery membership using ICCs; (4) identify indicators that co-vary across sites and those with high site-to-site volatility; (5) assess the extent and distribution of small-number suppression; and (6) present PCA/clustering and a resilience index.

Methods

Data source and study design

We used the 2025 GMC NTS post-speciality-by-site extract for medicine, benchmarked against all medical speciality posts (17). Rows in the extract corresponded to aggregated site × speciality × indicator combinations rather than individual trainee-level records. The Defence Postgraduate Medical Deanery and unspecified deanery were excluded before the analytic deanery comparisons. For each scored observation, a gap score was computed as follows: gap = site mean – national mean (all medical speciality posts benchmark). Scores are reported on a 0–100 scale. Since the available analytic unit was an aggregated survey cell, the study was designed as an ecological, descriptive analysis and not as an individual-level causal or multilevel modelling study.

Units of analysis and included groups

A total of 16 deaneries were included in the analysis. A total of 18 NTS indicators were analysed: adequate experience, clinical supervision, clinical supervision out of hours, educational governance, educational supervision, facilities, feedback, handover, induction, local teaching, overall satisfaction, regional teaching, reporting systems,

rota design, study leave, supportive environment, teamwork, and workload. Speciality-level profiling, PCA, clustering, within-deanery inequality, and speciality heterogeneity analyses were restricted to 16 major medical specialities with at least 50 scored observations.

Analytical framework and rationale

First, deanery-level summaries and deanery × indicator heatmaps described where mean gaps were above or below the national benchmark. Second, indicator-level comparisons tested whether gap distributions differed across deaneries. The Kruskal–Wallis tests were used as the primary non-parametric comparison because normality of aggregated gap distributions could not be assumed; a one-way analysis of variance (ANOVA) was reported as a complementary parametric sensitivity analysis (22–24). *p*-values across the 18 indicators were adjusted using the Benjamini–Hochberg false discovery rate procedure (25).

Third, effect sizes were prioritised to avoid overinterpreting statistical significance in this large dataset. For η^2 and ω^2, values of approximately 0.01, 0.06, and 0.14 were interpreted as small, medium, and large effects, respectively; for Mann–Whitney *r*, values of approximately 0.10, 0.30, and 0.50 were interpreted as small, medium, and large effects (26). These thresholds were used as interpretative guides. Fourth, ICCs were computed for each indicator to estimate the proportion of site-level variance attributable to deanery membership, with an n_0 adjustment for unequal deanery sizes (27). ICCs were interpreted as variance-partitioning descriptors. Fifth, for the 247 sites with complete data across all 18 indicators, site-level means were calculated for each indicator, and Spearman's ρ and Pearson's *r* correlations were estimated for all 153 indicator pairs. This analysis was intended to identify domains that tended to move together across sites. Site-to-site volatility was summarised using SD and interquartile range (IQR) of site-level gap scores for each indicator.

Sixth, suppression was examined because GMC outputs do not report scores for cells with fewer than three responses. Suppression rates were summarised overall and by deanery, speciality, indicator, and deanery × indicator. Outcome categories (below, Q1-not-below, within IQR, Q4-not-above, above) were compared across deaneries using Pearson χ^2 tests, with Cramér's *V* used as an effect size descriptor (28). Finally, the resilience index and PCA/clustering analyses were retained as exploratory summaries. The resilience index combined the breadth of above-national indicators with consistency across indicators and should not be interpreted as a validated psychometric scale. PCA/clustering was used for descriptive pattern recognition only; components are ordered by explained variance by definition, and cluster labels were not externally validated (29, 30).

Results

Deanery-level performance

The analytic dataset comprised 26,190 scored observations across 16 deaneries, 31 medical specialities, and 18 indicators (Table 1). Deanery sample size ranged from 613 scored observations [the

TABLE 1 Deanery-level summary statistics. Gap was calculated as the site mean minus the national mean for all medical speciality posts. The *N* score indicates the number of unsuppressed scored site × speciality × indicator observations included after exclusions.

Deanery	Sites	Specialities	*N* (scored)	Mean score	Mean gap	SD gap	Median gap
North East	18	17	1,477	73.77	+3.23	11.11	4.06
North West London	11	16	1,076	71.71	+1.43	12.03	2.62
NIMDTA	13	16	613	71.76	+1.37	10.88	1.74
NES	31	19	2039	71.70	+1.34	11.23	1.96
South London	17	19	1,536	71.44	+1.14	11.15	1.64
North Central and East London	16	22	1781	71.30	+0.99	11.65	1.54
North West	36	20	3,353	70.82	+0.56	10.91	1.14
South West	18	18	1915	70.78	+0.39	10.45	1.05
Thames Valley	8	18	702	70.53	+0.09	10.25	1.11
Wessex	12	17	1,185	70.47	+0.07	10.69	0.35
West Midlands	22	19	2,102	69.75	−0.82	10.06	−0.27
Health Education and Improvement Wales	20	14	1,075	68.56	−1.79	12.18	−1.70
Kent, Surrey, and Sussex	20	17	2053	68.46	−1.93	10.18	−1.55
East Midlands	15	19	1,480	67.74	−2.74	11.50	−2.20
Yorkshire and the Humber	22	19	1841	67.52	−2.90	11.59	−1.71
East of England	18	17	1962	67.27	−3.13	10.56	−2.50

Northern Ireland Medical and Dental Training Agency (NIMDTA)] to 3,353 (North West). The mean deanery score ranged from 67.27 in the East of England to 73.77 in the North East. Across all indicators and specialities, the mean deanery gap ranged from +3.23 (North East) to −3.13 (East of England), a spread of 6.36 points.

Deanery × indicator profile: breadth versus domain-specific deficits

The 16 × 18 deanery indicator matrix showed broad variation across domains (Figure 1). The North East was positive on 17 of 18 indicators, with particularly large gaps in regional teaching (+7.34), rota design (+4.55), and reporting systems (+4.44). The East of England showed the largest negative gap for study leave (−10.17). Yorkshire and the Humber; the East Midlands; and Kent, Surrey, and Sussex were below the national benchmark across all 18 indicators.

Speciality-level signals

Among the 16 major specialities (≥50 scored observations), speciality-indicator fingerprints showed wide dispersion and distinct profiles (Figure 2). Palliative medicine was above the national benchmark for all indicator means, with gaps ranging from +5.01 (facilities) to +22.49 (rota design). General (internal) medicine showed predominantly negative gaps (17 of 18 indicators), including rota design (−6.51), overall satisfaction (−4.80), and study leave (−10.56); its only positive indicator mean was facilities (+0.32). Stroke medicine demonstrated a mixed but often adverse pattern, including a large negative mean for regional teaching (−21.31).

Deanery effects by indicator

After false discovery rate correction across 18 indicators, 17 indicators showed statistically significant differences in gap distributions across deaneries (Table 2); local teaching did not show statistically significant differences in gap distributions across deaneries. Interpretation was based on effect size and statistical significance. The largest deanery effects (Figure 3) were observed for regional teaching ($H = 129.66$, $\eta^2 = 0.1021$; $\omega^2 = 0.0906$), clinical supervision out of hours ($\eta^2 = 0.0704$; $\omega^2 = 0.0675$), teamwork ($\eta^2 = 0.0672$; $\omega^2 = 0.0622$), and reporting systems ($\eta^2 = 0.0607$; $\omega^2 = 0.0567$).

Between-deanery variance by intraclass correlation coefficient

ICC values showed that only a modest proportion of total site-level variance was attributable to deanery membership for most indicators (Figure 4). The highest ICC was observed for regional teaching (0.0973; approximately 9.7% of variance between deaneries), followed by clinical supervision out of hours (0.0724), teamwork (0.0667), and reporting systems (0.0609). Lower ICC values were observed for domains described as more site-driven, including local teaching (0.0068), feedback (0.0143), and adequate experience (0.0148).

Post-hoc extreme contrasts

Extreme-group contrasts were used to illustrate the magnitude of separation between the three highest and three lowest deaneries for four prespecified indicators; these contrasts were not intended to replace the full indicator-level analyses. All four indicators showed statistically significant bottom-versus-top differences with small-to-moderate effect sizes: study leave: the East of England (mean gap −10.17) versus NIMDTA (+5.31) ($U = 718.0$, $p < 0.001$, $r = 0.532$); rota design: the East of England (−5.46) versus the North Central and East London (+5.45) ($p < 0.001$, $r = 0.430$); clinical supervision: the East Midlands (−2.54) versus North West London (+3.35) ($p < 0.001$, $r = 0.442$); and overall satisfaction: the East of England (−3.28) versus NHS Education for Scotland (NES) (+2.79) ($p < 0.001$, $r = 0.336$).

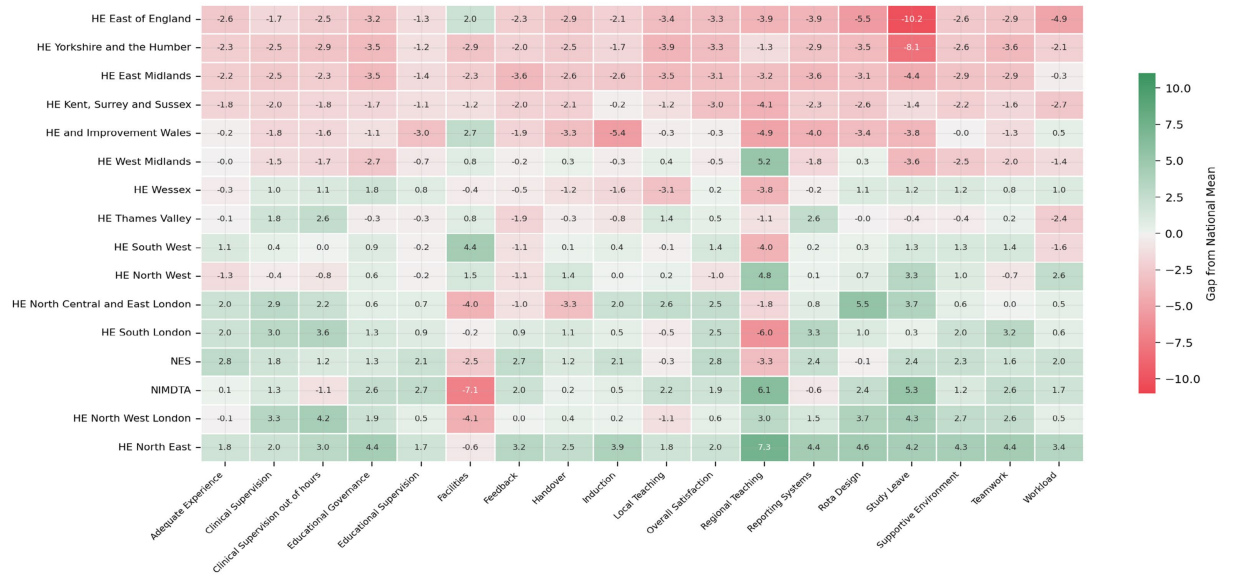


FIGURE 1
Deanery × indicator heatmap. Each cell shows the mean gap score for a deanery and indicator, calculated as the site mean minus the national benchmark. Green indicates an above-national-mean gap, and red indicates a below-national-mean gap.

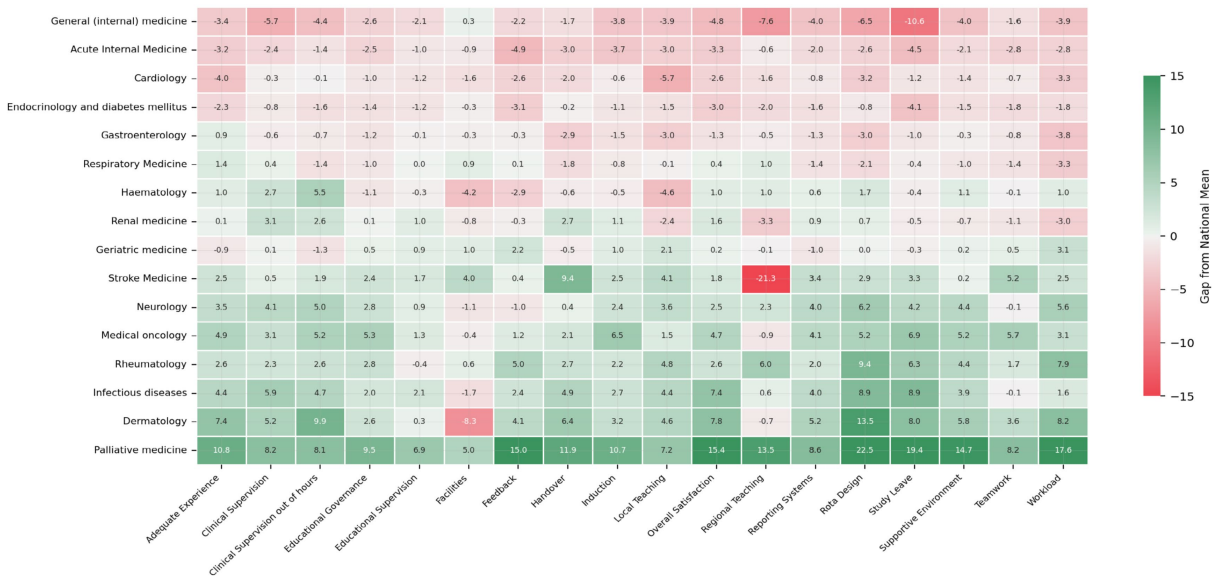


FIGURE 2
Speciality training domain fingerprints. Each cell shows the mean gap from the national benchmark for a speciality and an indicator among the 16 major medical specialities with at least 50 scored observations. Rows are sorted by the overall average gap.

TABLE 2 Deanery effects by indicator. H denotes the Kruskal–Wallis statistic; η^2 denotes eta-squared; F denotes the one-way ANOVA statistic; ω^2 denotes omega-squared; BH adj denotes the Benjamini–Hochberg adjusted p-value. Effect sizes should be interpreted as magnitude descriptors rather than proof of causal deanery effects.

Indicator	N	H	p (BH adj)	η^2	F	p (ANOVA BH adj)	ω^2
Regional teaching	1,139	129.66	<0.0001	0.1021	8.57	<0.0001	0.0906
Clinical supervision out of hours	1,441	115.35	<0.0001	0.0704	7.96	<0.0001	0.0675
Teamwork	1,586	120.52	<0.0001	0.0672	8.02	<0.0001	0.0622
Reporting systems	1,457	102.45	<0.0001	0.0607	6.84	<0.0001	0.0567
Clinical supervision	1,557	102.78	<0.0001	0.0570	6.44	<0.0001	0.0498
Study leave	1,337	85.29	<0.0001	0.0532	5.83	<0.0001	0.0514
Facilities	1,283	75.76	<0.0001	0.0480	5.68	<0.0001	0.0519
Educational governance	1,587	80.55	<0.0001	0.0417	5.37	<0.0001	0.0397
Supportive environment	1,588	76.52	<0.0001	0.0391	5.06	<0.0001	0.0370
Rota design	1,556	66.12	<0.0001	0.0332	4.32	<0.0001	0.0310
Overall satisfaction	1,588	65.55	<0.0001	0.0322	4.07	<0.0001	0.0282
Induction	1,588	61.21	<0.0001	0.0294	3.99	<0.0001	0.0275
Workload	1,588	60.34	<0.0001	0.0288	4.51	<0.0001	0.0321
Educational supervision	1,588	52.90	<0.0001	0.0241	3.18	<0.0001	0.0202
Handover	1,253	38.64	0.0008	0.0191	2.51	0.0014	0.0178
Feedback	1,327	37.61	0.0011	0.0172	2.19	0.0056	0.0133
Adequate experience	1,588	39.96	0.0006	0.0159	2.48	0.0015	0.0138
Local teaching	1,139	22.62	0.0925	0.0068	1.48	0.1040	0.0063

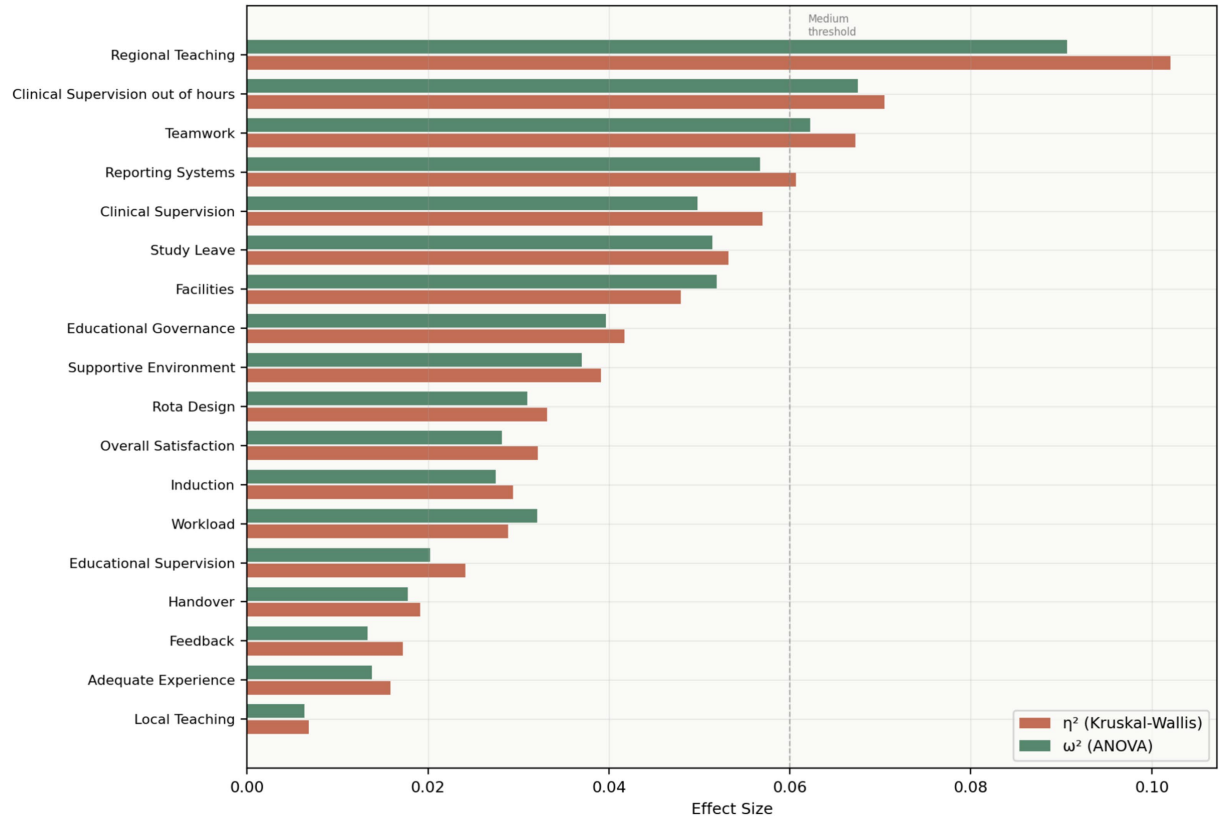


FIGURE 3
Indicator-level effect sizes for deanery differences. Bars show non-parametric η^2 and parametric ω^2. The dashed reference line marks the conventional medium-effect threshold of approximately 0.06; thresholds are interpretative guides rather than educational cutoffs.

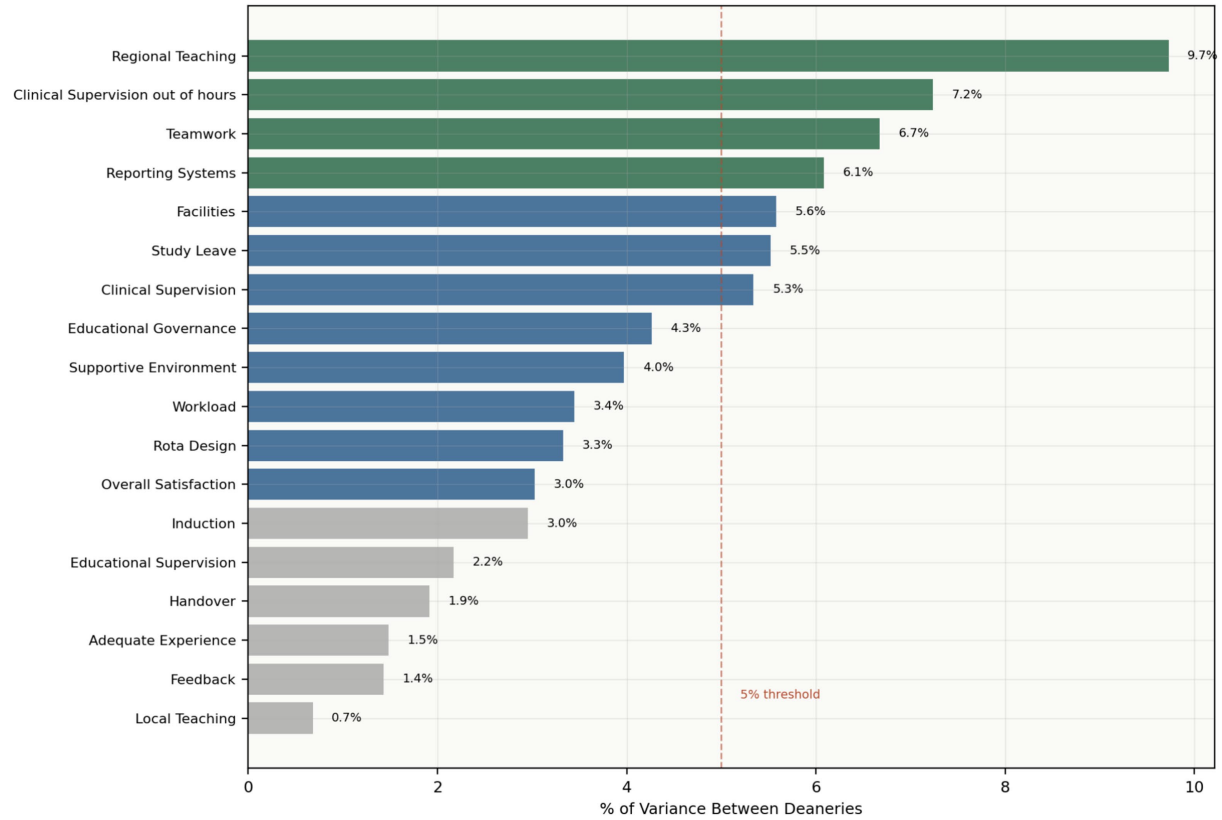


FIGURE 4
Intraclass correlation coefficient (ICC) by indicator. Higher values indicate a greater proportion of site-level variance attributable to deanery membership. The dashed 5% line is a pragmatic visual reference; ICCs should be interpreted as variance descriptors, not causal effects.

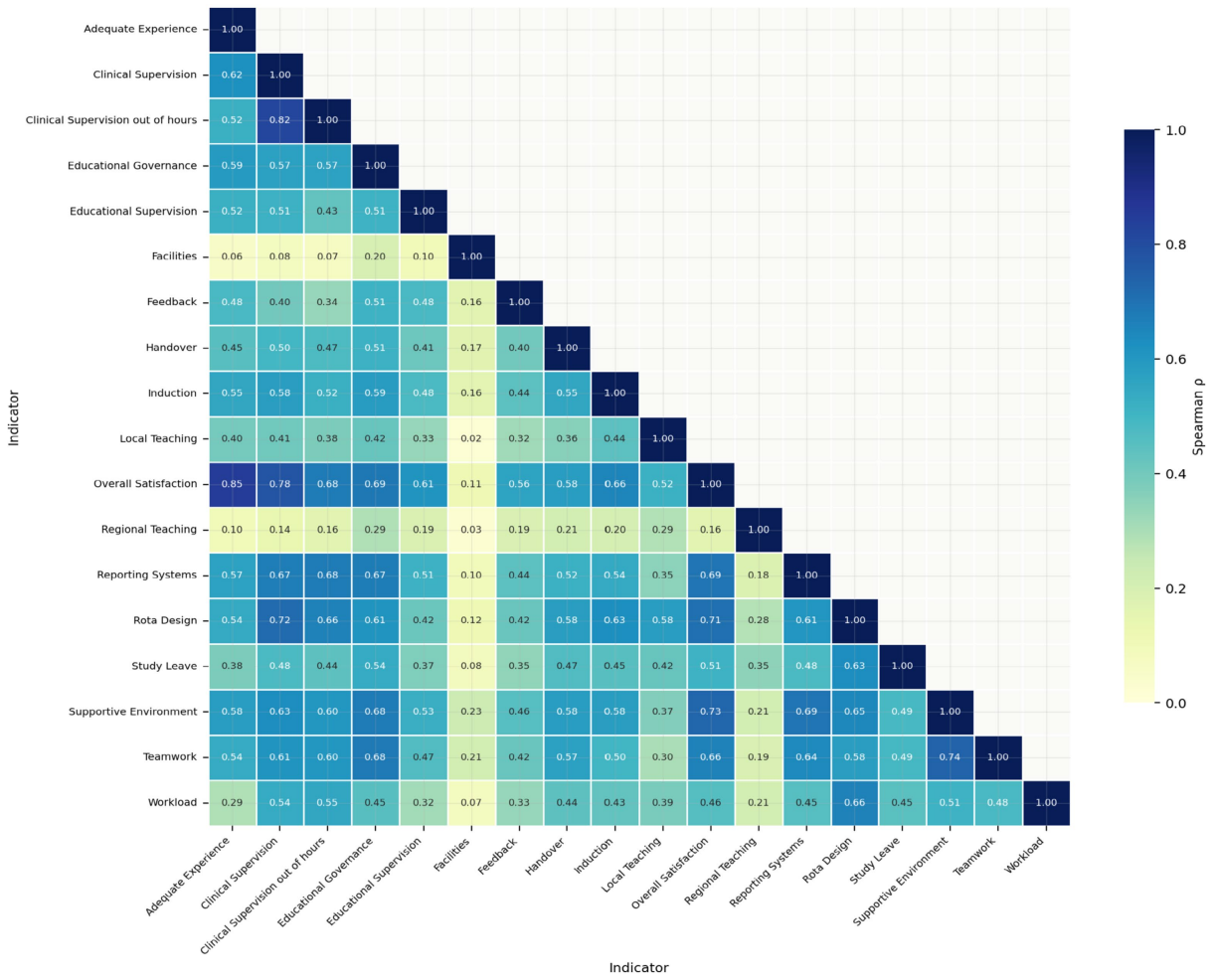


FIGURE 5
Indicator coupling across the sites. The lower-triangle matrix shows Spearman's correlation coefficients between site-level indicator means for the 247 sites with complete data across all 18 indicators. Darker shading indicates a stronger positive correlation; correlations do not establish directionality.

Indicator coupling and volatility

A correlation analysis used the 247 sites with complete data for all 18 indicators (Figure 5). The strongest positive Spearman correlations were adequate experience with overall satisfaction ($\rho = 0.848$; Pearson $r = 0.845$), clinical supervision with clinical supervision out of hours ($\rho = 0.818$), clinical supervision with overall satisfaction ($\rho = 0.781$), and supportive environment with teamwork ($\rho = 0.737$). The weakest couplings involved facilities and teaching metrics, including facilities with local teaching ($\rho = 0.023$) and facilities with regional teaching ($\rho = 0.027$). Indicator volatility rankings (Figure 6) identified study leave as the most site-variable domain (SD = 17.82), followed by rota design (SD = 14.71) and local teaching (SD = 13.53). Supervision domains were comparatively stable (educational supervision, SD = 7.54; clinical supervision, SD = 8.05).

N < 3 suppression

Suppression affected 21,060 of 47,442 medicine rows (44.4%) in the extract (Figure 7). After the prespecified exclusion of the Defence Postgraduate Medical Deanery and an unspecified deanery, the data used for deanery comparisons comprised 26,190 scored observations. Suppression was highest in Health Education and Improvement Wales (HEIW) (57.9%) and NIMDTA (57.6%). Speciality-level suppression was very high for palliative medicine (88.5%), stroke medicine (71.1%), and rheumatology (69.8%). Teaching indicators showed the highest suppression rates, including regional teaching and local teaching (each 52.2%), whereas clinical supervision out of hours had the lowest suppression rate (39.1%).

Outcome distributions

Outcome classification distributions differed by deanery (Figure 8) ($\chi^2 = 597.9$, df = 60, $p < 0.001$), with Cramér's $V = 0.0755$. Notable contrasts included the North East, where 12.32% of scored observations were classified as above and 4.47% as below (above-below $\approx +7.85$ percentage points); Yorkshire and the Humber, where 10.97% were below and 4.89% above (above-below ≈ -6.08 percentage points); and the East of England, which had the lowest above rate (3.47%).

The resilience index was treated as a composite summary rather than a validated outcome measure (Figure 9). The North East had the highest resilience score (92.7), corresponding to 17 of the 18 indicators above the national benchmark. Three deaneries had negative

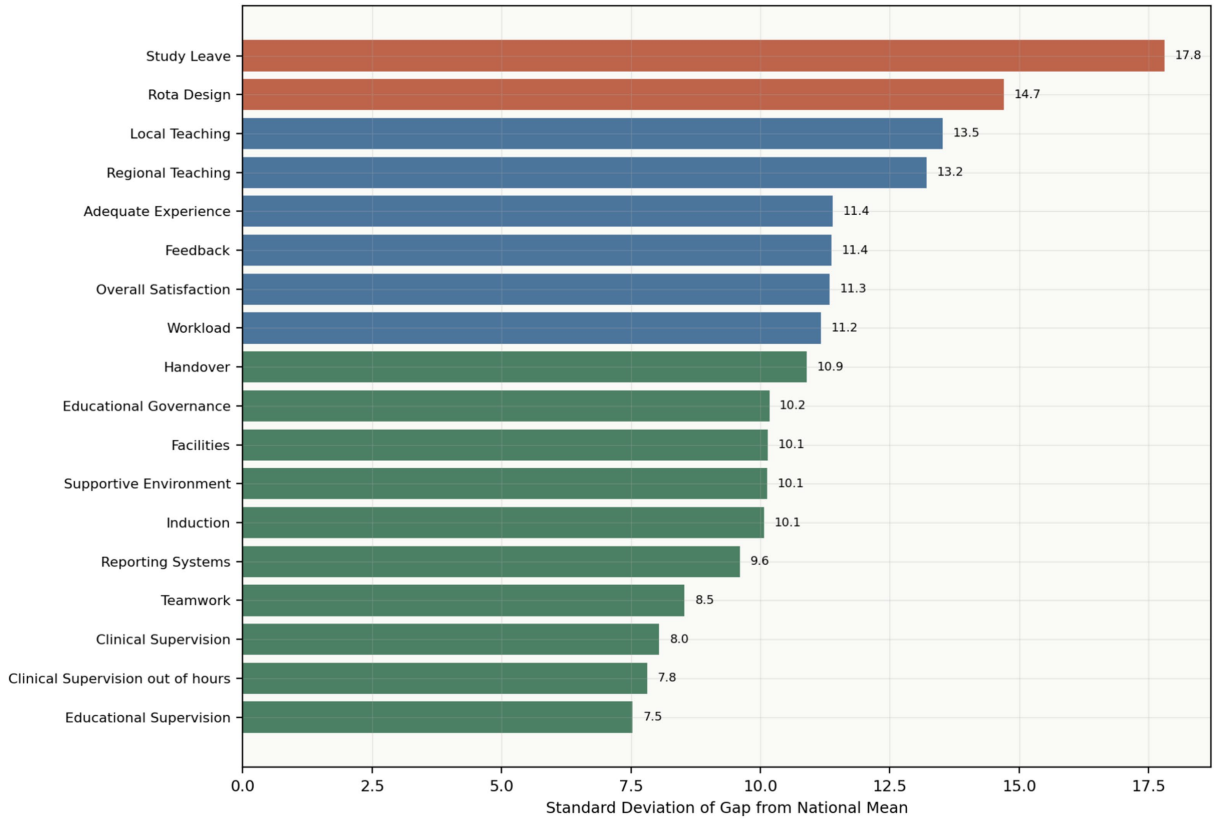


FIGURE 6
Indicator volatility ranking. Volatility was quantified using the standard deviation of site-level gap scores for each indicator. Red indicates the most site-variable indicators, blue indicates intermediate variability, and green indicates the most stable indicators.

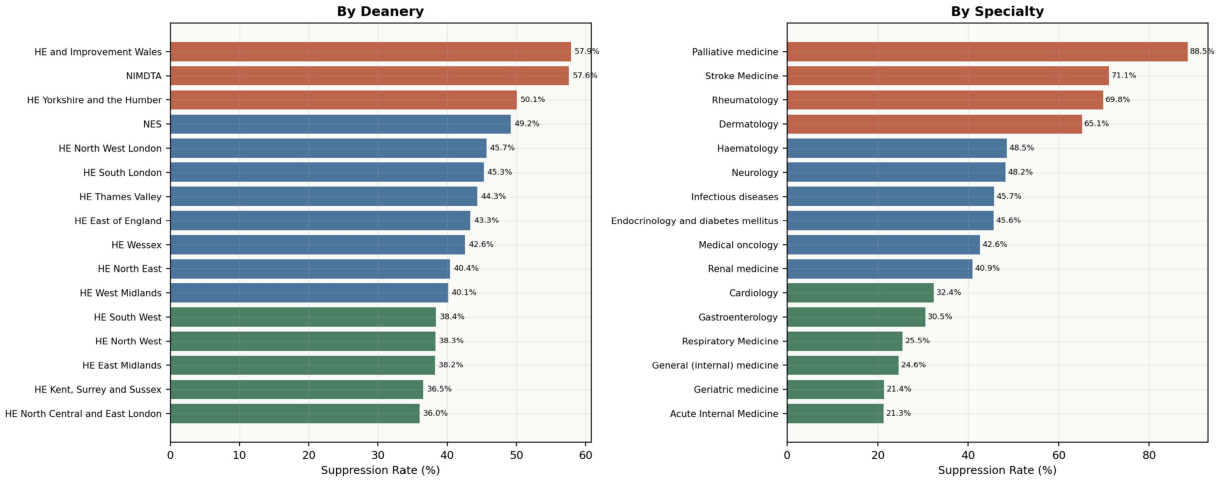


FIGURE 7
Small-number suppression by deanery and speciality. Suppression indicates cells with fewer than three responses for which a score was not reported. Higher suppression may reduce the visibility of performance in smaller deaneries, specialities, or site-speciality cells.

resilience scores: Kent, Surrey, and Sussex (−0.8); the East Midlands (−0.9); and Yorkshire and the Humber (−1.5).

PCA and clustering analyses were exploratory and hypothesis-generating (Supplementary Figures S1, S2). Speciality PCA found that PC1 accounted for 77.08% of the variance explained, with smaller contributions from PC2 (9.16%) and PC3 (6.21%). Since PCA components are ordered by explained variance, this distribution was interpreted descriptively rather than as a separate finding. Silhouette scores supported k = 3–4 cluster solutions (k = 3: 0.401; k = 4: 0.390; k = 5: 0.234). Deanery PCA found that PC1 accounted for 66.53% of the variance explained and PC2 accounted for 9.03% of the variance explained.

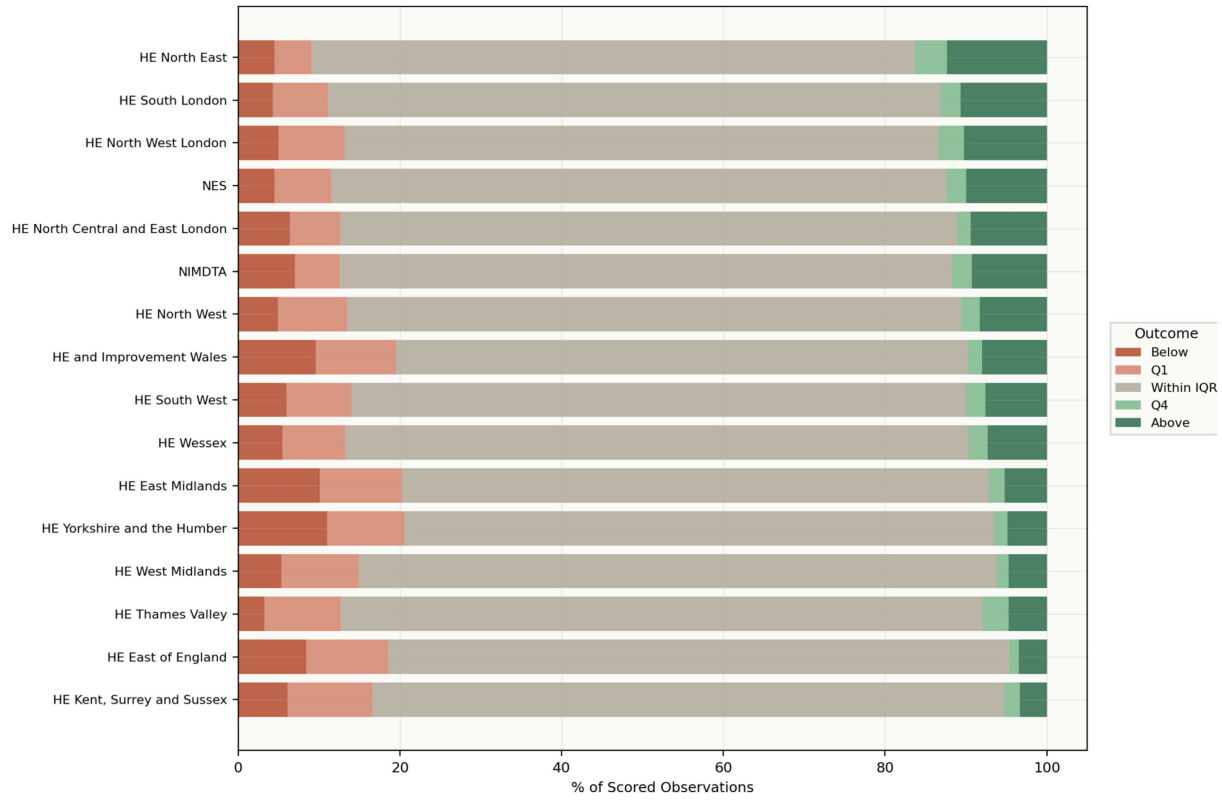


FIGURE 8
Stacked NTS outcome distribution by deanery. Bars show the proportion of scored observations in each official NTS outcome category: below, Q1-not-below, within IQR, Q4-not-above, and above. The comparison is descriptive and is based only on unsuppressed scored observations.

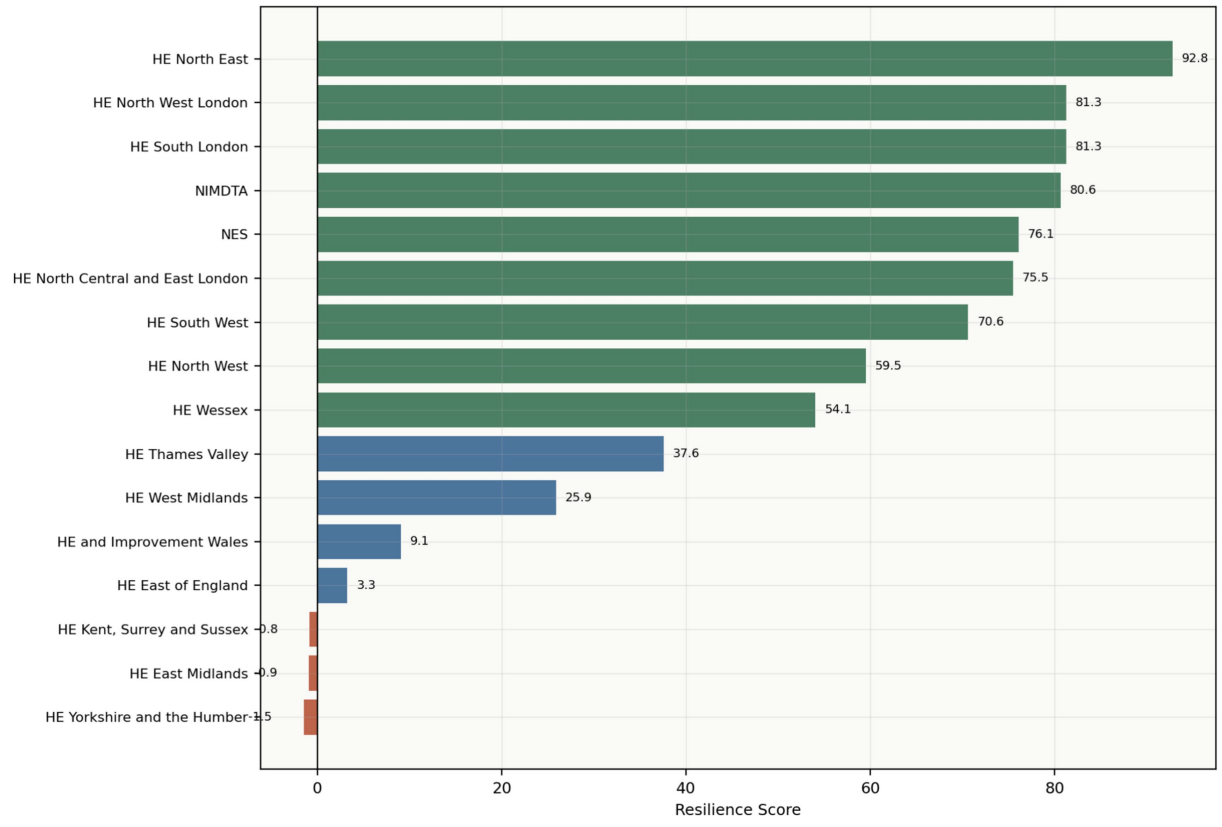


FIGURE 9
Exploratory resilience score by deanery. The score combines the percentage of indicators above the national benchmark with consistency across indicators. Green, blue, and red indicate higher, intermediate, and lower values, respectively. This composite is exploratory and not a validated performance scale.

Within-deanery inequality and speciality heterogeneity

Within-deanery cross-speciality inequality was relatively uniform across deaneries (average SD range: 9.77–11.79), with Health Education and Improvement Wales being the highest (11.79) and the Thames Valley being the lowest (9.77). Inequality was most pronounced in study leave (average SD 17.09) and rota design (average SD 14.17). Across specialities, heterogeneity (average cross-site SD) was highest for rheumatology (12.16) and lowest for palliative medicine (8.53). General (internal) medicine showed lower heterogeneity (9.40) despite consistently low mean gaps.

Two-way interaction (overall satisfaction)

In the deanery × speciality analysis of overall satisfaction for six large specialities (cells restricted to $N \geq 3$), general (internal) medicine was below the national benchmark in every deanery (range −0.3 to −7.5) and showed the least between-deanery SD (2.39). Respiratory medicine showed the greatest between-deanery variation (SD 4.09), indicating larger deanery dependence for this speciality's overall satisfaction gaps (Figure 10).

Discussion

This study used the 2025 GMC NTS post-speciality-by-site extract to examine whether variation in medicine training experience was more consistent with deanery-level patterns or local site/speciality variation. The main finding was not simply that deaneries differed, but that the degree of deanery-level patterning varied substantially by indicator. Regional teaching, clinical supervision out of hours, teamwork, and reporting systems showed the largest deanery effects, whereas local teaching, feedback, and adequate experience had low ICCs and appeared to be more site-driven.

The modest ICC values require careful interpretation. Even the highest ICC, for regional teaching, was 0.0973, indicating that a majority of variation remained outside deanery-level clustering. Therefore, statistically significant deanery differences should not be interpreted as evidence that deanery policy alone determines trainee experience. Instead, the findings support a dual-action model: Deanery-wide interventions may be appropriate where regional patterning is stronger, but site-level diagnostics remain necessary for domains with low ICCs or high local volatility.

The correlation and volatility analyses provide practical prioritisation signals but should not be interpreted causally. Adequate experience was strongly correlated with overall satisfaction, suggesting that trainees' perceptions of sufficient clinical exposure may be a useful lead indicator of broader training sentiment. However, this association cannot determine whether adequate experience drives satisfaction, reflects the same underlying workplace conditions, or both. Study leave and rota design were prominent because they showed large site-to-site variability and clear extreme-group contrasts, suggesting that both regional policy and local implementation may influence these domains.

The results are consistent with wider UK literature indicating that supervision, rota design, teamwork, workload, and work–life balance are meaningful components of the training environment (11–13). The GMC NTS report also highlights continued concerns around workload, burnout, and trainee escalation culture (16). These data support the relevance of NTS indicators as system-level signals.

Study leave warrants particular attention because it had the greatest site-to-site variability in this analysis. Contemporary trainee survey work on learning experiences abroad describes perceived educational value, accessibility concerns, and funding barriers, which aligns with the interpretation that leave processes may affect perceived fairness and opportunity (31). Rota design is similarly actionable: Recent scheduling studies report associations between rota structures and resident wellness or fatigue, supporting rota design as a plausible organisational lever while recognising that these associations are context-dependent (32, 33). The suppression audit is important for interpreting the visibility of training quality. Approximately 44% of medicine rows were suppressed because fewer than three trainees responded. This may disproportionately reduce visibility in smaller specialities, smaller deaneries, and low-count site-speciality cells, potentially masking both excellence and concern. The present study summarised suppression by deanery, speciality, indicator, and deanery × indicator but did not perform additional suppression-performance correlation analyses. Such analyses should be performed in future studies if row-level denominators and sufficient unsuppressed data are available.

The PCA/clustering and resilience analyses were conducted as exploratory summaries. PCA is useful for visualising multivariate profiles; however, the ordering of components by explained variance is inherent to the method and should not be overinterpreted. Cluster groupings were not externally validated, and terms such as "tiers" have therefore been avoided in favour of cautious descriptive wording. Similarly, the resilience index is an *ad hoc* composite that may help summarise breadth and consistency, but it should not be treated as a validated measure of training quality.

Strengths and limitations

The strengths of this study include the use of a national GMC dataset, a transparent gap score definition, correction for multiple testing, and consistent reporting of effect sizes alongside *p*-values. The analysis also adds ICC, coupling, volatility, and suppression summaries to make the results more interpretable for quality improvement rather than simple ranking.

The limitations are substantial. First, the study used aggregated site × speciality × indicator cells rather than individual trainee responses. Rows therefore cannot be assumed to represent independent trainees, and non-hierarchical tests may overstate statistical precision. Multilevel or mixed-effects modelling was not performed in this revision; future analyses with appropriate individual-level or weighted aggregated data should model trainees nested within sites, specialities, and deaneries. Second, the design is cross-sectional and perception-based, so associations cannot establish direction or causality. Third, a negative gap represents deviation from a national benchmark and should not be interpreted as an absolute unacceptable threshold. Fourth, small-number suppression caused substantial information loss and may cause bias visibility towards larger sites or specialities. Finally, exploratory PCA/clustering and the resilience index require external validation before being used for performance classification.

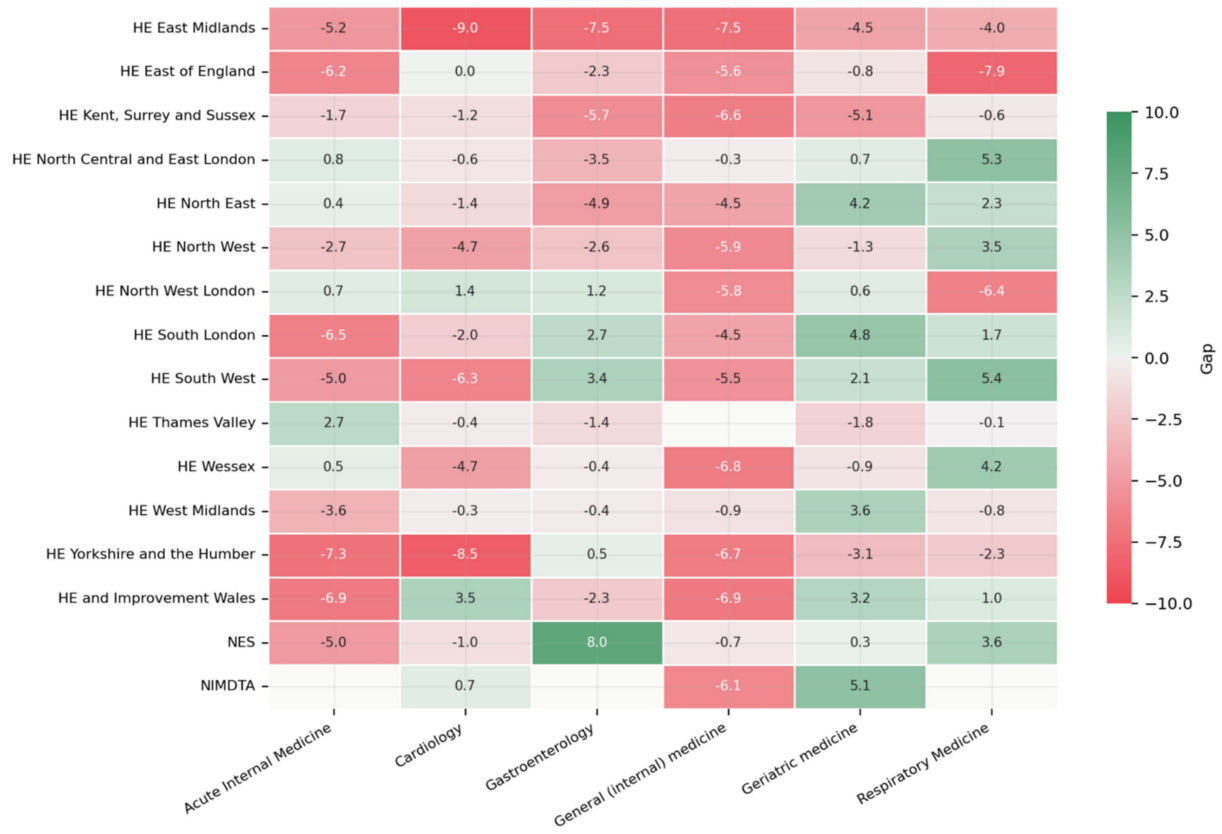


FIGURE 10
Two-way deanery × speciality heatmap for overall satisfaction. Cells show the mean overall satisfaction gap for six large specialities within deaneries, restricted to cells with $N \geq 3$. Green indicates above the national gap, and red indicates below the national gap.

Future directions

Future studies should evaluate whether targeted interventions in high-leverage domains, such as rota design, study leave processes, and out-of-hours supervision, improve subsequent NTS outcomes. Organisational interventions can reduce burnout, although effects vary by context (34). Qualitative follow-ups at outlier sites could clarify the operational mechanisms behind extreme gaps, and triangulation with objective measures such as rota gaps, teaching provision, exception reporting, and study-leave approval data would strengthen causal interpretation.

In summary, the study suggests that training experience in UK medicine varies across deaneries; however, much of the actionable variation remains at the site level. Separating where differences exist from where variance is concentrated provides a more cautious and practically useful interpretation of NTS data than ranking deaneries alone.

Conclusion

This analysis of the 2025 GMC National Training Survey for medicine shows that trainees' experiences vary across the UK, with some indicators showing modest deanery-level clustering and others appearing predominantly site-driven. Regional teaching, out-of-hours supervision, teamwork, and reporting systems showed the clearest deanery-level patterns, while study leave and rota design varied most between sites. The findings support coordinated deanery action for domains that reflect regional systems, alongside targeted site-level review where local implementation dominates. Interpretation should remain cautious because the analysis used aggregated cross-sectional survey data, did not include multilevel modelling, and was affected by substantial small-number suppression.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

AD: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing, Writing – original draft. SM: Writing – original draft, Writing – review & editing. GA: Writing – review & editing, Writing – original draft. DS: Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2026.1833862/full#supplementary-material

References

1. Billett S. Learning through health care work: premises, contributions and practices. *Med Educ*. (2016) 50:124–31. doi: 10.1111/medu.12848

2. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: a literature review. *Med Educ*. (2000) 34:827–40. doi: 10.1046/j.1365-2923.2000.00758.x

3. Rothwell C, Kehoe A, Farook SF, Illing J. Enablers and barriers to effective clinical supervision in the workplace: a rapid evidence review. *BMJ Open*. (2021) 11:e052929. doi: 10.1136/bmjopen-2021-052929

4. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare staff wellbeing, burnout, and patient safety: a systematic review. *PLoS One*. (2016) 11:e0159015. doi: 10.1371/journal.pone.0159015

5. Dewa CS, Loong D, Bonato S, Trojanowski L, Rea M. The relationship between resident burnout and safety-related and acceptability-related quality of healthcare: a systematic literature review. *BMC Med Educ*. (2017) 17:195:195. doi: 10.1186/s12909-017-1040-y

6. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ*. (2016) 50:132–49. doi: 10.1111/medu.12927

7. Naji L, Singh B, Shah A, Naji F, Dennis B, Kavanagh O, et al. Global prevalence of burnout among postgraduate medical trainees: a systematic review and meta-regression. *CMAJ Open*. (2021) 9:E189–200. doi: 10.9778/cmajo.20200068

8. Hodkinson A, Zhou A, Johnson J, Geraghty K, Riley R, Zhou A, et al. Associations of physician burnout with career engagement and quality of patient care: systematic review and meta-analysis. *BMJ*. (2022) 378:e070442. doi: 10.1136/bmj-2022-070442

9. Degen C, Weigl M, Glaser J, Li J, Angerer P. The impact of training and working conditions on junior doctors' intention to leave clinical practice. *BMC Med Educ*. (2014) 14:119:119. doi: 10.1186/1472-6920-14-119

10. Maisonneuve JJ, Lambert TW, Goldacre MJ. Doctors' views about training and future careers expressed one year after graduation by UK-trained doctors: questionnaire surveys undertaken in 2009 and 2010. *BMC Med Educ*. (2009) 14:270. doi: 10.1186/s12909-014-0270-5

11. Barton JC, Richardson D, Corrigan AE, Solomons MR, Kuri A, Round J. Supervision, scheduling, satisfaction and shared working: how experiences of junior doctors relate to excess mortality within the NHS. *Clin Med (Lond)*. (2023) 23:582–7. doi: 10.7861/clinmed.2023-0129

12. Atraszkiewicz D. Redesigning a general surgery foundation doctor rota to improve shift equity, staffing levels, doctors' pay and patient care. *Future Healthc J*. (2025) 12:100253:100253. doi: 10.1016/j.fhj.2025.100253

13. Parida S, Aamir A, Alom J, Rufai TA, Rufai SR. British doctors' work-life balance and home-life satisfaction: a cross-sectional study. *Postgrad Med J*. (1169) 99:198–206. doi: 10.1136/postgradmedj-2021-141338

14. Hockey P, Vaithianathan R, Baeker A, Beer F, Goodall AH, Hammerton M, et al. Measuring the working experience of doctors in training. *Future Healthc J*. (2020) 7:e17–22. doi: 10.7861/fhj.2020-0005

15. Uthayanan M, Szram J, Mehta A, Menon G, Round J. The GMC national training survey: does it have an impact? *Future Healthc J*. (2020) 7:205–7. doi: 10.7861/fhj.2020-0031

16. General Medical Council. National training survey 2025 results report. London: General Medical Council (2025). Available online at: https://www.gmc-uk.org/cdn/documents/national-training-survey-2025-results-report_pdf-111657596.pdf (Accessed March 12, 2026)

17. General Medical Council. Indicator - trainee: post specialty by site. London: General Medical Council. Available online at: https://edt.gmc-uk.org/indicator---trainee/post-specialty-by-site (Accessed March 12, 2026)

18. Donald N, Lindsay T. Surgical trainee experiences from 2013 to 2023 within the United Kingdom as reported by the general medical council National Training Survey. *Surgeon*. (2013) 22:74–9. doi: 10.1016/j.surge.2023.11.005

19. Chan J, Oo S, Yeatman M. Is there a national variation in cardiothoracic trainees' satisfaction? Evaluation of the United Kingdom general medical council national training surveys. *J Card Surg*. (2019) 34:1069–74. doi: 10.1111/jocs.14211

20. O'Hanlon JR, Chan J, Fudulu DP, Angelini GD. The impact of COVID-19 on the job satisfaction of cardiothoracic trainees in the United Kingdom: an insight from the general medical council National Training Survey. *Cureus*. (2024) 16:e73682. doi: 10.7759/cureus.73682

21. Lamprou V, Karthikeyan VJ, Campbell NG, Sastry S. Utilising the general medical council survey as an incentive to improve training satisfaction in cardiology: a case study. *Future Healthc J*. (2025) 12:100454:100454. doi: 10.1016/j.fhj.2025.100454

22. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc*. (260) 47:583–621. doi: 10.1080/01621459.1952.10483441

23. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. (1965) 52:591–611. doi: 10.1093/biomet/52.3-4.591

24. Levene H. "Robust tests for equality of variances". In: Olkin I and Hotelling H, editors. Contributions to Probability and Statistics. Stanford: Stanford University Press (1960). p. 278–92.

25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

26. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. HillsdaleLawrence. United States of America: Erlbaum Associates (1988).

27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

28. Cramer H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press (1946).

29. Ward JH Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. (301) 58:236–44. doi: 10.1080/01621459.1963.10500845

30. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. (1987) 20:53–65. doi: 10.1016/0377-0427(87)90125-7

31. Waugh D, Hayburn A, Johnston M. Learning experiences abroad adding value in medical education (LEAVE): results from the LEAVE trainee experience survey. *Cureus*. (2025) 17:e82779. doi: 10.7759/cureus.82779

32. Heppe D, Baduashvili A, Limes JE, Suddarth K, Mann A, Gottenborg E, et al. Resident burnout, wellness, professional development, and engagement before and after new training schedule implementation. *JAMA Netw Open*. (2024) 7:e240037. doi: 10.1001/jamanetworkopen.2024.0037

33. Yuan JH, Huang Y, Rosgen BK, Donnelly S, Lan X, Katz SJ. Burnout and fatigue amongst internal medicine residents: a cross-sectional study on the impact of alternative scheduling models on resident wellness. *PLoS One*. (2023) 18:e0291457. doi: 10.1371/journal.pone.0291457

34. Panagioti M, Panagopoulou E, Bower P, Lewith G, Kontopantelis E, Chew-Graham C, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med*. (2017) 177:195–205. doi: 10.1001/jamainternmed.2016.7674