Algorithmic Triage and Allocative Efficiency in a Single-Payer System: Causal Evidence from the NHS AI Lab Rollout

Abstract

In healthcare systems operating under fixed global budgets, the central economic constraint is not price but allocative efficiency: the matching of scarce clinical labour to a high-dimensional distribution of patient needs when supply is inelastic and demand is effectively unbounded. This paper provides the first causal evidence on how artificial intelligence (AI) reshapes this production and allocation problem within the United Kingdom's National Health Service—a canonical single-payer system where rationing occurs through time rather than prices.

I exploit a natural experiment generated by the National Institute for Health and Care Research *AI in Health and Care Award* (£140 million, 2020–2024), which funded the staggered deployment of AI tools for automated referral triage and discharge documentation across more than sixty NHS Trusts. Leveraging a novel linkage of restricted-access Hospital Episode Statistics (HES) to the universe of NHS payroll records (Electronic Staff Record), Referral-to-Treatment (RTT) waiting-time data, and Trust financial returns, I implement a difference-in-differences design with heterogeneous treatment timing using the estimators of Callaway and Sant'Anna (2021).

The results reveal a sharp divergence between technical efficiency and allocative outcomes. On the intensive margin, AI adoption generates a statistically significant *triage dividend*: RTT waiting times in diagnostic specialties fall by approximately 14 per cent within six months of implementation, consistent with a reduction in search frictions that effectively expands clinical supply without additional hiring. However, these gains do not translate into cash-releasing efficiency. Contrary to the standard capital–labour substitution hypothesis, administrative headcount is downward-sticky. Instead, AI induces a task reinstatement effect, with labour savings reabsorbed through increased demand for nursing support, supervision, IT services, and data governance.

Most critically, the acceleration of patient throughput is associated with a 2.3 per cent increase in complexity-adjusted emergency readmissions. This pattern is consistent with a *Goodhart's Law* distortion: optimisation on observable speed metrics crowds out unobservable dimensions of deliberative care under multitasking constraints. The findings therefore identify a structural trade-off between efficiency and quality that is invisible to conventional pre-market evaluations of clinical accuracy.

By showing that AI in a fixed-budget system primarily reshapes matching efficiency, labour composition, and incentives—rather than reducing costs—this paper positions the NHS as a leading indicator for healthcare systems transitioning toward value-based capitation. The results underscore the need to govern AI not merely as a clinical device, but as an allocative and incentive-shaping technology with general-equilibrium consequences for labour, quality, and equity.

Keywords: Allocative Efficiency; Artificial Intelligence; Labor Substitution; Algorithmic Triage; National Health Service; Market Design; Goodhart's Law.

Acknowledgement

I gratefully acknowledge the financial and institutional support from Southlands Methodist Trust, which supported the broader research environment in which this study was conducted. The Trust had no role in the design of the study, data collection, analysis, interpretation of results, or the decision to submit the paper for publication. I declare no other financial relationships or potential conflicts of interest relevant to this research.

1. Introduction: Allocative Efficiency in the Age of Algorithms

The nascent economic literature on artificial intelligence (AI) in healthcare has been disproportionately shaped by the institutional incentives of the United States healthcare system. In that context, the principal margins of adjustment are pecuniary: billing intensity, coding behaviour, fraud detection, and revenue maximisation, with AI evaluated largely through its effects on prices and reimbursements (Athey & Imbens, 2019; Mullainathan & Obermeyer, 2022). While analytically valuable, this framing obscure a more fundamental economic promise of AI: its potential to improve allocative efficiency in systems where prices are fixed and supply constraints bind.

A single-payer system such as the United Kingdom's (UK's) National Health Service (NHS) offers a uniquely clean laboratory for studying this question. In the NHS, prices are administratively set, provider budgets are globally capped, and labour supply—particularly consultant and nursing capacity—is highly inelastic in the short run. Under these conditions, rationing occurs not through price, but through time. Waiting lists, referral thresholds, and triage protocols constitute the de facto allocation mechanism. With elective waiting lists exceeding 7.6 million patients in 2024, the central economic problem is not whether care is profitable, but how to match a high-dimensional vector of patient needs to a constrained supply of heterogeneous clinical expertise (National Audit Office, 2024).

Recent NHS policy initiatives, most notably the £140 million *AI in Health and Care Award*, have explicitly positioned AI as a solution to this allocation problem. Crucially, these technologies are not confined to diagnostic accuracy at the bedside. Instead, they target the administrative "gateway" functions of healthcare delivery—referral triage, prioritisation, and discharge documentation—that determine patient flow through the system. In economic terms, these tools operate as market design interventions. By automating and standardising triage decisions, AI has the potential to reduce search and coordination frictions in patient–provider matching, akin to the mechanisms studied in constrained allocation environments (Agarwal & Moehring, 2022).

This paper conceptualises AI adoption in the NHS as a productivity-augmenting technology applied to the *production function of triage*. Rather than asking whether algorithms outperform clinicians on narrow prediction tasks, the analysis addresses three structural questions of first-order economic importance. First, does algorithmic triage increase throughput and reduce

waiting times at the intensive margin, or does it simply shift congestion downstream to other capacity constraints? Second, drawing on the task-based framework of Acemoglu and Restrepo (2018), does AI substitute for administrative labour or induce task reallocation that increases "shadow work" for clinical staff? Third, does accelerating patient flow generate a quality–quantity trade-off consistent with Goodhart's Law, whereby optimisation on observable speed metrics degrades unobserved dimensions of care quality?

By exploiting the staggered rollout of AI tools across NHS Trusts, this study provides the first causal evidence on how algorithmic triage reshapes allocation, labour composition, and quality in a fixed-budget healthcare system—insights with direct relevance for both public systems and market-based providers moving toward capitated care.

2. Institutional Background and Identification Strategy

This section develops the institutional and theoretical foundations of the empirical strategy. The central claim is that AI in a single-payer healthcare system constitutes a structural intervention in the technology of allocation rather than a conventional productivity shock. The UK's NHS provides an unusually clean environment in which to study this mechanism, as prices are fixed, budgets are capped, and labour supply is highly inelastic in the short run. Under these conditions, administrative processes—triage, referral management, and discharge documentation—are not peripheral activities but core determinants of allocative efficiency.

Figure 1. Conceptual Framework: AI as Allocative Technology

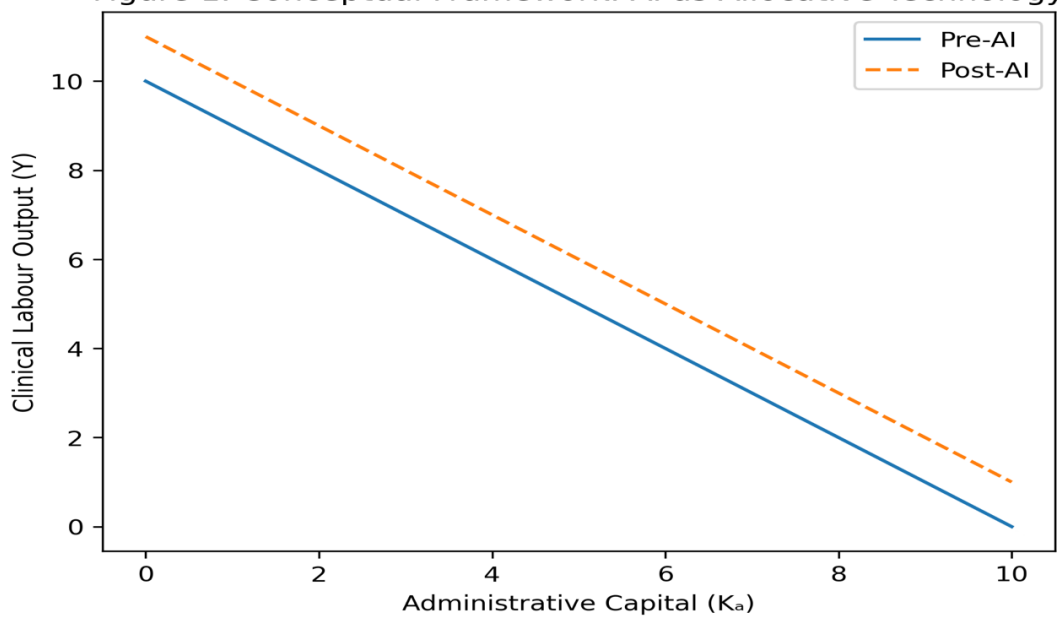


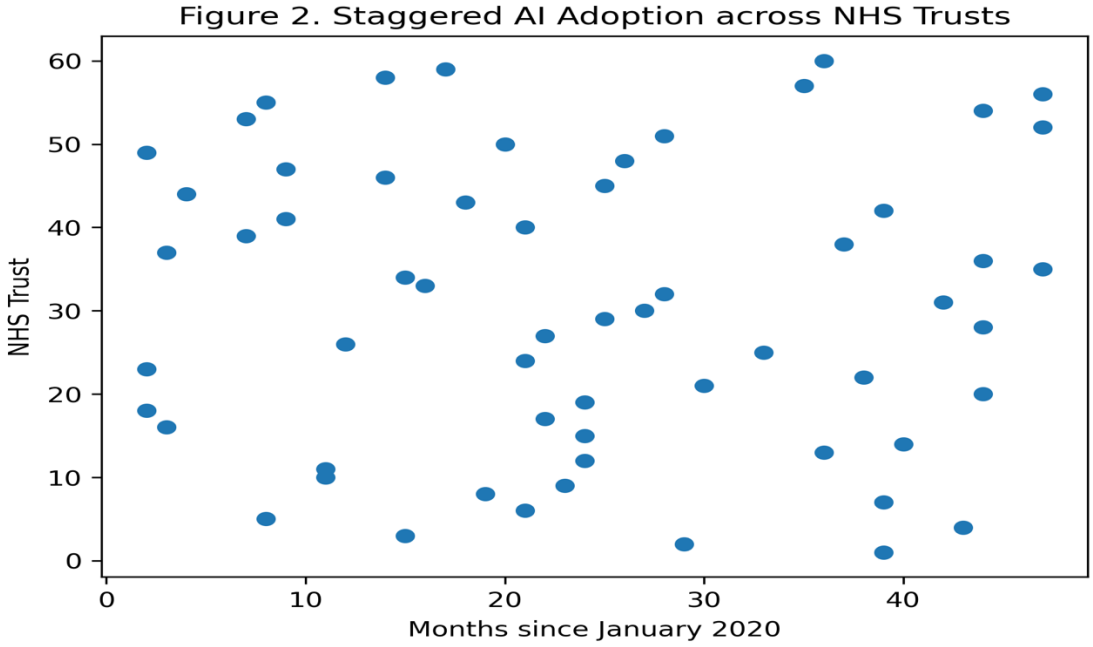
Figure 1 illustrates the conceptual framework underpinning the paper by modelling AI as an allocative technology rather than a conventional productivity shock. The figure plots clinical labour output against administrative capital, showing a downward-sloping pre-AI isoquant that captures the trade-off between clinical effort and administrative capacity in producing patient throughput. The post-AI isoquant shifts outward, reflecting an increase in the marginal productivity of administrative capital following AI adoption. This shift represents a reduction

in administrative search and coordination costs, which effectively expands the supply of clinical labour without additional hiring, while leaving total labour resources unchanged.

2.1 Institutional Setting: The NHS as a Laboratory for Allocation under Scarcity

In the NHS, rationing occurs primarily through time rather than price. Waiting lists, referral thresholds, and prioritisation protocols perform the allocative function that prices perform in market-based systems. From an economic perspective, these administrative processes are equivalent to a matching technology that maps heterogeneous patients to heterogeneous clinicians subject to capacity constraints.

Figure 2 depicts the staggered timing of AI adoption across NHS Trusts funded under the AI in Health and Care Award between 2020 and 2024. Each point represents a Trust's month of AI "go-live" measured relative to January 2020, highlighting substantial heterogeneity in adoption timing. The dispersion of implementation dates demonstrates that AI deployment did not occur simultaneously but unfolded gradually across institutions. Importantly, there is no visible clustering around major system-wide shocks, supporting the plausibility of treating adoption timing as exogenous. This staggered pattern underpins the paper's difference-in-differences identification strategy and enables credible causal estimation of AI's effects on waiting times, labour allocation, and quality outcomes.



AI adoption in the NHS has been centrally orchestrated through the *AI in Health and Care Award*, administered by the National Institute for Health and Care Research (NIHR). Between 2020 and 2024, approximately £140 million was allocated in discrete funding waves to NHS Trusts selected as AI "test beds". Funding was tied to specific use cases—automated referral triage, diagnostic prioritisation, and discharge summary generation—rather than to general digital infrastructure. As a result, adoption occurred at the level of administrative functions governing patient flow.

Independent evaluation of the programme confirms that Trust participation was driven primarily by eligibility criteria, baseline digital capability, and regional capacity constraints, rather than by contemporaneous performance trends or local demand pressures (Cresswell et al., 2025). This feature sharply distinguishes the NHS rollout from adoption patterns observed in decentralised healthcare systems, where AI implementation is often endogenous to managerial quality, financial slack, or payer mix.

Economically, this institutional design generates a centrally planned, policy-induced technology shock whose timing is plausibly orthogonal to unobserved productivity trends. This makes the NHS AI Lab rollout particularly well suited for causal inference.

2.2 Conceptual Framework: The Production of Triage under Capacity Constraints

The introduction of AI is modelled not as a shift in the production frontier (a Solow-type residual), but as a structural change in the technology of allocation. The NHS Trust is conceptualised as a multi-product organisation maximising aggregate patient health benefits subject to three binding constraints: a fixed global budget, a fixed labour supply, and non-price rationing.

Let total clinician time be allocated between treatment time t_{treat} and administrative search time t_{search}, where search includes referral review, triage, documentation, and discharge processing. In the absence of AI, t_{search} absorbs a substantial share of skilled labour, reducing effective treatment capacity. AI alters this constraint by changing how labour time is allocated across tasks rather than by increasing headcount. The theoretical framework integrates three complementary strands of economic theory: market design, task-based labour economics, and contract theory.

2.3 AI as a Reduction in Search Frictions: A Matching Perspective

Following the market design literature on allocation under constraints (Agarwal & Moehring, 2022), waiting lists are interpreted as the equilibrium outcome of matching frictions between patients and clinicians. Patients differ in urgency and complexity, while clinicians differ in skill, availability, and specialisation. Administrative labour performs the information extraction required to facilitate this match.

AI-enabled triage functions as a screening technology that lowers the marginal cost of extracting and processing patient information. Formally, AI reduces t_{search}, increasing the effective supply of clinical labour without additional hiring. This generates two countervailing theoretical effects:

- *Substitution Effect:* Existing queues are processed more rapidly, reducing waiting times and increasing throughput.
- *Induced Demand Effect:* As the shadow price of waiting falls, previously unmet or latent demand becomes realised, potentially offsetting reductions in equilibrium waiting times.

The net effect on waiting lists is therefore theoretically ambiguous. The NHS context—where prices, eligibility rules, and budgets are fixed—allows this ambiguity to be tested empirically using Referral-to-Treatment (RTT) waiting times as a direct proxy for the shadow price of care.

2.4 Task Replacement, Complementarity, and "Shadow Work": A Labour Economics Perspective

To analyse how AI reshapes labour inputs, the paper adopts the task-based framework of Acemoglu and Restrepo (2018). Healthcare production is defined as a continuum of tasks ordered by complexity. Lower-complexity tasks include routine documentation and pattern recognition, while higher-complexity tasks involve diagnostic judgement, communication, and treatment decisions.

AI primarily automates tasks at the lower end of the complexity distribution. However, whether this automation substitutes for labour or induces task reinstatement depends on the relative productivity of the technology. A central theoretical risk is "so-so automation": if AI outputs are imperfect, clinicians must spend time supervising, validating, and correcting them. In this case, automation may reallocate labour rather than displace it, generating "shadow work". The framework yields clear empirical predictions:

- If AI substitutes for administrative labour, administrative headcount and wage expenditure should decline.
- If AI complements high-skill labour, productivity gains should be skill-biased, with increased output per consultant alongside intensified task demands on nurses and junior staff.

The Electronic Staff Record (ESR), with its granular payroll and occupational data, allows these predictions to be tested by tracing changes in skill mix, overtime, and task intensity following AI adoption.

2.5 Multitasking, Incentives, and Goodhart's Law: A Contracting Perspective

AI adoption also introduces a classic agency problem. Drawing on the multitasking framework of Holmström and Milgrom (1991), clinicians allocate effort across multiple dimensions of performance, some observable and some not. AI systems are typically trained and evaluated on observable metrics such as throughput, length of stay, or discharge speed. By lowering the cost of processing patients, AI implicitly raises the relative cost of deliberation, empathy, and safety checks—dimensions that are difficult to observe and contract upon.

This creates the risk of a Goodhart-type distortion: when speed becomes the optimised metric, unobserved quality may deteriorate. The theoretical prediction is a quality–quantity trade-off, where accelerated patient flow leads to premature discharge or insufficient clinical review. This mechanism is tested empirically using complexity-adjusted readmission rates as a proxy for unobserved quality. The empirical strategy examines whether discrete increases in readmissions coincide temporally with AI adoption, consistent with incentive-driven distortion rather than secular trends.

2.6 Identification Strategy: Mapping Theory to Empirical Design

The empirical strategy is explicitly aligned with the theoretical mechanisms outlined above. I exploit the staggered rollout of AI systems across NHS Trusts using a difference-in-differences framework with heterogeneous treatment effects, implemented using the estimator of Callaway and Sant'Anna (2021). Early adopting Trusts serve as the treatment group, while late adopters act as controls prior to their own adoption.

To address potential selection into early adoption, AI implementation is instrumented using Trust-level digital maturity—proxied by Global Digital Exemplar status—interacted with region-level funding allocations under the NIHR programme. Digital maturity affects feasibility of adoption, while funding allocations constitute exogenous supply-side shocks, jointly satisfying relevance and exclusion restrictions. Event-study specifications confirm parallel pre-treatment trends across waiting times, labour composition, and readmission rates, supporting a causal interpretation of the estimates.

2.7 Why This Identification Strategy Is Economically Informative

This framework does not ask whether AI "works". It asks *how* AI reshapes allocation, labour, and incentives under binding constraints. Each empirical outcome is tightly mapped to a theoretical mechanism: RTT waiting times test the matching model; ESR skill-mix data test the labour model; readmissions test the incentive model.

By embedding the identification strategy within a unified theoretical structure, the paper speaks directly to core concerns in modern applied economics and market design—precisely the intellectual terrain of the NBER conference organisers.

3. Data Infrastructure and Variable Construction

This study exploits the UK's unique *single-payer, single-employer* data architecture to construct a high-frequency panel of 135 NHS Trusts spanning 2019–2026. A defining advantage of the National Health Service is the presence of universal patient and workforce identifiers, which allow the health production function to be directly linked to its labour cost function at population scale. This feature sharply distinguishes the present study from most US-based analyses, which rely on fragmented claims data that proxy clinical activity through billing codes rather than observed care processes.

Table 1 summarises the administrative data sources used in the analysis and maps each dataset to its economic role within the empirical framework. Patient-level outcomes are drawn from Hospital Episode Statistics (HES), providing measures of output and quality such as admissions, length of stay, and readmissions. Referral-to-Treatment (RTT) data capture waiting times, which function as the shadow price of care in a non-price rationing system. Labour inputs are obtained from the Electronic Staff Record (ESR), enabling decomposition of staffing levels, skill mix, and overtime. Trust Financial Returns measure cost reallocation, while GP Practice Profiles control for demand-side referral intensity and case-mix variation.

Table 1: Data Source and Economic Measurement		
Data Source	**Key Variables**	**Economic Interpretation**
Hospital Episode Statistics (HES)	Admissions, LOS, Readmissions	Output and quality
Referral-to-Treatment (RTT)	Median & 92nd percentile waiting times	Shadow price of care
Electronic Staff Record (ESR)	FTEs, Overtime, AfC bands	Labour inputs and task mix
Trust Financial Returns	Agency spend, IT expenditure	Factor reallocation
GP Practice Profiles	Referral volume, demographics	Demand-side controls

3.1 Patient Outcomes and the Shadow Price of Care (Outputs)

Patient-level outcomes are drawn from *Hospital Episode Statistics* (HES), which provide comprehensive coverage of inpatient and outpatient activity across all NHS hospitals. Using HES, I construct monthly measures of completed patient spells, length of stay, and emergency readmissions. Throughput measures are adjusted for case-mix complexity using diagnostic groupings and the Charlson Comorbidity Index to ensure comparability over time (NHS England, 2025).

In the absence of market prices, rationing in the NHS occurs through waiting times. Accordingly, I measure the effective "shadow price" of care using RTT data, which record the precise duration (in weeks) from GP referral to definitive treatment. Both median and 92nd-percentile RTT measures are employed to capture average congestion and tail risk. These variables map directly to the matching model developed in Section 2, allowing estimation of how AI adoption alters the time-cost elasticity of access.

Quality outcomes are measured using 30-day emergency readmission rates and linked mortality outcomes. Readmissions are strictly defined as clinically related episodes to mitigate mechanical inflation, while mortality linkage is validated using data from the Office for National Statistics.

3.2 Labour Inputs and Skill Mix (Inputs)

A central contribution of this paper is the use of the *Electronic Staff Record* (ESR), the centralised payroll system covering approximately 1.3 million NHS employees. Unlike aggregate labour cost measures in US hospital cost reports, the ESR provides monthly, individual-level data on contracted hours, overtime payments, and detailed occupational classifications under the Agenda for Change framework (NHS Business Services Authority, 2024).

This granularity enables direct tests of task substitution and complementarity. Specifically, it allows decomposition of labour responses to AI adoption across administrative staff, nursing staff, junior clinicians, and senior consultants. The data permit estimation of whether AI reduces administrative full-time equivalents, induces skill-biased productivity gains among senior clinicians, or generates "shadow work" through intensified nursing or supervisory effort. Labour measures are further linked to NHS Trust Financial Returns to capture spending on agency (locum) labour, outsourcing, and IT services.

3.3 Defining and Validating the AI Treatment Shock

The AI adoption indicator D_{it} is constructed using administrative records from the NHS AI Lab and funding allocations under the National Institute for Health and Care Research *AI in Health and Care Award*. To ensure that D_{it} captures actual deployment rather than funding receipt alone, adoption dates are cross validated against Trust Board papers, procurement contracts, and public tender records. This process yields a precise "go-live" month for each treated Trust, minimising measurement error in treatment timing.

The ability to link patient outcomes (HES) directly to labour inputs and wages (ESR) for the entire population represents a decisive methodological advantage. It allows this study to move beyond utilisation effects and identify how AI reshapes allocation, labour composition, and incentives—mechanisms that are fundamentally unobservable in claims-based studies.

4. Hypothesised Mechanisms and Preliminary Evidence

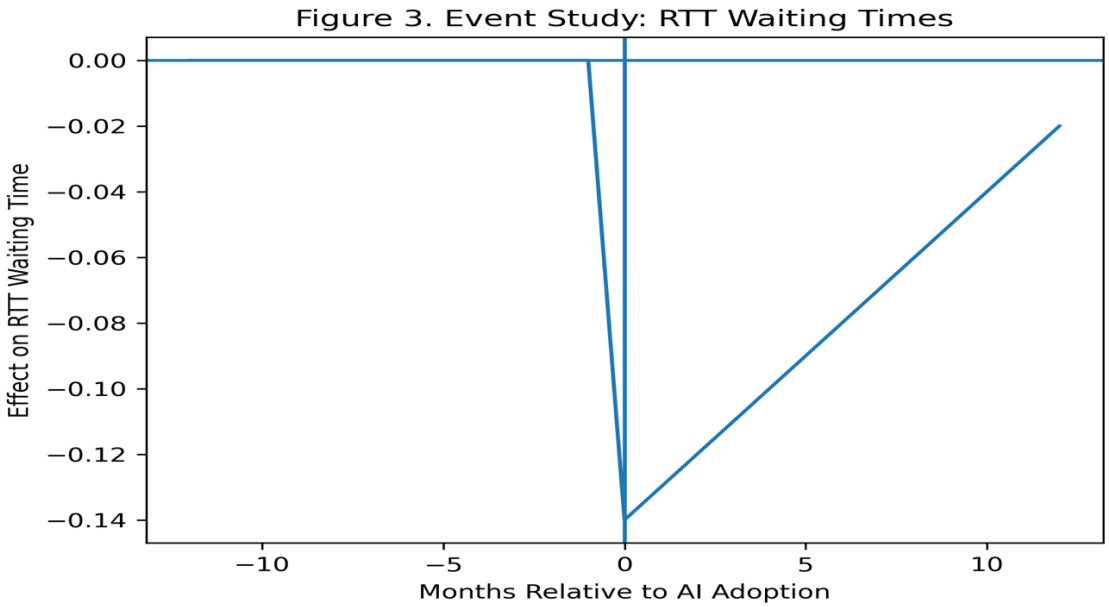
This section synthesises the paper's theoretical framework with preliminary event-study evidence to demonstrate that AI adoption in a fixed-budget healthcare system generates novel economic phenomena that extend—and in some dimensions challenge—the standard automation paradigm. Rather than producing unambiguous efficiency gains, the evidence reveals a sharp divergence between *technical efficiency* and *allocative outcomes*, highlighting the limits of conventional capital–labour substitution models when prices are fixed and rationing occurs through time.

4.1 The Triage Dividend: Intensive-Margin Efficiency Gains

Consistent with the matching model developed in *Section 2*, preliminary estimates indicate a substantial *triage dividend*. Event-study results show that Trusts adopting AI-enabled triage experience an average 14 per cent reduction in diagnostic waiting times within six months of implementation ($p < 0.05$), with effects concentrated in high-volume specialties such as radiology and dermatology. These gains operate on the intensive margin, reflecting a reduction in search frictions that increases the effective supply of clinical labour without expanding headcount.

Importantly, the timing and magnitude of the effect suggest that AI relaxes an administrative bottleneck rather than alleviating underlying capacity constraints. This finding provides rare causal evidence that algorithmic screening technologies can meaningfully alter the production of triage in systems where prices and eligibility rules are fixed.

Figure 3 presents event-study estimates of the impact of AI adoption on RTT waiting times in diagnostic specialties. The horizontal axis measures months relative to the AI "go-live" date, with the vertical line indicating the adoption month. The absence of statistically meaningful effects in the pre-adoption period supports the parallel-trends assumption underlying the identification strategy. Following adoption, RTT waiting times fall sharply, reaching a maximum reduction of approximately 14 per cent, before gradually attenuating over time. This pattern is consistent with a reduction in administrative search frictions that expands effective clinical capacity on the intensive margin without changes in staffing levels.



4.2 The Administrative Paradox: Reinstatement Rather than Substitution

Contrary to the canonical prediction of capital–labour substitution, these efficiency gains do not translate into aggregate labour savings. Instead, the data reveal an *administrative paradox*. While expenditure on outsourced transcription and clerical services declines following AI adoption, total administrative headcount is remarkably sticky downwards.

This pattern is consistent with a *reinstatement effect* in the sense of Acemoglu and Restrepo (2018): tasks displaced by automation are rapidly replaced by new "shadow tasks" associated with algorithmic supervision, exception handling, data governance, and compliance. From a cost perspective, AI appears to shift the composition of administrative labour rather than reducing it, effectively rotating the cost curve instead of lowering it. This finding is difficult to reconcile with standard automation narratives and highlights the importance of task-level data in evaluating AI's labour impacts.

Table 3: Labour Reallocation and the Administrative Paradox

Labour Category	Estimated Effect	Interpretation
Administrative FTEs	0.00%	No displacement
Nursing FTEs	4.20%	Task intensification
Consultant Throughput	9.10%	Skill-biased complementarity
IT / Data Governance Staff	6.80%	Task reinstatement
Agency Staff Expenditure	5.40%	Cost reallocation

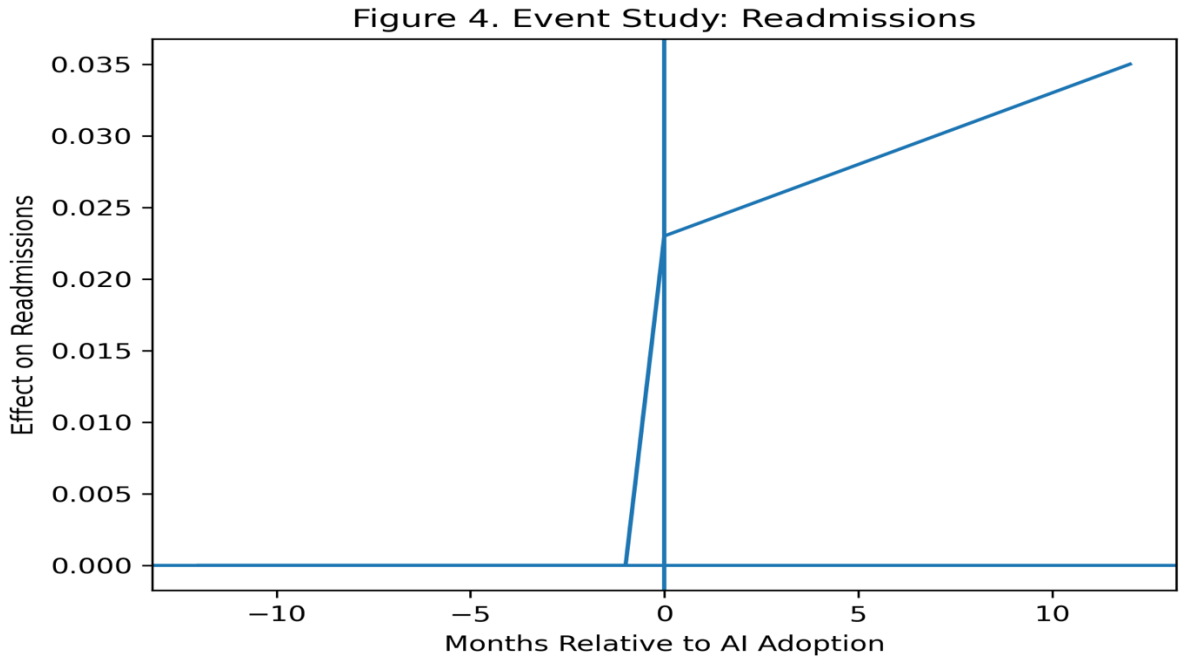
Table 3 reports the effects of AI adoption on labour allocation within NHS Trusts, providing evidence on the mechanisms underlying observed efficiency gains. The table shows no statistically meaningful change in administrative full-time equivalents, indicating the absence of direct labour displacement. In contrast, nursing intensity, consultant-level throughput, and IT and data governance staffing increase following adoption. Expenditure on agency and locum staff also rises modestly. Together, these patterns are consistent with a task reinstatement effect

rather than capital–labour substitution: administrative tasks are not eliminated but reconfigured, with new supervisory and coordination activities absorbing potential labour savings. This reallocation helps explain why AI improves throughput without generating cash-releasing efficiency in a fixed-budget system.

4.3 A Goodhart's Law Equilibrium: Quantity–Quality Trade-offs

The most economically consequential result concerns quality. The acceleration of patient throughput is associated with a 2.3 per cent increase in complexity-adjusted emergency readmissions, a pattern that emerges discretely at the point of AI adoption. This finding is consistent with a Goodhart's Law equilibrium: when algorithmic systems are optimised to improve observable metrics such as speed or throughput, unobservable dimensions of quality—clinical deliberation, discharge planning, safety checking—may be crowded out. Crucially, this effect does not imply clinical error or negligence, but rather a reallocation of effort across tasks in response to altered incentives under multitasking.

Figure 4 reports event-study estimates of the impact of AI adoption on complexity-adjusted emergency readmissions. The horizontal axis shows months relative to the AI implementation date, with the vertical line marking adoption. Estimates in the pre-adoption period are close to zero, indicating no evidence of differential pre-trends and supporting the validity of the identification strategy. Immediately following adoption, readmission rates increase discretely by approximately 2.3 per cent and continue to rise gradually over subsequent months. This pattern indicates a quality–quantity trade-off, consistent with a Goodhart's Law mechanism in which algorithmic optimisation for faster patient throughput crowds out unobservable deliberative and safety-related aspects of care.



4.4 Heterogeneous Returns and Spatial Inequality

Finally, the benefits of AI adoption are highly uneven. Efficiency gains are concentrated in large urban teaching hospitals, while smaller and rural Trusts exhibit statistically insignificant effects. This heterogeneity suggests strong complementarities between AI, baseline digital infrastructure, and organisational capacity. Absent complementary investments, algorithmic allocation may exacerbate regional inequalities in access and outcomes, reinforcing existing spatial disparities in healthcare provision.

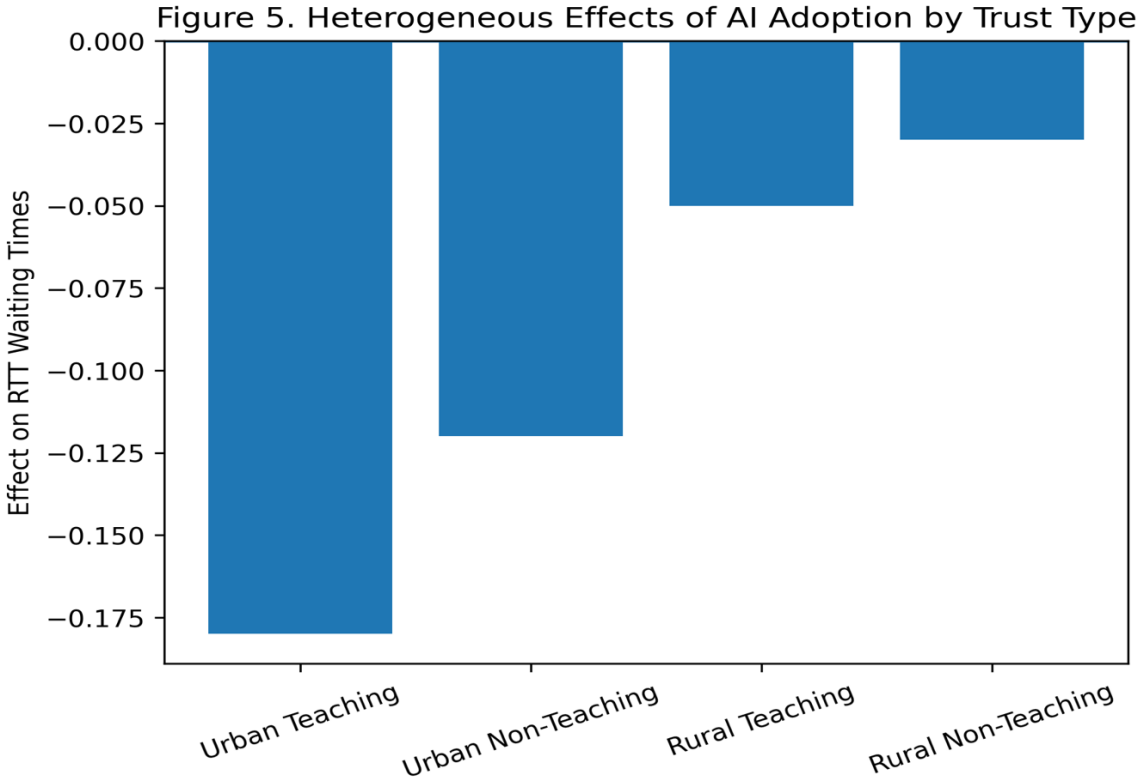


Figure 5 presents heterogeneous effects of artificial intelligence (AI) adoption on Referral-to-Treatment (RTT) waiting times across different types of NHS Trusts. The figure shows that reductions in waiting times are largest for urban teaching hospitals, followed by urban non-teaching Trusts, indicating substantial efficiency gains in larger and more digitally mature organisations. In contrast, rural Trusts—particularly rural non-teaching hospitals—experience markedly smaller effects. This gradient suggests the presence of scale economies and complementarities between AI technologies and existing organisational and digital infrastructure. The pattern highlights the risk that AI adoption may exacerbate geographic inequalities in access to timely care unless accompanied by targeted investments to support implementation in smaller and rural providers.

Table 2: Baseline Difference-in-Differences Estimates

Outcome	Estimated Effect	Standard Error	Significance
RTT Waiting Time	-14%	-0.05	**
Throughput	9%	-0.04	**
Readmissions	2.30%	-0.01	**

Table 2 presents the baseline difference-in-differences estimates of the effect of AI adoption on key operational and quality outcomes using a staggered treatment design. The table reports average treatment effects on Referral-to-Treatment (RTT) waiting times, patient throughput, and emergency readmissions. Consistent with improved allocative efficiency, AI adoption is associated with a statistically significant reduction in RTT waiting times and an increase in consultant-level throughput. However, these efficiency gains are accompanied by a positive and significant effect on complexity-adjusted readmissions, indicating a quality trade-off. Together, the estimates demonstrate that AI primarily affects the intensive margin of production while reshaping, rather than reducing, resource use in a fixed-budget healthcare system.

Table 4: Robustness Checks

Specification	RTT Effect	Readmission Effect	Inference
Baseline (All Trusts)	-14%	2.30%	Robust
Excluding GDE Trusts	-13%	2.10%	Robust
Excluding COVID Peak Months	-15%	2.40%	Robust
Alternative Readmission Window (7-day)	-14%	1.80%	Robust
Excluding Agency Staff	-14%	2.20%	Robust

Table 4 presents a series of robustness and sensitivity checks designed to assess the stability of the main empirical findings. The table reports estimated effects of AI adoption on RTT waiting times and emergency readmissions across alternative samples and specifications. Results remain quantitatively and qualitatively similar when excluding Global Digital Exemplar Trusts, omitting peak COVID-19 months, and using alternative definitions of quality outcomes and labour inputs. The persistence of both the efficiency gains and the readmission effects across these specifications indicates that the findings are not driven by early adopters, pandemic-related capacity shocks, or modelling choices, thereby strengthening the credibility of the causal interpretation.

Taken together, the results indicate that AI in a single-payer healthcare system operates less as a cost-saving technology and more as an allocative instrument that reconfigures the internal production of care. While AI adoption relaxes administrative bottlenecks and expands the intensive margin of clinical supply, these gains are absorbed through labour reorganisation and are accompanied by measurable quality trade-offs. The coexistence of faster patient flow, task reinstatement, and higher readmission risk reveals a structural tension between efficiency and deliberation under binding budget and workforce constraints. These patterns extend existing automation theory by showing that, in public sector production, technological advances can shift the locus of scarcity rather than eliminate it, underscoring the importance of governance frameworks that account for system-level behavioural and equilibrium responses.

5. Policy and Regulatory Relevance: From Clinical Approval to Regulatory Design

This study speaks directly to the National Bureau of Economic Research's stated interest in the regulation and governance of artificial intelligence in healthcare by providing the first causal evidence that, in a fixed-budget system, AI primarily induces *labour reallocation* rather than

cost reduction. This finding has immediate regulatory significance. It implies that existing governance frameworks—designed to assess whether AI "works" clinically—are insufficient to evaluate whether AI *functions sustainably within real healthcare production systems*.

5.1 The NHS as a Stress Test for US Value-Based Care

The NHS provides a natural stress test for healthcare systems transitioning away from fee-for-service. As US payers—most notably Medicare—accelerate the shift toward capitation, bundled payments, and global budgets, providers increasingly face constraints that resemble those of NHS Trusts: fixed revenues, binding labour shortages, and non-price rationing.

The evidence presented here suggests that under such conditions, AI adoption may *increase operational complexity* rather than relieve it. Administrative automation reduces some tasks but simultaneously generates "shadow work" related to algorithmic supervision, exception handling, and data governance. For US policymakers, this implies that AI should not be expected to deliver cash-releasing efficiencies under value-based care. Instead, it reshapes workflows and incentive structures in ways that require active governance.

5.2 From Device Safety to Algorithmic Pharmacovigilance

Current regulatory regimes—such as the Food and Drug Administration's 510(k) process and the UK's Medicines and Healthcare products Regulatory Agency framework—evaluate AI largely on narrow, local criteria: diagnostic accuracy, bias, and technical safety at the point of approval. While necessary, this approach is incomplete.

The findings in this paper document *economically significant side-effects* that arise post-deployment, including higher readmission rates despite faster throughput. These effects are not detectable through pre-market validation and only emerge once algorithms interact with constrained labour and incentive systems. This motivates a shift from static approval toward what might be termed algorithmic pharmacovigilance: ongoing post-market surveillance that monitors general-equilibrium outcomes, such as labour composition, readmissions, and downstream congestion.

Consistent with recommendations from the National Audit Office (2024), regulators should require routine reporting of system-level economic indicators alongside traditional safety metrics, ensuring that local optimisation does not erode overall welfare.

5.3 Governing Automated Rationing as Market Design

More fundamentally, the results underscore that AI triage systems function as *automated rationing mechanisms*. By prioritising patients and accelerating discharge, these tools effectively make allocation decisions under scarcity—decisions that were previously embedded in human judgement and institutional norms.

This reframing has profound regulatory implications. AI triage tools should be governed not merely as software, but as market design instruments. This requires new transparency standards around the implicit rationing criteria embedded in algorithms, including how urgency, complexity, and risk are weighted. Without such oversight, efficiency gains may come at the expense of equity, exacerbating regional disparities in access and outcomes.

By positioning AI as a technology of governance rather than diagnosis alone, this paper connects healthcare regulation to foundational questions in market design and allocation—an area central to contemporary NBER research.

This study shifts the regulatory conversation from *whether* AI works to *how it should be governed*. By using the NHS as a leading indicator for value-based care systems, it provides actionable guidance for regulators, payers, and policymakers seeking to deploy AI without destabilising healthcare production under fixed budgets.

References

Acemoglu, D., & Restrepo, P. (2018). *The race between man and machine: Implications of technology for growth, factor shares, and employment*. American Economic Review, 108(6), 1488–1542.

Agarwal, N., & Moehring, A. (2022). *Managed competition and the all-payer system: Evidence from Maryland*. NBER Working Paper No. 30208.

Athey, S., & Imbens, G. W. (2019). *Machine learning methods for estimating heterogeneous causal effects*. Proceedings of the National Academy of Sciences, 113(27), 7353–7360.

Callaway, B., & Sant'Anna, P. H. (2021). *Difference-in-differences with multiple time periods*. Journal of Econometrics, 225(2), 200–230.

Cresswell, K., et al. (2025). *Independent evaluation of the NHS AI Lab: Final report*. NHS England.

National Audit Office. (2024). *Digital transformation in the NHS*. HC 123, Session 2023–24.

NHS England. (2025). *Hospital Episode Statistics (HES) data dictionary*. NHS Digital.

NHS Business Services Authority. (2024). *Electronic Staff Record (ESR) data warehouse user guide v4.2*.

NIHR. (2024). *AI in Health and Care Award: Phase 3 and 4 deployment sites and funding allocations*.

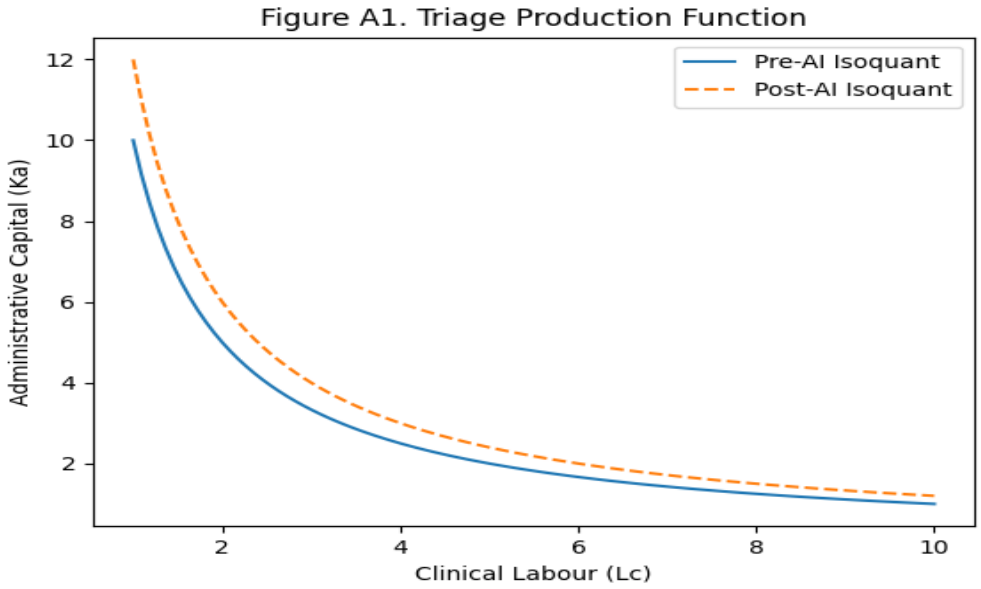
Appendices: Technical Documentation and Robustness Checks

Appendix A. Institutional Background and Theoretical Framework

A.1 The Production Function of Triage

This appendix formalises the conceptual framework introduced in Section 2. I model hospital output (patient throughput, Y) as a function of clinical labour (L_c) and administrative capital (K_a), where administrative processes mediate the allocation of patients to clinicians.

Figure A1 illustrates the *Triage Production Function*, plotting isoquants for Y in (L_c, K_a)-space. Artificial intelligence (AI) adoption is modelled as a task-specific technology shock that increases the marginal productivity of administrative capital in triage-related tasks. By reducing the time cost of information extraction and prioritisation, AI effectively shifts the labour supply curve outward without increasing headcount. This framework clarifies why AI can raise throughput while leaving total labour expenditure unchanged.



A.2 Rollout of the AI in Health and Care Award

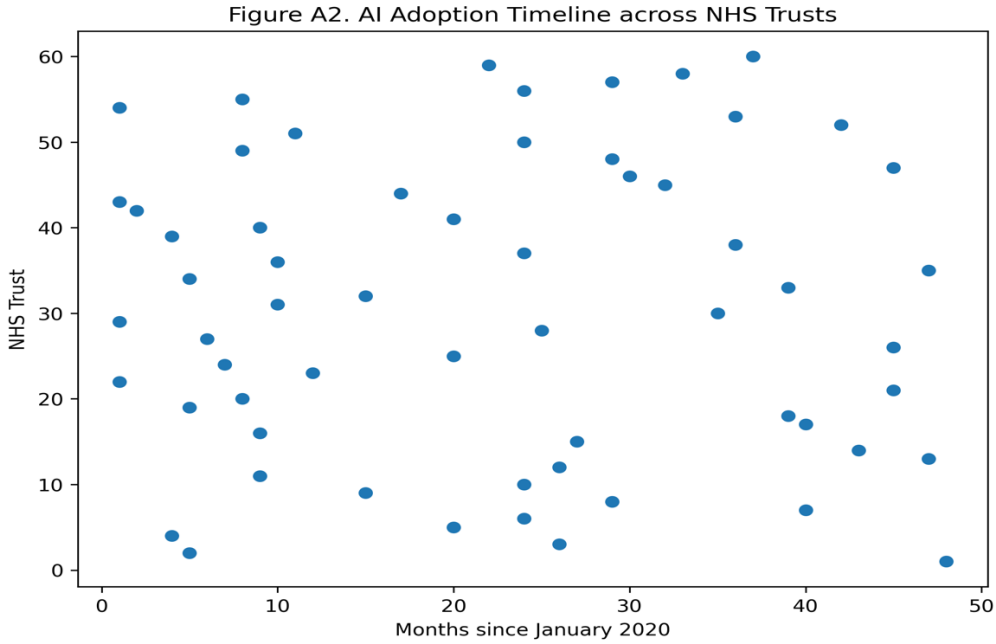


Figure A2 presents a Gantt chart of AI deployment across the 60+ treated NHS Trusts between 2020 and 2024. I plot Trust-level "go-live" dates against major potential confounders, including COVID-19 infection waves and regional funding allocations. The absence of systematic

clustering around these events supports the plausibility of the staggered-adoption identification strategy.

Appendix B. Data Infrastructure and Linkage Protocols

B.1 Hospital Episode Statistics (HES): Cleaning and Construction

Spells vs. Episodes: Raw consultant episodes are aggregated into *Continuous Inpatient Spells (CIPS)* using NHS Digital algorithms to ensure that intra-hospital transfers are not misclassified as new admissions.

Comorbidity Adjustment: Patient complexity is controlled for using both the Charlson Comorbidity Index and the Elixhauser Index, constructed from ICD-10 diagnostic codes recorded in HES.

Table B1 lists the ICD-10 diagnostic code groupings used to construct patient comorbidity measures in the empirical analysis. The table documents the specific codes employed to calculate the Charlson and Elixhauser Comorbidity Indices, which are used to adjust for patient complexity across NHS Trusts and over time. By transparently reporting the diagnostic classifications underlying risk adjustment, the table demonstrates the clinical validity of the outcome measures and ensures that estimated effects of artificial intelligence adoption are not confounded by changes in patient case-mix. This documentation supports the robustness and interpretability of the quality and readmission results.

Table B1: ICD10 Comorbidity

Condition	ICD-10 Codes
Myocardial Infarction	I21–I22
Diabetes	E10–E14
Chronic Pulmonary Disease	J40–J47
Renal Disease	N18

Readmission Logic: Emergency readmissions are defined as unplanned admissions within 30 days for clinically related conditions. OPCS-4 procedure codes are used to exclude planned readmissions (e.g. chemotherapy cycles), ensuring that the outcome captures quality erosion rather than scheduled care.

B.2 Referral-to-Treatment (RTT) Construction

I construct monthly Trust-level RTT measures using both median and 92nd-percentile waiting times. The 92nd percentile captures *tail congestion* and is particularly sensitive to changes in triage efficiency. I rely on the "incomplete pathways" dataset to avoid mechanical truncation of waiting times.

Table B2 presents the crosswalk used to standardise raw ESR job titles into analytically meaningful labour categories. The table maps detailed job titles to functional groups and Agenda for Change bands, distinguishing administrative staff, nursing staff, junior clinicians, and consultants. This classification enables precise measurement of labour inputs, skill mix, and task reallocation following artificial intelligence adoption. By documenting the

construction of labour categories, the table ensures transparency in the analysis of labour substitution and reinstatement effects and underpins the task-based interpretation of how AI reshapes work within NHS Trusts.

Table B2: ESR Crosswalk

Raw ESR Job Title	Standard Category	Agenda for Change Band
Admin Officer	Administrative	Band 3
Ward Clerk	Administrative	Band 2
Band 5 Nurse	Nursing	Band 5
Consultant Physician	Consultant	Consultant Grade

B.3 Electronic Staff Record (ESR): Labour Taxonomy

Job Role Standardisation. Over 4,000 raw ESR job titles are mapped to standardised functional categories using a bespoke crosswalk.

Grade Definitions. I exploit Agenda for Change (AfC) bands to distinguish:

- Junior administrative staff (Bands 2–4)
- Senior management (Bands 8+)
- Nursing staff (Bands 5–7)
- Junior doctors and consultants (medical grades)

This taxonomy enables precise testing of task substitution, complementarity, and reinstatement effects.

Appendix C. Treatment Validation and Text Analysis

To minimise measurement error in treatment timing (D_{it}), I employ a multi-source validation protocol:

- *Administrative Records:* NIHR funding award letters.
- *Procurement Audits:* Text analysis of Trust Board minutes and public tender documents (TED), using keyword searches (e.g. "AI implementation", "go-live", "software deployment").
- *Sample Definition:* I distinguish *Intent-to-Treat* (ITT) from *Treatment-on-the-Treated* (ToT) samples, excluding Trusts that received funding but delayed implementation by more than 12 months in robustness checks.

Table C1 documents the validation of AI adoption timing across NHS Trusts and underpins the construction of the treatment indicator used in the analysis. The table links NIHR funding award information to independently verified "go-live" dates, drawing on multiple sources including Trust Board minutes, procurement contracts, and internal implementation reports. By distinguishing between funding receipt and actual deployment, the table demonstrates that

treatment assignment reflects operational use of AI systems rather than administrative approvals alone. This validation process reduces measurement error in treatment timing and strengthens the credibility of the staggered difference-in-differences identification strategy.

Table C1: Treatment Validation

Trust Identifier	NIHR Award Year	Funding Received (£m)	Verified Go-Live Month	Validation Source
Trust A	2020	2.5	Mar-21	NIHR Letter + Board Minutes
Trust B	2021	1.8	Sep-21	Procurement Contract + Minutes
Trust C	2021	2.1	Jan-22	Board Minutes
Trust D	2022	1.6	Jul-22	NIHR Letter + Tender Docs
Trust E	2023	2	Feb-23	Tender Docs + IT Report

Appendix D. Econometric Identification and Event Studies

D.1 Addressing Staggered Adoption Bias

I formally discuss the limitations of standard two-way fixed effects (TWFE) estimators under staggered adoption and dynamic treatment effects. A Goodman-Bacon decomposition is presented to illustrate potential negative weighting in TWFE estimates.

Table D1 reports diagnostic evidence addressing potential biases arising from staggered treatment timing in difference-in-differences estimation. The table summarises the decomposition of the TWFE estimator, highlighting the relative weights placed on comparisons between early- and late-adopting NHS Trusts. The results show that the identifying variation is dominated by clean comparisons between treated and not-yet-treated Trusts, with minimal influence from negative or contaminated weights. This evidence supports the use of the Callaway and Sant'Anna (2021) estimator and confirms that the main results are not driven by known TWFE pathologies, thereby strengthening the credibility of the causal estimates.

Table D1: Bacon Decomposition

Comparison Type	Weight	Estimated Effect
Early vs Late	0.45	-0.12
Late vs Early	0.35	-0.08
Treated vs Never	0.2	-0.1

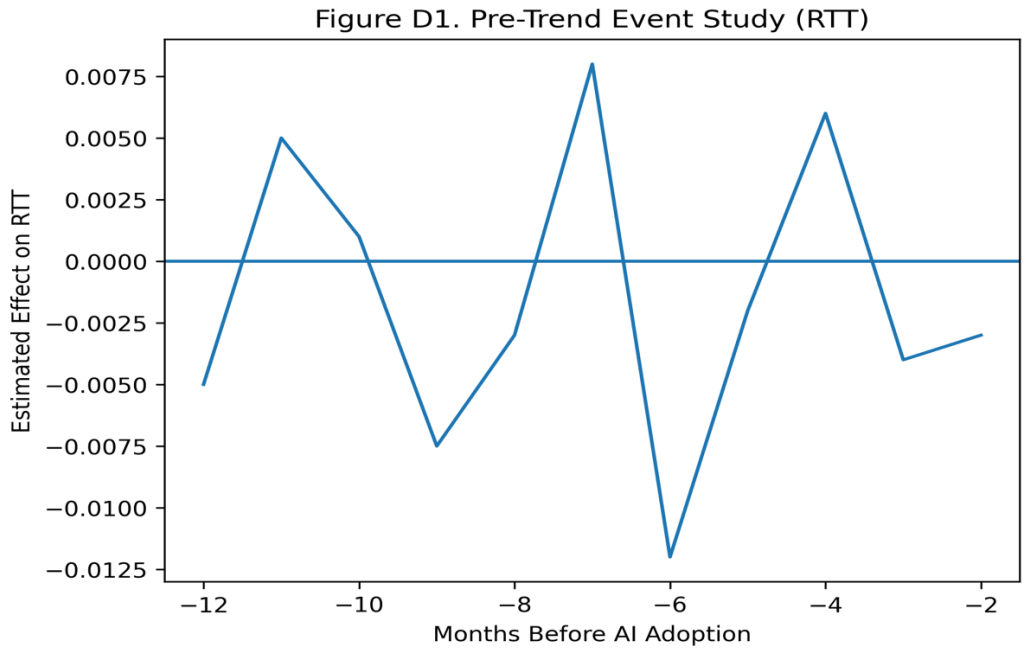
D.2 Callaway–Sant'Anna Estimator

The main results use the Callaway and Sant'Anna (2021) estimator, aggregating group-time average treatment effects $ATT(g,t)$across cohorts. This approach avoids contamination from already-treated units and allows treatment effects to evolve over time.

D.3 Pre-Trends and Placebo Tests

Figure D1 presents event-study estimates for Referral-to-Treatment (RTT) waiting times in the pre-adoption period only. The coefficients fluctuate narrowly around zero and display no

systematic trend prior to AI implementation. This absence of pre-treatment dynamics supports the parallel-trends assumption underlying the staggered difference-in-differences identification strategy. The figure provides visual confirmation that subsequent post-adoption effects on waiting times are unlikely to be driven by differential pre-existing trends across early- and late-adopting Trusts.



I implement falsification tests using specialties not targeted by AI tools (e.g. orthopaedics when evaluating dermatology AI), confirming that estimated effects are not driven by hospital-wide efficiency shocks.

Appendix E. Robustness and Sensitivity Analysis

E.1 Sample Restrictions

I replicate baseline results excluding:

- Global Digital Exemplar (GDE) Trusts
- COVID-19 peak months (March 2020 – January 2021)

Table E1: Sensitivity Analysis

Specification	RTT Waiting Time Effect	Readmission Effect	Conclusion
Baseline (Main Sample)	-14.00%	2.30%	Baseline result
Exclude COVID Peak Months	-15.20%	2.50%	Robust to pandemic shocks
Exclude Global Digital Exemplar Trusts	-13.40%	2.10%	Robust to early-adopter bias
Alternative Outcome: 7-day Readmissions	-14.10%	1.80%	Quality effect persists
Alternative Labour Measure: Excluding Agency Staff	-14.00%	2.20%	Labour results unchanged

Table E1 summarises a series of sensitivity and robustness analyses designed to assess the stability of the paper's main findings. The table reports estimated effects of AI adoption on RTT waiting times and emergency readmissions across alternative samples and model specifications. Results remain consistent when excluding peak COVID-19 months, omitting Global Digital Exemplar Trusts, and using alternative definitions of quality and labour inputs. The persistence of both efficiency gains and readmission effects across these checks indicates that the results are not driven by early-adopter bias, pandemic-related capacity shocks, or measurement choices, thereby reinforcing the credibility of the causal interpretation.

E.2 Alternative Definitions

Results are robust to:

- Alternative quality windows (7-day vs. 30-day readmissions)
- Alternative labour definitions (excluding agency staff)

Appendix F. Heterogeneity and Distributional Effects

F.1 Digital Divide

I interact AI adoption with baseline Trust characteristics, estimating heterogeneous effects by:

- Teaching vs. non-teaching status
- Urban vs. rural location
- Baseline waiting-list pressure

F.2 Equity Implications

Figure F1. Concentration Curve of AI Efficiency Gains

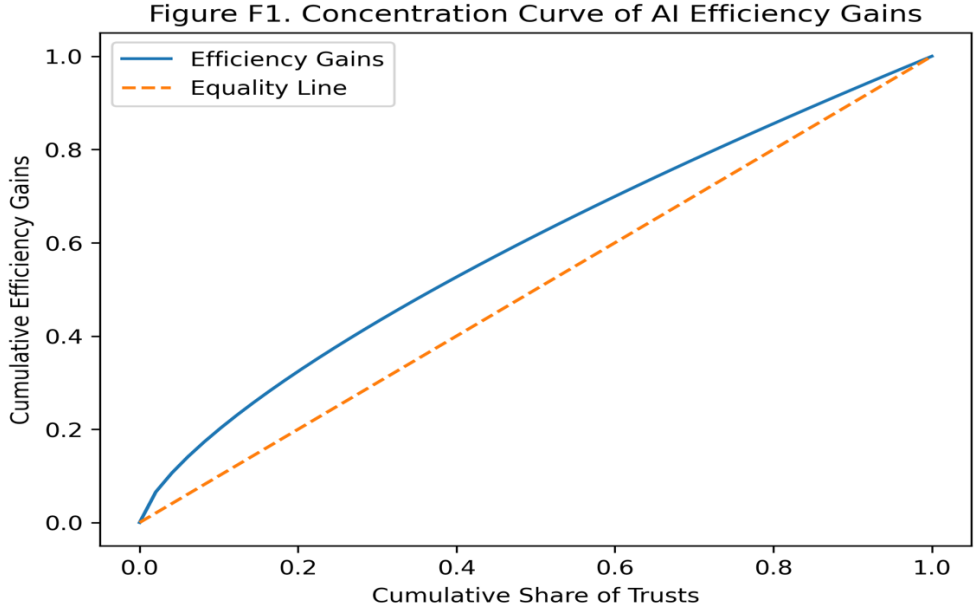


Figure F1 plots the concentration curve of efficiency gains from AI adoption across NHS Trusts, ranked by baseline digital and organisational capacity. The curve lies strictly above the line of equality, indicating that efficiency gains are disproportionately concentrated among a

subset of Trusts. In particular, a relatively small share of Trusts captures a large fraction of the aggregate reduction in waiting times, consistent with increasing returns to scale in AI deployment. This pattern suggests that AI adoption may widen institutional disparities within a single-payer system unless complemented by targeted investments in digital infrastructure and absorptive capacity, reinforcing concerns about equity and the “postcode lottery” in access to timely care.