

Heathrow Airport Limited: Technical Review and Critique of Commissioned Benchmarking Studies

Prepared by

Oxford Global Projects (OGP)

Dr. Alexander Budzier, Dr. Sam Franzen, Andreas Leed, Dirk Pöker
November 2025

FOR PUBLICATION

1. EXECUTIVE SUMMARY	1
KEY FINDINGS	1
2. INTRODUCTION	2
2.1. PURPOSE AND CONTEXT	2
2.2. HEATHROW'S STRATEGIC AND REGULATORY SETTING	2
2.3. COMMISSIONED STUDIES UNDER REVIEW	3
2.4. ANALYTICAL APPROACH AND PHILOSOPHY	3
3. OVERVIEW OF HAL'S COMMISSIONED STUDIES	4
4. METHODOLOGICAL REVIEW	6
4.1. COMPARATOR SELECTION AND REPRESENTATIVENESS	6
4.2. INFLATION AND COST YEAR TREATMENT	6
4.3. FOREIGN EXCHANGE AND PURCHASING POWER TREATMENT	7
4.4. COST NORMALISATION AND PRODUCTIVITY ADJUSTMENT	9
4.5. TREATMENT OF SCOPE AND QUALITY DIFFERENCES	9
4.6. TREATMENT OF ASSUMPTIONS AND UPLIFTS	10
4.7. STATISTICAL TESTING AND SENSITIVITY ANALYSIS	11
4.8. DATA TRANSPARENCY AND REPRODUCIBILITY	11
4.9. SUMMARY OF METHODOLOGICAL WEAKNESSES	12
5. DETAILED STUDY CRITIQUES	13
5.1. STEER (2024) – CAPEX BENCHMARK OF TERMINAL BUILDING PROJECTS	13
5.1.1. OVERVIEW	13
5.1.2. KEY METHODOLOGICAL WEAKNESSES	13
5.1.3. IMPLICATIONS	16
5.1.4. RECOMMENDED CORRECTIONS	17
5.2. KPMG (AERONAUTICAL REVENUE BENCHMARKING, 2025)	17
5.2.1. OVERVIEW	17
5.2.2. KEY METHODOLOGICAL WEAKNESSES	17
5.2.3. IMPLICATIONS	20
5.2.4. RECOMMENDED CORRECTIONS	20
5.3. ARCADIS (CAPEX EFFICIENCY REVIEW, 2020)	20
5.3.1. OVERVIEW	20
5.3.2. KEY METHODOLOGICAL WEAKNESSES	21
5.3.3. IMPLICATIONS	21
5.3.4. RECOMMENDED CORRECTIONS	22
6. FINDINGS AND IMPLICATIONS	22
7. INCENTIVE MISALIGNMENT AND REGULATORY RISK	23

1. Executive Summary

Heathrow Airport Limited (HAL) has commissioned a series of external benchmarking and efficiency studies to demonstrate that its costs are justified within the context of London's construction market and the airport's operational complexity. These studies form the core of HAL's evidence for the H8 regulatory period and aim to demonstrate to the Civil Aviation Authority that Heathrow's higher capital and revenue costs are justified.

This interim critique, prepared by Oxford Global Projects (OGP) on behalf of the airline community, evaluates the credibility of that evidence.

The review focuses on Steer's Capex Benchmark of Terminal Building Projects and is supplemented by targeted assessments of KPMG's Aeronautical Revenue Benchmarking and the CAA-commissioned Arcadis Capex Efficiency Review.

Taken together, these studies form the basis of HAL's cost efficiency case, yet they share methodological weaknesses that reduce the reliability of that case.¹

Key Findings

- **Selective datasets.**

All three studies use small samples without published inclusion criteria. The resulting benchmarks are not statistically representative and tend to align with Heathrow's own cost levels.

- **Inconsistent inflation and FX treatment.**

Costs are converted using nominal exchange rates and then inflated with UK indices, applying two inflation systems to the same values and biasing results toward Heathrow.

- **Overlapping uplifts.**

Adjustments for location, congestion and complexity are applied cumulatively, often capturing the same underlying drivers. In some cases, these uplifts almost double comparator values.

- **Input-based rather than output-based metrics.**

Measures such as cost per square metre or revenue per passenger do not capture capacity or throughput. Output-based metrics are required to assess delivery efficiency and space use.

¹ This report is not intended to provide an exhaustive review of all issues previously raised by Heathrow or discussed with the Civil Aviation Authority. It focuses only on methodological points relevant to the present benchmarking assessment.

- **No statistical or sensitivity testing.**

The studies do not provide statistical evidence or test how results change under alternative assumptions.

- **Lack of independent validation.**

The analyses draw on the same data sources and assumptions, giving the appearance of corroboration without providing independent verification.

- **Omission of design efficiency.**

The studies do not assess whether projects were required at their delivered scale or specification. Without testing alternative design choices, the benchmarks cannot distinguish efficient delivery from over-design.

2. Introduction

2.1. Purpose and Context

This report provides a critical technical review of the benchmarking studies commissioned by Heathrow Airport Limited (HAL) to justify its capital efficiency and high aeronautical charges in preparation for the H8 regulatory period (2027–2031).

HAL relies on external studies by Steer, KPMG and Arcadis to argue that Heathrow’s higher costs reflect London’s construction market and the airport’s operational complexity. These studies claim that, once adjusted for these factors, Heathrow’s performance lies within the international range.

However, the Civil Aviation Authority and the airline community have raised concerns that these conclusions rely on selective assumptions and opaque methods. The purpose of this report is to test the robustness of HAL’s cost benchmarking evidence, identify methodological weaknesses, and assess whether the studies provide a reliable foundation for regulatory and pricing decisions.

The findings will inform the design of a stage 2: Independent Benchmarking, in which OGP will apply its international reference dataset and Reference Class Forecasting (RCF) methods to assess Heathrow’s actual cost performance across terminals, baggage systems, car parks, and runways.

2.2. Heathrow’s Strategic and Regulatory Setting

Heathrow remains the UK’s most important aviation hub, handling over 83 million passengers in 2024 and 21% of the nation’s total goods trade by value. Yet, it is also the most

expensive major airport in the world and its passenger charges are roughly twice those of comparable European hubs such as Frankfurt, Paris CDG, and Amsterdam.

Under the RAB model, HAL's income rises with asset growth rather than efficiency. Each pound invested expands the regulated asset base and boosts future revenues through depreciation and return on capital. The model therefore risks encouraging overinvestment and weak cost control, an issue consistently highlighted by the airline community and the CAA. Heathrow's £10 billion H8 programme, despite a nominal £500 million "efficiency target," thus merits rigorous examination.

2.3. Commissioned Studies Under Review

HAL's benchmarking case draws on three studies, two commissioned by HAL (Steer 2024 and KPMG 2025) and one commissioned by the CAA (Arcadis 2020)::

- **Steer (2024):** claims Heathrow's terminal construction costs are within international norms after location and congestion adjustments.
- **KPMG (2025):** argues that Heathrow's aeronautical revenues per passenger are justified when adjusted for London's cost structure, tax regime, and hub complexity.
- **Arcadis (2020):** a CAA-commissioned review of Q6 project delivery, concluding that most projects were delivered efficiently based on internal criteria.

These studies underpin Heathrow's claim that its capital plan offers "value for money" despite its global price premium.

2.4. Analytical Approach and Philosophy

Our critique applies OGP's established international benchmarking methodology. This approach has been developed through extensive academic and applied work and aligns with recognised good practice used by organisations such as the OECD, World Bank, and national audit bodies. The key principles are:

- all cost data should be harmonised to a consistent base year and expressed in a common international price level, typically using purchasing power parity or equivalent price-level adjustments;
- inflation and foreign exchange adjustments should be applied in an economically coherent sequence, with local inflation applied first, followed by translation into a common price level, and only then the application of UK inflation to the comparison year;

- scope and quality differences should be addressed through explicit functional controls, such as passenger capacity, baggage throughput, or design specification;
- comparator samples should be reasonably representative and constructed with transparent inclusion and exclusion criteria;
- results should be reproducible, with sufficient documentation of data sources, cost years, exchange rates, and adjustment methods.

These principles ensure that benchmarking results reflect genuine differences in efficiency rather than artefacts of data treatment or methodological choices. The review finds that the HAL-commissioned studies diverge from these principles in several respects, which affects the reliability of their conclusions.

In particular, they conflate cost estimates and outturns, apply cumulative and unverified adjustments, and handle inflation inconsistently, practices that give an illusion of analytical rigour while biasing results in Heathrow's favour. Adjustments for local complexity are often introduced without evidence, and productivity differences are not normalised, undermining the validity of cost-per-unit comparisons.

3. Overview of HAL's Commissioned Studies

HAL's cost-efficiency case rests primarily on three consultant reports produced between 2020 and 2025. Each report examines one dimension of performance, but together they form the evidence base HAL uses to justify its investment plans and regulatory submissions.

Across the three reports, several common issues appear:

- small comparator samples without transparent inclusion criteria;
- reliance on shared data sources, which reduces independence;
- limited transparency that prevents replication.

These issues weaken the overall evidence base.

Study	Year	Author	Scope	Key Claim	Dataset	Notes / Observations
Capex Benchmark of Terminal Buildings	2024	Steer	Cost benchmarking (Capex/m ²) for terminal projects	Heathrow within the international range once adjusted for London costs and congestion	≈ 20 projects (mix of estimates and out-turns)	Small, opaque sample; arbitrary 25 % “congestion” uplift; weak inflation sequencing; no statistical testing.
Aeronautical Revenue Benchmarking	2025	KPMG	Comparison of aeronautical revenue per passenger	Heathrow not unusually expensive after adjustment for local conditions	7 airports	Heavy assumption-stacking; reliance on Arcadis ICC indices; potential double-counting of location and constraint factors.
Capex Efficiency Review (CAP1964a)	2020	Arcadis	Review of HAL’s project delivery (Q6 period)	HAL projects generally efficient; most overruns driven by external factors	10 projects	CAA-commissioned; qualitative assessment; limited evidence base; relies on HAL data.

4. Methodological Review

This section assesses the methodological foundations of the commissioned studies in order of their likely impact on Heathrow's benchmarked position. Although each study has its own scope, they share a common methodological architecture and, consequently, a common set of limitations. These are not minor technical details; they materially affect the credibility of the findings and the confidence that can be placed in the resulting policy conclusions.

4.1. Comparator Selection and Representativeness

Comparator selection is the most consequential methodological choice because it determines the benchmark distribution. If the dataset is selective, narrow or skewed toward atypical markets, the benchmark becomes biased. While small samples are common in international benchmarking due to limited public data, this increases rather than reduces the need for transparent and well-justified selection criteria. Quality and relevance of the sample is key.

All three studies use small comparator sets and do not explain how projects were selected. They mix design estimates and out-turns, which introduces optimism bias. The samples also omit major international terminals. These factors weaken representativeness and may skew results in Heathrow's favour.

The samples also over-represent mature European markets while omitting high-capacity peers in Asia and North America. They conflate projects of differing scope and functionality, reducing comparability further. A transparent benchmarking exercise should document inclusion criteria, data sources and scope definitions to enable replication and independent scrutiny. None of the reviewed studies meet this standard. Their benchmarks are therefore selective rather than representative.

4.2. Inflation and Cost Year Treatment

A second methodological issue concerns the treatment of inflation and cost-year adjustments. Across the studies, inflation and currency conversions are often applied in a sequence that is not economically coherent. In several cases, foreign project costs are converted into pounds using nominal exchange rates and then inflated using UK construction indices. This applies two inflation systems to the same value and raises comparator costs.

Nominal exchange rates already incorporate differences in inflation across countries because currencies tend to depreciate or appreciate in line with domestic price growth. When a foreign cost is converted into GBP using nominal FX, these cumulative inflation

effects are already embedded in the converted value. Applying UK inflation on top introduces a second inflation layer, overstating comparator costs and narrowing Heathrow's apparent cost gap.

A related concern is the reliance on a single construction index when performing these adjustments. Construction indices differ in scope, data sources and sensitivity to local wage, productivity and material-market conditions. Dependence on one source risks embedding its assumptions into the results. A more robust method would triangulate across several reputable indices or, at minimum, test the sensitivity of Heathrow's position to alternative sources.

An economically sound benchmark must separate geographic normalisation from temporal inflation and apply these steps in that order. The correct sequence is as follows:

1. Convert costs to UK-equivalent prices at the project's cost year, using PPP or a sectoral construction index to express what the project would have cost if built in the United Kingdom at that time.
2. Inflate these UK-equivalent values to a common comparison year using a cost index.

This approach ensures that inflation is applied once within a consistent UK price system. As a result, the benchmark reflects genuine differences in cost efficiency rather than artefacts created by index choice or inflation sequencing.

4.3. Foreign Exchange and Purchasing Power Treatment

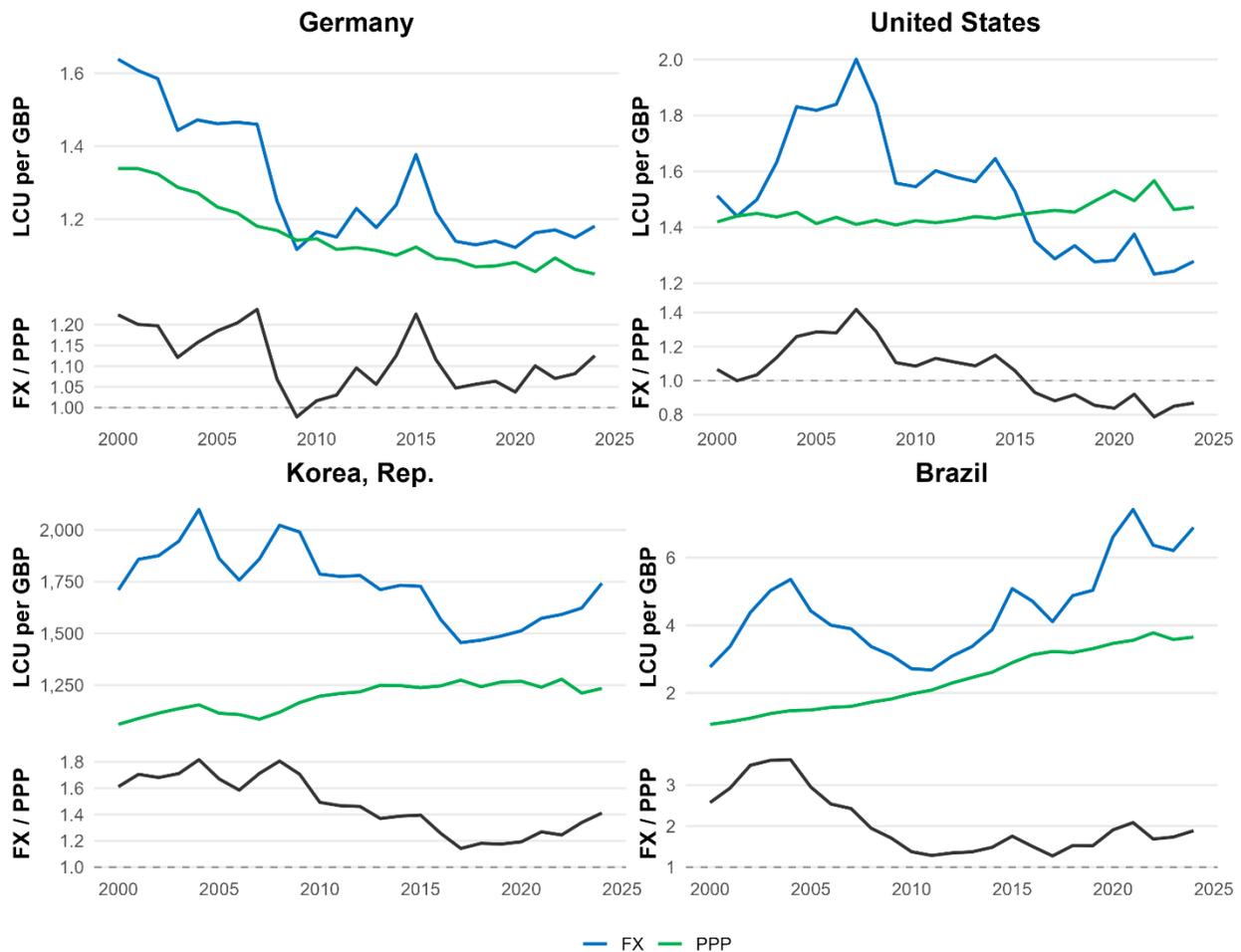
Currency conversion is also handled inconsistently. Both Steer and KPMG rely on nominal market exchange rates rather than PPP-adjusted equivalents, even though nominal FX reflects short-term capital flows and macro-financial conditions rather than construction purchasing power. Applying nominal FX to long-lived infrastructure costs introduces volatility unrelated to real input prices.

In parallel, the studies apply the Arcadis International Construction Cost (ICC) index as a proxy for local construction prices. Because this index already embeds productivity, wage and material-cost differentials, combining it with nominal FX conversion captures overlapping effects and can overstate London's price premium.

Figure 1 compares nominal exchange rates (FX) and PPP conversion factors in local currency units per GBP (LCU per GBP). The lower panels show the ratio of FX to PPP. Values above one indicate that FX converts foreign costs into GBP at a higher price level than PPP, making comparator projects appear more expensive relative to the United Kingdom. Values below one imply that FX understates local prices and narrows the apparent CAPEX gap.

The charts show that FX and PPP diverge substantially for several countries, particularly the United States, Korea and Brazil. In these cases, using FX instead of PPP inflates the GBP-denominated cost of comparator projects for reasons unrelated to construction markets. This reinforces the central critique: FX-based conversions tend to overstate Heathrow’s apparent cost parity, whereas PPP-based conversions are more likely to reveal a wider and more economically grounded cost gap.

Figure 1: Divergence Between Nominal Exchange Rates and PPP Conversions (2000–2024)



A practical example illustrates the scale of this issue. One study claims that constructing in Madrid is about 65 percent cheaper than in London. When costs are recalculated using PPP at the project’s cost year and adjusted for construction-sector productivity, the gap narrows to roughly 35-40 percent. The difference, while still substantial, is far below the headline figure and reflects methodological distortion rather than genuine market variation. Because this inflated “London premium” becomes the baseline for further uplifts, it magnifies bias throughout the subsequent analysis.

4.4. Cost Normalisation and Productivity Adjustment

Another shortcoming across the studies is the failure to normalise costs for productivity or output. Cost per square metre measures inputs but does not capture how effectively space delivers passenger or baggage capacity. Two terminals of identical area may support very different passenger volumes or operational outputs; benchmarking them on a cost-per-square-metre basis therefore rewards low utilisation rather than efficient design. Metrics such as cost per annual passenger capacity or passengers per square metre would provide a more accurate assessment of efficiency.

Effective normalisation requires metrics that relate cost to performance, for example:

- **Cost per annual passenger capacity (Cost/MPPA)** for terminal buildings.
- **Passengers per square metre (pax/m²)** as an indicator of spatial efficiency and design density.
- **Cost per bag throughput (Cost per tph)** for baggage systems.
- **Cost per additional aircraft movement** for runways and aprons.
- **Cost per parking space** for car parks.

Using such denominators shifts the analysis from how much infrastructure was built to what that infrastructure achieves. None of the HAL-commissioned benchmarks apply this logic. Steer's Capex benchmark, for example, remains focused on cost per square metre and does not test whether throughput, design density or system capacity explain cost variation. As a result, differences in productivity remain unexamined.

This omission is particularly relevant in Heathrow's regulatory context, where the Regulated Asset Base model links allowable revenues to the size of the asset base rather than to the efficiency with which that asset delivers capacity. Without normalising for output, benchmarking results may interpret high capital intensity as normal even when it reflects design choices that increase asset size without delivering equivalent functional benefit. A credible normalisation framework would statistically control for both input (size, specification, region) and output (passenger or system capacity) variables, allowing efficiency gaps to be identified and quantified.

4.5. Treatment of Scope and Quality Differences

HAL's commissioned benchmarks claim to adjust for differences in scope, specification, and quality, but the adjustment logic is neither documented nor evidence-based. Projects that include baggage systems, stands, or integrated piers are compared directly to those that

do not. Quality factors are acknowledged but not quantified. As a result, the studies cannot separate *functional need* from *strategic choice*. A sound approach would establish standard scope templates for each asset class and apply bounded adjustments (e.g., $\pm 10\text{--}15\%$) based on documented cost breakdowns.

4.6. Treatment of Assumptions and Uplifts

A recurring methodological weakness across the studies is the stacking of adjustment factors that are not tested for independence. Location, congestion, hub complexity, non-Schengen processing and retail-mix uplifts are treated as separate allowances, even though they reflect many of the same underlying drivers. When these factors are applied sequentially, they risk double counting and inflate comparator results. Each uplift may appear reasonable in isolation, but taken together they materially raise the adjusted values.

A further distortion arises from the interaction between uplift factors and embedded contingencies within comparator cost estimates. Many international project estimates already include allowances for risks such as night-work disruption, constrained site access or interface management. Applying additional “London” or “site-constraint” uplifts on top of these provisions compounds similar risk elements twice. This reduces comparability by overstating the adjustments required to align foreign projects with Heathrow’s conditions.

A related compounding effect appears in KPMG’s aeronautical-revenue model. Comparator asset values are first uplifted to reflect the higher construction prices associated with London. These enlarged asset values are then multiplied by standard return allowances. Because the return is calculated as a proportion of the asset base, the initial uplift feeds directly into the return calculation. This applies the construction-cost differential twice: once through the uplift to the asset value and again through the return derived from that higher base. The result is a systematic magnification of the adjustment that narrows the gap between Heathrow and its comparators.

A particularly visible example of assumption-stacking is the 25 per cent “congestion uplift” applied by Steer to account for site constraints and construction phasing at Heathrow. This factor is introduced as a blanket allowance across all comparator projects without empirical derivation or reference to measured productivity impacts from recent Heathrow works. No evidence is provided that congestion imposes a uniform 25 per cent cost penalty, nor that such a relationship is consistent across trades or project types. Elements of this penalty are also likely to be embedded already within location factors, productivity adjustments or project-level contingencies. Applying an additional fixed 25 per cent uplift therefore double counts existing provisions and does so in a way that disproportionately benefits Heathrow.

Robust benchmarking practice requires each adjustment factor to be evidence-based, independent and tested for interaction effects. A more rigorous approach would identify potential overlaps between location adjustments, risk allowances and operational constraints, and would estimate the marginal contribution of each factor through regression or sensitivity analysis. Only by isolating non-overlapping drivers can the resulting benchmarks be regarded as fair, unbiased and suitable for regulatory decision-making.

4.7. Statistical Testing and Sensitivity Analysis

None of the benchmarking studies provide the formal statistical testing needed to validate the precision of their comparative results. While the KPMG operating-cost work includes a set of regression models, these are not used to test the robustness of the cost adjustments or to establish confidence intervals for the benchmarking outputs. The Steer Capex study presents a simple correlation, but again without testing the significance of the relationship or its explanatory strength. Across the studies, the benchmarking results are presented as point estimates, with no confidence ranges, uncertainty bands or sensitivity analysis. Given the small sample sizes and the number of assumptions applied, the absence of these tests means the apparent precision of the findings cannot be independently verified.

4.8. Data Transparency and Reproducibility

Across the evidence base, data transparency is minimal. No study provides a full data appendix listing project names, cost years, exchange rates, or inflation indices. Analytical steps, such as how “scope adjustments” or “constraint uplifts” are calculated, are described only qualitatively. This lack of documentation prevents independent replication and limits regulatory confidence in the results.

Reproducibility is a cornerstone of any credible benchmarking methodology. Without it, findings cannot be peer-tested or validated. An analyst attempting to reconstruct the Steer or KPMG calculations would find insufficient information to do so.

4.9. Summary of Methodological Weaknesses

Analytical Dimension	Observed Weakness	Consequence / Impact
Inflation and Cost-Year Treatment	Incorrect sequencing of deflators and FX conversion; inconsistent base years	Comparator costs artificially inflated; misaligned price levels across projects
Foreign Exchange and PPP Conversion	Use of nominal exchange rates and unadjusted Arcadis ICC indices	Exaggerates “London premium”; misrepresents real purchasing power; inflates Heathrow’s relative normality
Productivity Normalisation	Reliance on cost per m ² ; no control for delivered output (passengers, baggage tph, ATMs)	Measures inputs not efficiency; ignores throughput and design density
Comparator Selection and Representativeness	Small, curated samples; mix of estimates and out-turns; unclear inclusion criteria	Selection bias (“cherry-picking”); compressed cost range that favours Heathrow
Scope and Quality Control	Unspecified adjustment logic for functional scope and finish levels	Blurs line between functional need and design choice; weakens comparability
Assumptions and Uplifts	Multiple overlapping factors (location, constraint, hub complexity, retail mix); arbitrary 25 % “congestion uplift”	Double counting; arbitrary inflation of comparator costs; benefits Heathrow disproportionately
Contingency and Risk Treatment	Uplifts applied on top of costs already containing contingencies	Compounding of risk allowances; exaggeration of cost parity
Statistical Analysis and Sensitivity Testing	No regression, correlation, or uncertainty analysis	Reported differences lack significance testing; apparent precision unsupported by evidence
Data Transparency and Reproducibility	Missing project-level data, FX sources, and inflation indices	Results cannot be replicated or validated; low regulatory confidence

5. Detailed Study Critiques

This section examines in detail the three main studies that underpin HAL's benchmarking narrative:

1. *Steer (2024) – Capex Benchmark of Terminal Building Projects*
2. *KPMG (2025) – Aeronautical Revenue Benchmarking*
3. *Arcadis (2020) – Capex Efficiency Review (CAP1964a) , commissioned by the CAA as part of H7*

Each report contributes to HAL's position that Heathrow's costs are justified once adjusted for location, congestion and complexity. A technical review of these studies, however, reveals methodological flaws that undermine their reliability. Across all three, the evidence suffers from selective data use, weak normalisation and opaque adjustment processes that cumulatively distort the findings.

5.1. Steer (2024) – Capex Benchmark of Terminal Building Projects

5.1.1. Overview

Steer's Capex Benchmark compares the construction cost of Heathrow's recent and planned terminals with a set of international comparator projects. The study presents results primarily on a cost-per-square-metre basis and applies several uplifts intended to reflect Heathrow-specific conditions. While the approach aims to provide a like-for-like comparison of terminal construction costs, the benchmark is shaped heavily by choices in comparator selection, the absence of output-based normalisation and the stacking of untested adjustments.

5.1.2. Key Methodological Weaknesses

5.1.2.1. *Limited and Selective Sample*

The comparator group consists of about twenty projects, drawn mainly from Europe and North America, with no published criteria for inclusion or exclusion. It combines projects with differing scope, specification and functionality, which further undermines comparability. A number of major international terminals are absent, while several design-stage or pre-Final Investment Decision (pre-FID) estimates are included. The resulting dataset is neither geographically representative nor balanced in terms of project maturity. Reliance on early-stage estimates also introduces optimism bias, which systematically distorts the comparison.

5.1.2.2. *Lack of Productivity Normalisation*

Steer relies on cost per square metre as its primary benchmark metric. This is a construction input measure, not a performance indicator. It ignores passenger throughput, baggage capacity, and other functional outputs that determine the economic productivity of terminal space. A smaller terminal with higher throughput per area can have higher Capex/m² yet be more efficient per passenger. Without adjusting for outputs such as cost per annual passenger capacity (Cost/MPPA) or cost per bag throughput, the analysis conflates design density with cost efficiency.

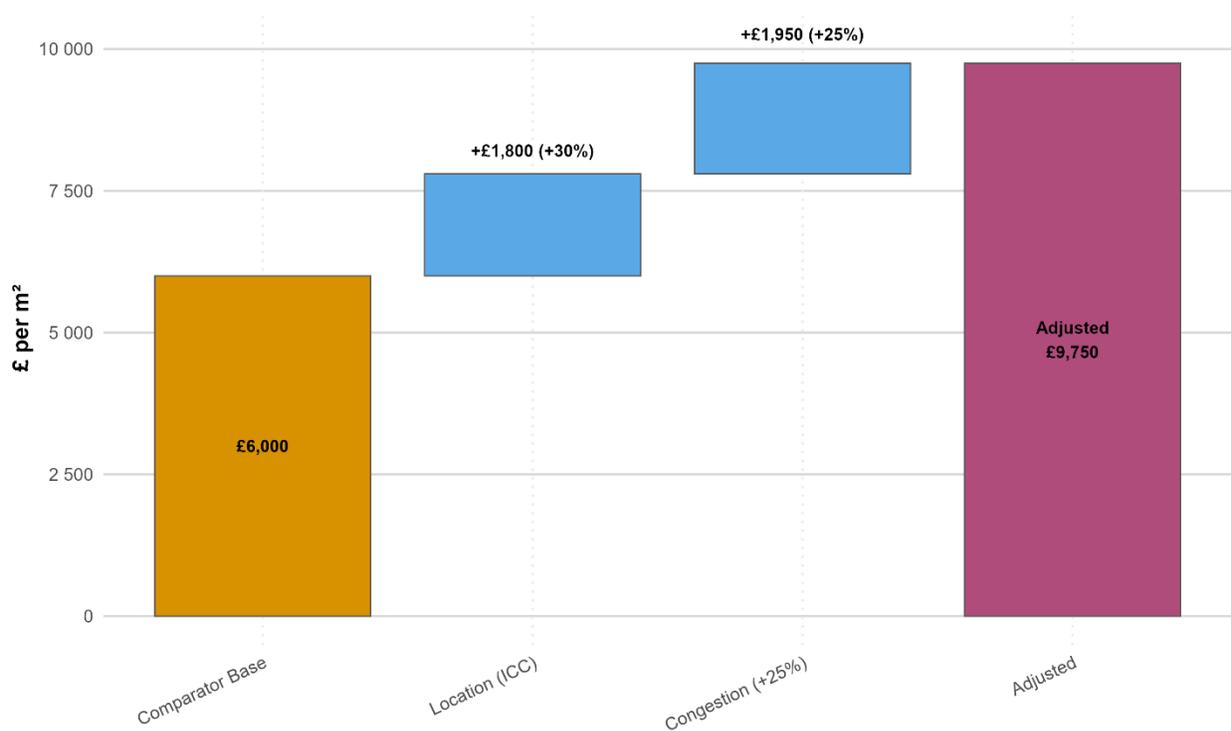
5.1.2.3. *Arbitrary 25 % “Congestion Uplift”*

A flat 25 per cent cost premium is applied to all comparator projects to account for Heathrow’s constrained site and construction phasing. This uplift lacks empirical justification and no evidence is provided to show how the 25 per cent figure was derived. The study does not present measured productivity penalties from recent Heathrow works, such as access restrictions, night working or interface management, that would support the magnitude of the adjustment.

The uplift also reflects assumption stacking. Construction costs and the indices used to adjust them are already correlated with local congestion, land constraints and market intensity. Many international cost indices implicitly capture the effect of congested working environments through higher labour costs, reduced productivity and regional market premia. Applying a further 25 per cent allowance therefore risks double counting congestion-related effects that are already embedded in the base cost levels of both Heathrow and the comparator projects.

The uplift is applied uniformly across all comparators despite substantial variation in site conditions. Airports with materially less congestion receive the same allowance as those with similar or greater constraints, which reduces the discriminating power of the benchmark. The cumulative impact is to inflate comparator costs asymmetrically and narrow the apparent gap between Heathrow and the comparator group.

Figure 2: Cumulative Adjustments Applied to Comparator Terminal Costs



Economies of Scale and Complexity

The study assumes a linear relationship between cost and size and reports no statistically significant scale effects. This finding is unsurprising because the analysis does not control for other variables that co-vary with terminal size, such as scope, system integration, design standard or passenger capacity. Without such controls, any underlying scale effect is masked by cross-project heterogeneity.

In practice, cost per unit of capacity often falls with size as fixed costs are spread over a larger base, before rising again as integration and complexity increase. Ignoring this non-linear pattern and treating all terminals as directly comparable conflates structural drivers and may allow Heathrow's large, multi-system terminals to appear more cost-competitive than they are.

5.1.2.4. Scope and Quality Control

Steer states that adjustments were made for scope and quality differences, yet no calculation or decomposition is shown. Projects including baggage systems, piers, or extensive MEP networks are benchmarked against simpler buildings without adjustment. Differences in interior specification, retail integration, or façade systems are acknowledged but not quantified. The absence of explicit scope templates means functional and discretionary costs are merged, weakening any inference about comparative efficiency.

5.1.2.5. *Inconsistent Inflation and Cost-Year Treatment*

Steer converts foreign project costs to GBP at base-year exchange rates, inflates them using local indices, and then adjusts them to UK construction prices. This sequencing is conceptually flawed. By applying the UK price-level adjustment after local inflation, the method compounds both local and UK cost growth, effectively applying two inflation systems to the same cost base. This raises comparator costs relative to Heathrow.

A consistent approach would convert foreign costs to UK-equivalent price levels at the project's cost year and then apply UK inflation to the comparison year. This ensures that all projects are expressed in real UK price terms before temporal inflation is applied.

Correcting the sequencing would generally reduce non-UK comparator costs, widening rather than narrowing the cost gap. This issue shows the need for consistent cost-normalisation.

5.1.2.6. *Lack of Statistical or Sensitivity Testing*

No regression or variance analysis is presented. Results are reported as single averages or bars without confidence ranges. Given the small sample, this omission is critical because the statistical uncertainty might exceed the reported differences. The claim that Heathrow sits “within the range” is therefore not statistically demonstrated but visually inferred from a narrow sample. Sensitivity tests using alternative inflation or uplift assumptions could shift Heathrow toward another position on the distribution.

5.1.3. Implications

Taken together, these weaknesses undermine the conclusion that Heathrow's costs are typical. Inflation mis-sequencing and double-counted uplifts alone are sufficient to move Heathrow's adjusted cost per m² from the apparent median to the P80–P90 range of a normalised dataset. The absence of output-based and scale-adjusted metrics further conceals inefficiencies in design density and system integration. While the precise effect remains to be quantified, the combined direction of these biases appears to narrow rather than widen the apparent cost gap between Heathrow and its peers.

5.1.4. Recommended Corrections

Issue	Recommendation
Inflation and FX treatment	Convert costs to GBP using price-level adjustments at the project's cost year, then rebase to 2024 GBP PPP using UK construction deflators.
Congestion uplift	Replace arbitrary 25 per cent factor with evidence-based productivity analysis (e.g., access restrictions, night-shift ratios).
Metric selection	Supplement cost per m ² with cost/MPPA, passengers per m ² and cost per bag throughput.
Scale and complexity	Apply multivariate regression to test non-linear cost-size relationships, controlling for scope and specification.
Scope differences	Develop standardised scope templates distinguishing core and non-core elements (baggage, MEP, retail).
Statistical validity	Present results as distributions or percentile ranges and run sensitivity tests on key assumptions.

5.2. KPMG (Aeronautical Revenue Benchmarking, 2025)

5.2.1. Overview

KPMG argues that Heathrow's aeronautical revenue per passenger is reasonable once adjusted for local cost conditions, hub complexity and regulatory structure. The analysis is ambitious but relies heavily on assumed adjustments rather than empirical testing.

5.2.2. Key Methodological Weaknesses

5.2.2.1. Assumption Stacking and Double Counting

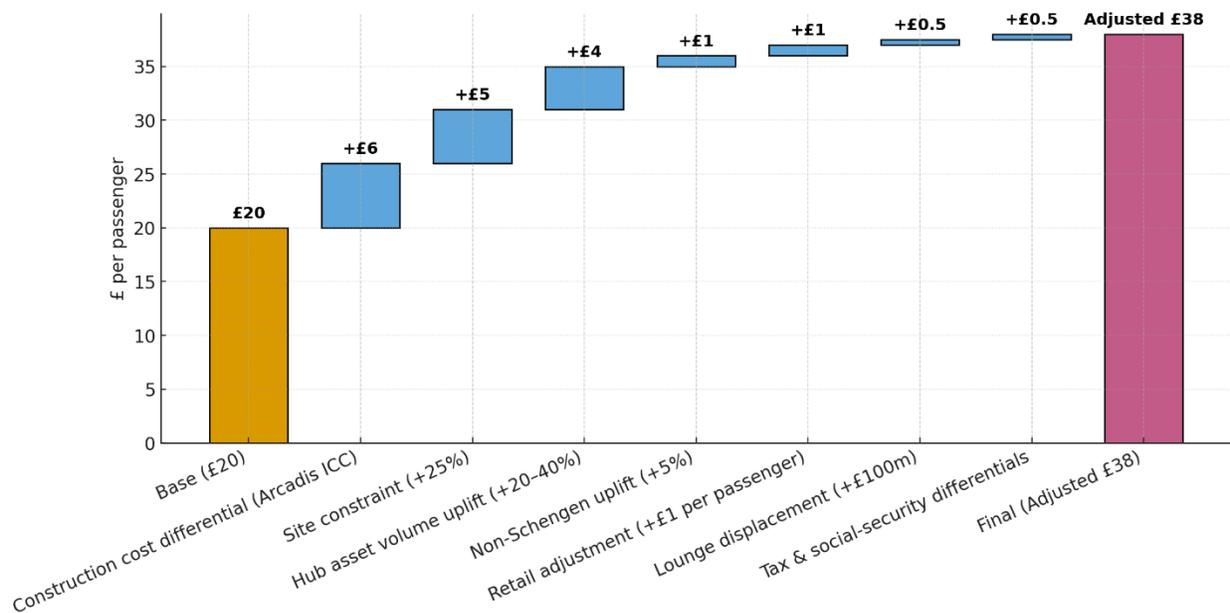
KPMG's benchmarking model normalises comparator airports to Heathrow by layering a sequence of upward adjustments designed to reflect Heathrow's unique operating and market conditions. While some adjustments may appear intuitive when viewed in isolation, many lack empirical justification and several overlap in what they measure. Applied sequentially, these adjustments raise comparator costs by nearly ninety per cent, transforming Heathrow from a statistical outlier into an apparent average performer. The cumulative structure introduces dependency risks and applies multiple uplifts to the same underlying characteristics.

The model's adjustment stack draws on both core and illustrative elements, including:

- **Construction cost differential (Arcadis ICC)** – London construction cost premium affecting depreciation and returns.
- **Site constraint (+25 %)** – illustrative assumption that Heathrow’s restricted footprint increases asset intensity.
- **Hub asset volume uplift (+20–40 %)** – higher fixed asset requirements for connecting hubs versus point-to-point airports.
- **Non-Schengen uplift (+5 %)** – higher screening and processing costs.
- **Retail adjustment (+£1 per passenger)** – effect of removing VAT-free shopping on retail yield.
- **Lounge displacement (+£100 million)** – opportunity cost of premium lounges reducing retail space.
- **Tax and social-security differentials** – higher UK property taxes and employer charges.

Each layer is applied sequentially, raising comparator values before the next is introduced. As Figure 3 illustrates, the cumulative effect increases the notional “justified” revenue per passenger from roughly £20 to £38. What appears to be incremental adjustment therefore results in a near-doubling of the base value and is driven by assumed uplifts rather than observed performance data.

Figure 3: Sequential Adjustments in Heathrow’s Revenue Benchmarking Model



Many of these uplifts capture the same claimed structural features of Heathrow's system. Infrastructure, hub-operation and premium-mix factors all stem from the same root condition: a wide-body, long-haul network. Claimed site constraints and regulatory conditions are treated as distinct factors even though they operate through similar mechanisms, such as reduced productivity and constrained access. Rail-infrastructure costs correlate closely with hub scale and passenger volumes, making them functionally dependent on the hub-scale uplift.

Through this stacking, the model presents Heathrow as expensive for a series of seemingly separate reasons, but those reasons are often repetitive. By systematically layering interdependent premiums, the analysis raises comparator costs until they mirror Heathrow's own cost structure. The result is a framework that demonstrates internal consistency rather than empirical independence. In practice, the model validates its own assumptions rather than testing Heathrow's efficiency.

5.2.2.2. Misuse of Cost Indices and FX Conversion

KPMG adopts the Arcadis ICC and nominal exchange rates to estimate relative construction costs. This combination exaggerates London's cost level for the same reasons described earlier in Section 4.5. The study asserts that Madrid's construction costs are around 65 % lower than London's, a difference inconsistent with other datasets and realistic productivity adjustments. When recalculated using PPP conversion and construction-sector productivity indices, the true differential falls to approximately 35–40 %. This methodological inflation spreads through the model, inflating Heathrow's relative normality.

5.2.2.3. Unverified Elasticities and Parameters

Key behavioural parameters, such as opex elasticity to scale, lounge displacement impacts, and retail elasticity, are introduced without empirical estimation. The elasticities appear to be assumed rather than statistically derived from multi-airport datasets. Without validation, these parameters act as levers to tune the model toward a desired outcome.

5.2.2.4. Circular Logic in Asset-Valuation Adjustments

In addition to assumption stacking, the asset-valuation adjustments create a separate problem of circular logic. Comparator asset values are first increased using London-specific uplifts for construction costs and site constraints. These inflated asset values are then passed through return ratios that already embed London-specific financing costs, tax conditions and risk premia. The same set of Heathrow-specific conditions therefore enters the model twice: once through the uplifted asset base and again through the return calculations. The resulting benchmark reflects a self-reinforcing framework in which Heathrow's own cost environment determines both the comparator uplift and the return allowed on that uplift, rather than providing an independent test of Heathrow's efficiency.

5.2.3. Implications

It may be argued that the adjustments are intended to ensure comparability by placing other airports into Heathrow’s operating environment and assessing whether they would face similar charges under equivalent conditions. However, the analysis does not achieve this in practice. The uplifts intended to construct these “like-for-like” conditions are not supported by evidence, are poorly specified, or overlap in what they measure. Several describe the same claimed structural features yet are applied as separate adjustments, while others rely on illustrative percentages rather than measured impacts. As a result, the exercise does not provide a credible test of how other airports would perform under Heathrow’s conditions. Instead, it mechanically elevates comparator values through assumption stacking and circular valuation logic, producing results shaped more by the structure of the model than by genuine differences in operating conditions or efficiency.

5.2.4. Recommended Corrections

Issue	Recommendation
Index and FX misuse	Use PPP-adjusted rates and construction productivity data
Parameter estimation	Derive elasticities empirically from multi-airport panel data.
Double counting	Test interdependence of location and complexity factors; apply only independent effects.
Scenario testing	Run low/base/high cases to show sensitivity of Heathrow’s position.

5.3. Arcadis (Capex Efficiency Review, 2020)

5.3.1. Overview

Arcadis was commissioned by the CAA to review the efficiency of ten Heathrow projects delivered during the Q6 period. This review is distinct from the “Heathrow Reimagined” work undertaken by Arcadis for HAL and should not be interpreted as part of that programme. The study concludes that most projects were delivered efficiently, attributing cost increases primarily to external factors such as regulatory changes and scope evolution. While the report offers useful descriptive insights, its qualitative approach limits the extent to which firm conclusions can be drawn.

5.3.2. Key Methodological Weaknesses

5.3.2.1. *Limited and Non-Representative Sample*

The ten projects were nominated by HAL and cover a narrow cross-section of the capital portfolio. Several high-value or schedule-challenged schemes were excluded. The sample therefore cannot be generalised to Heathrow's overall capital programme.

5.3.2.2. *Reliance on HAL-Supplied Information*

Arcadis explicitly states that much of the evidence came from HAL's own documentation and interviews. The absence of independent cost verification or external benchmarking undermines objectivity and introduces confirmation bias.

5.3.2.3. *Absence of Quantitative Counterfactuals*

The analysis lacks a formal baseline against which to judge efficiency. There is no reference class or external comparator showing what a prudent operator would have spent under similar conditions. Without such a counterfactual, "efficient" and "inefficient" become relative terms defined within HAL's own data context.

5.3.2.4. *Narrative Rather than Analytical Assessment*

Findings are descriptive using categories such as "efficient," "good practice," or "external factors", but without quantitative substantiation. No attempt is made to measure cost or schedule variance relative to a benchmark distribution. The result is a narrative of competence rather than a statistical test of efficiency.

5.3.2.5. *Omission of Design Efficiency*

The assessment does not consider design efficiency, namely whether the projects were required at their delivered scale or specification. It does not test whether alternative design choices could have achieved comparable functionality at lower cost. Excluding this dimension further limits the ability to distinguish efficient delivery from over-design or scope inflation.

5.3.3. Implications

Arcadis' work provides useful qualitative insights into project management practices but cannot serve as quantitative evidence of efficiency. The small, self-selected sample and lack of measurable counterfactuals mean its conclusions should be treated as illustrative, not evidential. When combined with the Steer and KPMG studies, it completes a circular narrative: each report validates Heathrow's position using overlapping data and assumptions, none of which are independently verified.

5.3.4. Recommended Corrections

Issue	Recommendation
Sample bias	Expand dataset to include full range of Q6 projects; apply value weighting.
Data independence	Require independent data verification and third-party validation.
Counterfactual definition	Establish benchmark using reference class or external comparator projects.
Quantification	Express results in cost/schedule variance terms rather than qualitative labels.

6. Findings and Implications

Across the three commissioned studies a consistent analytical pattern emerges. Each uses limited and partly opaque data, apply adjustment layers that are not always independent, and presents conclusions that appear confirmatory rather than interrogative. While these issues might individually be attributed to data constraints or professional judgement, taken together they form a body of evidence whose internal consistency reflects shared assumptions rather than empirical robustness.

The common features are clear. All studies rely on small comparator samples without transparent inclusion criteria, which makes their results difficult to generalise. Inflation and exchange rate adjustments are often sequenced inconsistently, with both local and United Kingdom deflators applied to the same costs, which reduces comparability between foreign projects and Heathrow. Several adjustments, including those for location, site constraint, hub complexity and premium passenger mix, are applied cumulatively even though they describe related underlying conditions. None of the reports applies formal statistical testing or uncertainty analysis, so the apparent precision of the results cannot be verified. Finally, data transparency is limited. Key project information, exchange rate dates and index sources are not disclosed, which prevents independent replication.

These methodological features reduce the apparent differences between Heathrow and its comparators. Whether or not this effect is intentional, the structure of the analyses tends to make Heathrow's costs appear more typical than the underlying data would suggest. The evidence base therefore provides a narrative rather than a statistically verified assessment.

The relationship between the studies further strengthens this weakness. Each draws on the same underlying indices and, in some cases, cites the others as confirmation. Steer's capital expenditure benchmark uses the Arcadis construction index that also supports the KPMG

revenue model, while KPMG refers to Steer's findings to support its claim of cost parity. Arcadis, in turn, evaluates Heathrow's delivery efficiency using similar underlying assumptions. This interdependence gives the impression of independent validation, when in reality the studies reference one another and amplify the same underlying assumptions.

The implications for regulatory confidence are significant because the analyses are neither transparent nor independently reproducible. They should therefore be interpreted with caution when assessing efficiency. The Civil Aviation Authority and other stakeholders would require an independent benchmark to determine Heathrow's position within international cost distributions. Until such benchmarking is available, the current studies cannot be relied upon as evidence of Heathrow's relative efficiency.

7. Incentive Misalignment and Regulatory Risk

These methodological issues must be viewed within the regulatory framework that shapes Heathrow's incentives. Heathrow operates under the Regulated Asset Base model, which links allowed revenues and investor returns directly to the size of the airport's asset base. Within this framework, additional capital expenditure increases the RAB and therefore future returns, regardless of whether the investment delivers proportional efficiency or service improvements.

This mechanism creates a structural tension between investor returns and consumer interests. The model was designed to encourage stable long-term investment, but over time it has also produced weak incentives for cost discipline. When the value of the asset base itself determines allowed revenue, the focus of management effort naturally shifts toward capital expansion rather than efficiency optimisation. Cost control becomes a secondary concern so long as spending can be presented as necessary or benchmarked as normal.

This structural misalignment creates several forms of regulatory risk. First, it allows information asymmetry to persist: Heathrow controls most of the data needed to assess efficiency and can shape the presentation of its own performance. Second, it blurs the distinction between legitimate investment and gold-plating. The regulator may approve expenditure that appears reasonable within the available evidence but which, in a more competitive or transparent environment, would have been challenged. Third, by linking returns to capital growth, the framework can perpetuate high passenger charges even where service quality or productivity lags behind international peers.

The CAA has sought to manage these risks through efficiency targets and ex-post adjustments, yet such mechanisms remain constrained by the quality of underlying evidence. Without independent benchmarks or open data, the regulator's ability to test

whether Heathrow's costs are proportionate is limited. As Heathrow enters the H8 period with an investment plan approaching £10 billion, this asymmetry becomes more material: small shifts in assumed efficiency translate into large changes in allowable revenue and consumer charges.

Rebalancing these incentives does not require the RAB model to be abandoned, but it does require its counterweights to be strengthened. Greater transparency of project level data, independent benchmarking of cost and productivity, and a clearer separation between efficiency and expansion would all enhance regulatory confidence. The aim is not to discourage investment but to ensure that capital spending delivers genuine public value. Capital projects should be shown to be efficient, necessary and well managed within a framework that rewards performance rather than scale.