© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

Forecasting NHS Emergency Demand: A Box-Jenkins ARIMA Approach to A&E Capacity Planning

PRECIOUS EJIBA[1], EFEMENA IKPRO[2]
[1]*Science & Engineering, DB Biomedical Ltd, University of Northampton, Northampton, United Kingdom*
[2]*Technology and Business, Amdari, London, United Kingdom, EC1V 2NX*

Abstract-** Accurate short-to-medium-term forecasting of Accident & Emergency (A&E) attendance demand is a precondition for safe capacity planning across the National Health Service (NHS), where staffing, bed allocation, and escalation decisions are made against a constitutional four-hour treatment standard that the system has not met since 2014-15. This study applies a Box-Jenkins ARIMA methodology, benchmarked against Holt's Linear Exponential Smoothing (LES), to 61 quarters (2011-12 Q1 to 2026-27 Q1) of NHS England quarterly total A&E attendance data. The series is confirmed non-stationary in levels (Augmented Dickey-Fuller statistic = -1.74) and stationary after first differencing (ADF = -4.56). Autocorrelation and partial autocorrelation analysis of the differenced series identifies a single significant lag-1 term consistent with an IMA(1,1) signature. An ARIMA(0,1,1) specification is selected over higher-order alternatives on an eight-quarter holdout backtest (RMSE = 154,563; MAE = 127,861; MAPE = 1.86%), outperforming both Holt's LES (MAPE = 2.22%) and a more heavily parameterised ARIMA(2,1,2) that achieves marginally better in-sample information criteria but materially worse out-of-sample accuracy (MAPE = 3.59%) which is a concrete demonstration of in-sample overfitting. Ljung-Box diagnostics confirm residual white noise at all tested lags. The selected model projects total quarterly attendances rising from approximately 7.18 million in 2026-27 Q2 to 7.72 million by 2031-32 Q1 (+7.5%), with a 95% prediction interval that widens from ±905,000 to ±5.33 million over the five-year horizon, a finding with direct implications for how far ahead NHS planners can responsibly commit capacity decisions to a single point forecast. The paper closes with recommendations for NHS operational planning teams and for healthcare data science practice more broadly.

***Keywords-** ARIMA; Box-Jenkins methodology; exponential smoothing; healthcare demand forecasting; NHS A&E attendances; capacity planning; forecast uncertainty.*

I. INTRODUCTION

Accident & Emergency departments occupy the most operationally exposed position in the NHS: demand arrives unscheduled, capacity is fixed in the short run, and the consequence of under-forecasting is measured in corridor care, breached waiting-time standards, and, at the margin, patient harm (Silva et al., 2023). Since 2011-12, NHS England quarterly A&E attendance volumes have risen from roughly 5.5 million to over 7.2 million, a trajectory disrupted only by the acute 2020 COVID-19 contraction and its subsequent overshoot. Winter 2024-25 saw only 73% of patients treated within the four-hour constitutional standard, against a 95% target last met in 2014-15, alongside bed occupancy persistently above the 85% safety threshold recommended by the Royal College of Emergency Medicine (Health Foundation, 2025). In December 2025 the UK government began national rollout of an AI-based A&E demand-forecasting tool across 50 NHS organisations specifically to address this planning gap (GOV.UK, 2025), underscoring that forecasting-technique selection is now an active operational question rather than a purely academic one.

This study applies the Box-Jenkins ARIMA methodology, benchmarked against Holt's Linear Exponential Smoothing, to NHS England's official quarterly A&E attendance time series. The objective is threefold: to identify a statistically adequate forecasting model for total quarterly attendance demand; to compare its accuracy and behaviour against a simpler trend-extrapolating alternative; and to translate the resulting forecast, including its uncertainty, into concrete implications for NHS capacity planning. The analysis follows the same five-stage Box-Jenkins framework (identification, estimation, diagnostic checking, forecasting, and model comparison) commonly applied to commodity

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

and financial series, testing whether a methodology built for continuous market prices transfers cleanly to a healthcare demand series shaped by seasonal illness cycles, policy interventions, and public health shocks. Figure 1 sets out the conceptual logic of the study: the operational tension between rising demand and fixed capacity that motivates it, the modelling approach and model selected, and the resulting forecast's implications for planning.

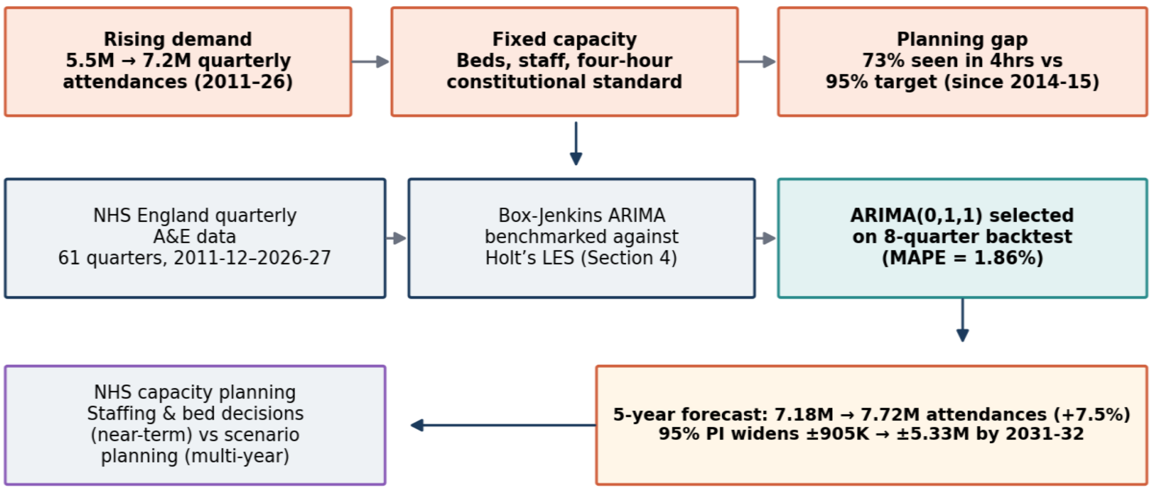

Figure 1: Conceptual Framework of This Study

Illustrates the operational tension between rising NHS demand and fixed capacity that motivates this study, the modelling approach and model selected, and the resulting forecast's implications for capacity planning.

II. Overview of Literature

2.1 ARIMA Forecasting in Healthcare Demand Contexts

ARIMA remains the most frequently applied technique in the emergency-department forecasting literature. A systematic review of seven Emergency Department (ED) forecasting studies found that ARIMA-based and other linear models performed well for short-horizon forecasts, with errors generally below 10% MAPE, while noting that ARIMA's relative advantage narrows as forecast horizon and department size increase (Silva et al., 2023). This horizon-sensitivity is directly relevant to the present study, which forecasts twenty quarters ahead, far beyond the daily and weekly horizons typical of the ED-forecasting literature and is one reason the widening prediction interval reported in Section 4.5 must be treated as a central finding rather than a technical footnote.

Vollmer et al. (2021) developed a unified machine-learning pipeline benchmarked against classical time-series methods for forecasting demand at a London NHS trust's emergency department, finding that ARIMA and related linear models remained competitive at shorter horizons even as more flexible learners were introduced. Tuominen et al. (2022), forecasting daily ED arrivals using high-dimensional multivariate feature sets, found that feature-engineered machine-learning models could outperform ARIMA where rich exogenous data (weather, calendar effects, local events) was available, a condition the present study's univariate quarterly series does not meet, and which is flagged as a limitation in Section 6.

2.2 Quantitative Forecasting Approaches and Model Selection Criteria

Where qualitative methods such as expert elicitation and scenario planning remain valuable for structurally novel situations (a new pathogen variant, an unprecedented policy shock), quantitative time-series

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

methods are preferred for routine operational forecasting precisely because they are auditable, reproducible, and can be formally validated against held-out data. Table 1 summarises this distinction as it applies to healthcare demand planning. Reviews of the broader time-series prediction literature reinforce that model selection should rest on demonstrated out-of-sample accuracy rather than in-sample fit alone (Kolambe & Arora, 2024), a principle this study applies directly in Section 4.7, where a model with the best information criteria is rejected in favour of one with materially better back test accuracy.

Table 1. Qualitative and Quantitative Forecasting Approaches in Healthcare Demand Planning
This table contrasts the two forecasting paradigms as they apply to healthcare demand planning and identifies which is used, and why, in this study.

Dimension	Qualitative methods	Quantitative methods
Basis	Expert judgement, scenario planning, Delphi elicitation	Historical data patterns (trend, seasonality, autocorrelation)
Best suited to	Structurally novel situations with no comparable history	Routine, recurring operational demand
Auditability	Low — judgement-dependent, difficult to replicate	High — formally validated against held-out data
Example in this study	Anticipating an unprecedented pathogen outbreak	ARIMA(0,1,1) forecast of routine quarterly A&E attendance

Adapted from the qualitative and quantitative forecasting method taxonomy in Hyndman and Athanasopoulos (2021) for a healthcare demand-planning context. This study applies quantitative methods throughout, consistent with the routine, recurring nature of the A&E attendance series analysed.

Kontopoulou et al. (2023) conducted an extensive review comparing ARIMA against machine-learning approaches across finance, health, weather, and network-traffic applications, concluding that machine-learning methods tend to dominate on raw post-sample accuracy, but that ARIMA retains structural advantages in transparency, computational cost, and auditability that matter disproportionately in regulated, resource-constrained settings such as public healthcare planning. This transparency argument underpins the present study's choice to build and report a fully interpretable Box-Jenkins model rather than a black-box learner, consistent with NHS operational teams needing to explain, not merely trust, a forecast.

2.3 Comparative Evaluation of Forecasting Techniques Across Healthcare Contexts

Direct comparisons of ARIMA against alternative techniques in healthcare settings show no universally superior method; performance is highly context-dependent, as summarised in Table 2. Farimani et al. (2024), reviewing models for predicting emergency-department length of stay, found wide variation in which technique performed best depending on data granularity and case-mix. Poursoltan et al. (2025) compared econometric, machine-learning, and foundation models for forecasting ED boarding-patient counts and found that simpler econometric methods remained competitive against far more complex foundation models on several accuracy metrics, reinforcing that model complexity does not guarantee forecasting advantage. Morbey et al. (2025), piloting a machine-learning forecasting pipeline for peak healthcare demand during the UK's 2022-23 winter respiratory season, demonstrated that ensembles of relatively simple models, correctly validated, can deliver operationally reliable short-term forecasts, a finding consistent with this study's own preference for a parsimonious, well-diagnosed

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

ARIMA specification over a more heavily parameterised alternative.

Table 2. Comparative Forecasting Techniques Across Healthcare Demand Contexts
This table summarises the typical horizon, strengths, and limitations of each technique as reported in the healthcare-forecasting literature

Technique	Typical horizon	Key strength	Key limitation	Source
ARIMA / Box-Jenkins	Short–medium	Transparent, auditable, low data requirement	Linear; interval widens quickly beyond fitted horizon	Kontopoulou et al. (2023)
Exponential smoothing	Short	Simple, fast to fit and update	Extrapolates trend even where none is statistically supported	*Hyndman and Athanasopoulos (2021)*
Feature-engineered ML	Short	Exploits exogenous data (weather, calendar)	Requires rich covariates; less transparent	Tuominen et al. (2022)
Foundation / deep learning	Short–medium	Can capture complex nonlinear patterns	High complexity not always reflected in accuracy gains	Poursoltan et al. (2025)

These findings collectively motivate the comparative design adopted here: rather than assuming ARIMA's dominance a priori, this study benchmarks it against Holt's LES and against a more complex ARIMA variant, selecting the final specification on out-of-sample evidence.

2.4 Ethical Considerations in Healthcare Demand Forecasting

Forecasting in a public healthcare context carries ethical weight beyond statistical accuracy. Clear communication of a model's limitations and uncertainty is identified as essential to maintaining trust and enabling informed operational decisions in AI-supported healthcare systems (Pham, 2025). A parallel strand of the ethics literature emphasises that model transparency and explainability are preconditions for legitimate deployment of predictive tools in healthcare governance, since decision-makers who cannot interrogate a forecast's assumptions cannot be meaningfully accountable for decisions built on it (Bouderhem, 2024). Applied to demand forecasting specifically, this implies that a point forecast presented without its prediction interval is not merely incomplete but potentially misleading, a risk this study addresses directly by reporting 95% prediction intervals throughout, and by explicitly flagging how rapidly those intervals widen over the five-year horizon in Section 5.

III. METHODOLOGY AND ANALYTICAL FRAMEWORK

This study follows the five-stage Box-Jenkins methodology; identification, estimation, diagnostic checking, forecasting, and model comparison, illustrated conceptually in Figure 2 and Figure 3, and applied step-by-step in Section 4.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

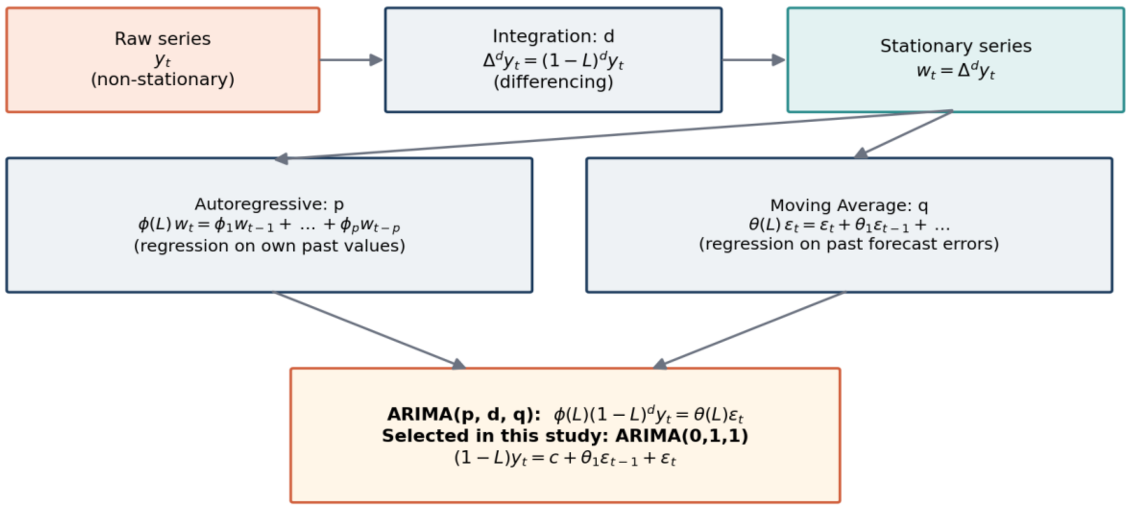

Figure 2: Conceptual Architecture of the Box-Jenkins ARIMA(p,d,q) Model

Shows how the integration (d), autoregressive (p), and moving-average (q) components combine into the general ARIMA(p,d,q) form, and the ARIMA(0,1,1) specification ultimately selected in Section 4.7. Author's conceptualisation based on Hyndman and Athanasopoulos (2021) and Box and Jenkins (1976).

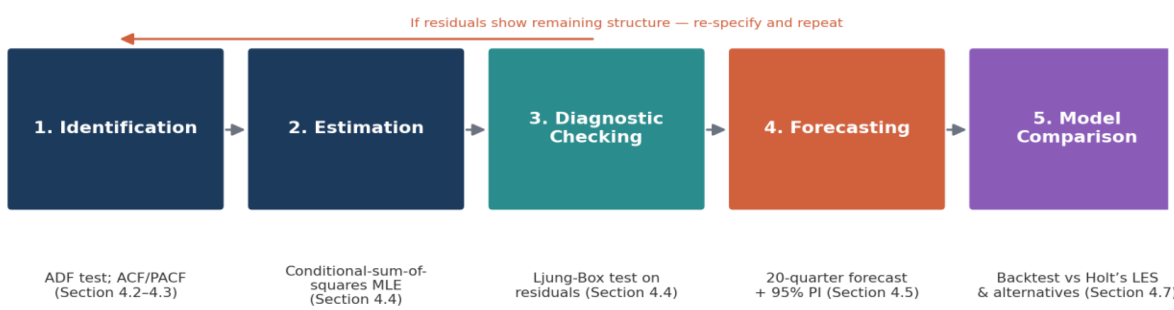

Figure 3: Five-Stage Box-Jenkins Methodology Applied in This Study

Maps each stage of the Box-Jenkins procedure — identification, estimation, diagnostic checking, forecasting, and model comparison — to the results section in which it is carried out.

3.1 Data Description and Preparation

The dataset comprises the official NHS England quarterly time series of total A&E attendances, published as part of the A&E Attendances and Emergency Admissions statistical work area under the Open Government Licence v3.0. Total attendances aggregate three provider types: Type 1 (major A&E departments), Type 2 (single-specialty departments), and Type 3 (other A&E, minor injury units, and urgent treatment centres). The full published series runs from 2004-05 Q1, but methodology changed twice over that period, from the QMAE collection (to 2010-11) to the WSitAE weekly collection (2011-12 to mid-2015) and finally the MSitAE monthly collection (from July 2015), which NHS England's own documentation flags as a source of discontinuity in the earliest years.

IV. ANALYSIS AND FINDINGS

To avoid conflating a genuine demand signal with a data-collection artefact, this study restricts the analytical window to 2011-12 Q1 through 2026-27

Q1: 61 consecutive quarterly observations under methodologically consistent (WSitAE/MSitAE-basis) collection. Table 3 summarises the resulting series.

Table 3: Descriptive Statistics of NHS England Quarterly Total A&E Attendances (2011-12 Q1 – 2026-27 Q1)
Reports the mean, standard deviation, minimum, and maximum for the full series and for pre-COVID and post-recovery subperiods, showing the underlying demand growth separately from the pandemic shock.

Statistic	Full series (N = 61)	Pre-COVID (2011-12 Q1–2019-20 Q4, n = 36)	Post-recovery (2021-22 Q2–2026-27 Q1, n = 21)
Mean	5,956,871	5,758,623	6,601,387
SD	693,409	353,504	390,637
Minimum	3,589,014 (2020-21 Q1)	—	—
Maximum	7,240,633 (2026-27 Q1)	—	—

Note. The four quarters most acutely disrupted by the initial COVID-19 pandemic response (2020-21 Q1–2021-22 Q1) are excluded from both subperiod columns so that the pre/post comparison reflects underlying demand growth rather than the acute shock itself; the full series (N = 61) retains all quarters, including the shock, and is the series used for all modelling in Section 4.

No missing values were present in the extracted window. Quarterly aggregation follows NHS England's own published quarterly time series (13-week aggregates to 2015-16 Q1; three-calendar-month aggregates thereafter), so no additional aggregation was performed by the author. One documented data-quality caveat applies to a subset of the series: between May 2019 and May 2023, fourteen NHS trusts field-tested revised four-hour performance standards (the Clinical Review of Standards, or CRS, pilot), which affects the four-hour performance percentage series but not the total-attendance counts used as the forecasting target in this study, and is therefore noted but does not affect the analysis.

All statistical computation; stationarity testing, autocorrelation analysis, ARIMA and Holt's LES parameter estimation, diagnostic testing, and forecast generation was implemented in Python (NumPy and SciPy) using conditional-sum-of-squares maximum-likelihood estimation for the ARMA components, in place of a packaged auto-ARIMA routine, so that every stage of the Box-Jenkins procedure documented in Section 4 is fully reproducible from the annotated source code in Appendix A.

4.1 Exploratory Analysis

Figure 4.1 plots total quarterly A&E attendances across the 61-quarter window. The series shows a persistent upward trend from approximately 5.5 million attendances in 2011-12 Q1 to a pre-pandemic peak of 6.5 million in 2019-20 Q3, an abrupt contraction to 3.6 million in 2020-21 Q1 (the COVID-19 national lockdown period, shaded), a rapid recovery and overshoot through 2021-22, and continued growth to 7.24 million by 2026-27 Q1. Visual inspection indicates a clear non-stationary mean (a trending level) and no obvious variance instability that would require a log or Box-Cox transformation prior to differencing.

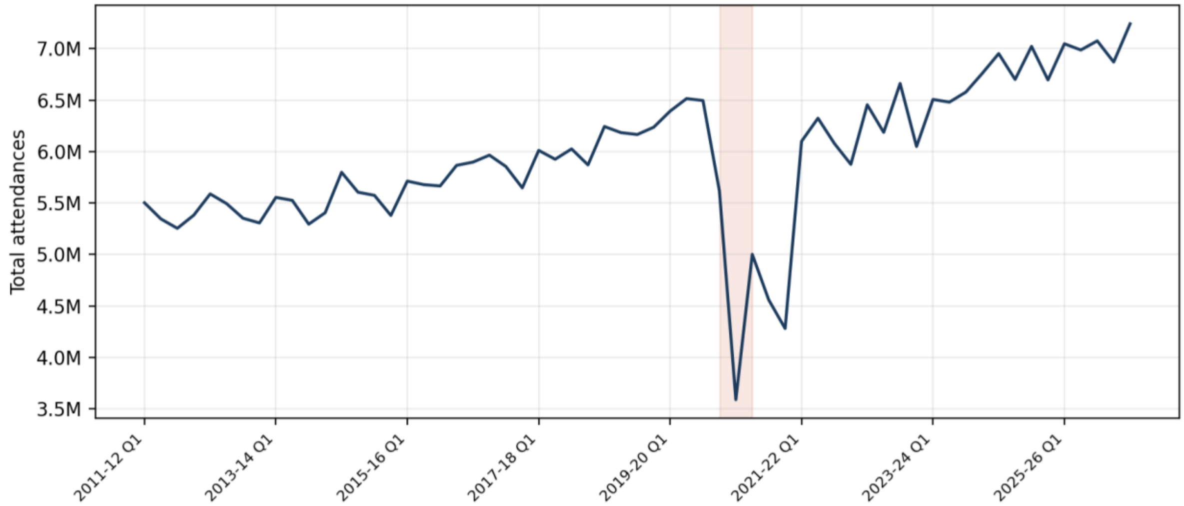

Figure 4.1. NHS England Quarterly Total A&E Attendances, 2011-12 Q1 – 2026-27 Q1
The shaded band marks the initial COVID-19 lockdown period, during which attendances fell sharply before recovering and overshooting pre-pandemic levels.

4.2 Stationarity Identification

An Augmented Dickey-Fuller (ADF) test on the level series (lag length selected by minimum AIC across lags 0–4) returns a test statistic of -1.74, which does not exceed the approximate 5% critical value of -2.91 for a series of this length with a constant term, the null hypothesis of a unit root cannot be rejected, confirming the series is non-stationary in levels. First-differencing produces an ADF statistic of -4.56, which exceeds the approximate 1% critical value of -3.55, rejecting the unit-root null and confirming stationarity after one order of differencing ($d = 1$). Figure 4.2 plots the first-differenced series, which fluctuates around a small positive mean with no visible trend, consistent with the ADF result.

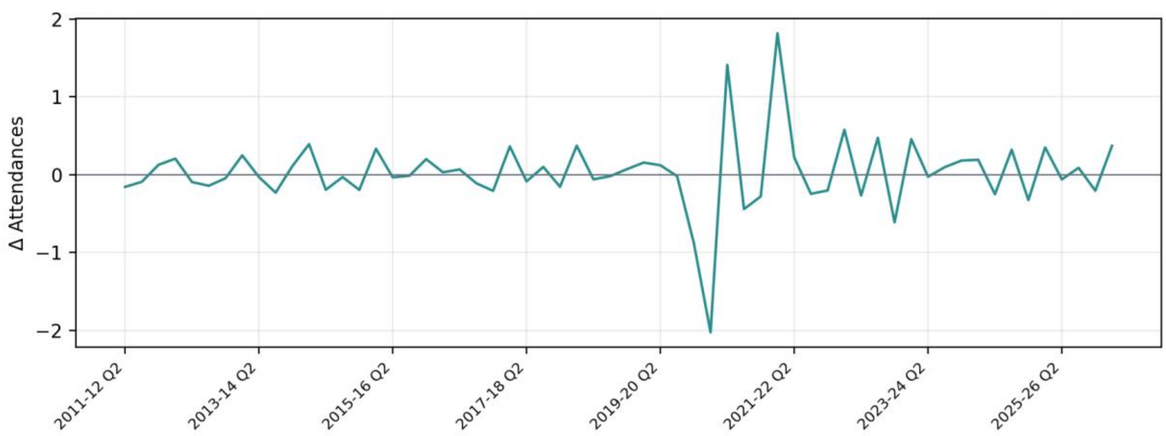

Figure 4.2. First-Differenced Quarterly Total A&E Attendances

The differenced series fluctuates around a small positive mean with no visible trend, consistent with the stationarity confirmed by the ADF test in Section 4.2.

4.3 Model Identification

Figure 4.3 presents the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the first-differenced series against 95% confidence bands

DOI: https://doi.org/10.64388/IREV10I3-1722770

(±0.253 for n = 60). Both functions show a single clearly significant spike at lag 1 (ACF = -0.251, PACF = -0.251) followed by values within or close to the confidence band at all subsequent lags, with a secondary borderline spike at lag 5 that does not repeat at the seasonal lag (lag 4) and is treated as sampling noise rather than a structural seasonal signature. A single significant lag-1 autocorrelation that does not decay geometrically is the textbook signature of a first-order moving-average process, indicating a tentative ARIMA(0,1,1) specification. For completeness, higher-order alternatives, ARIMA(1,1,0), (1,1,1), (2,1,1), (1,1,2), and (2,1,2), were also estimated and evaluated on an eight-quarter (two-year) holdout backtest, with results reported in Section 4.7.

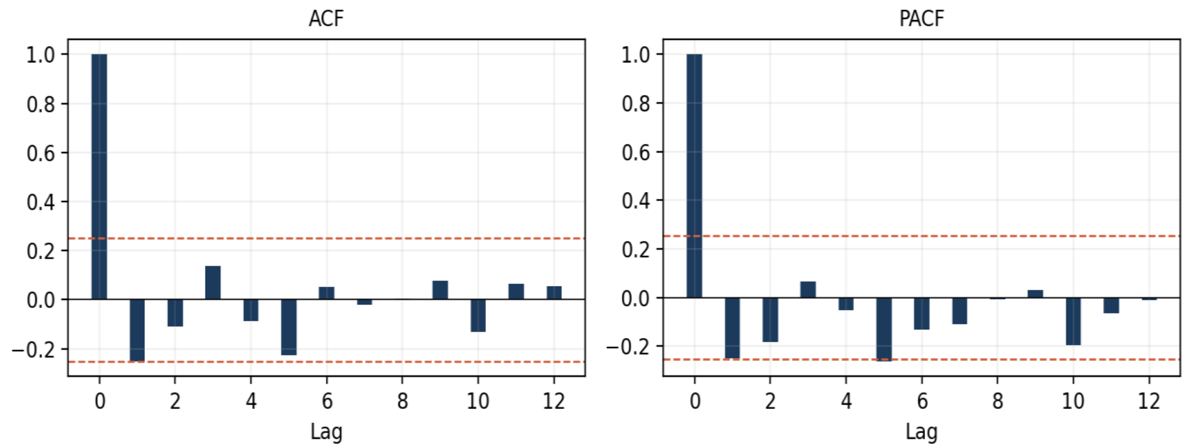

Figure 4.3. ACF and PACF of the First-Differenced Series (95% Confidence Bands)

Both functions show a single significant spike at lag 1 with no geometric decay, the signature used to identify the tentative ARIMA(0,1,1) specification in Section 4.3.

4.4 Model Estimation and Diagnostics

The ARIMA(0,1,1) model was estimated on the full 61-quarter sample by conditional-sum-of-squares maximum likelihood, yielding a constant (drift) term c = 28,348 attendances per quarter and a moving-average coefficient θ1 = -0.333 (within the invertibility bound |θ1| < 1). The residual variance is $\sigma^2 \approx 2.13 \times 10^{11}$. Ljung-Box portmanteau tests on the residuals show no evidence of remaining autocorrelation at any tested lag: Q(8) = 7.76 (df = 7), Q(12) = 9.32 (df = 11), and Q(16) = 10.43 (df = 15), all comfortably below their respective chi-squared critical values at the 5% level, confirming the residuals are statistically indistinguishable from white noise and the model is adequate for forecasting. Residuals show a mild negative skew (-1.53), attributable to the single large negative outlier at the 2020-21 Q1 COVID-19 contraction rather than a systematic model misspecification; this outlier is noted as a limitation in Section 6 rather than corrected for, since removing it would understate genuine tail risk in future crisis planning.

4.5 Five-Year ARIMA Forecast: 2026-27 Q2 – 2031-32 Q1

With model adequacy confirmed, a 20-quarter forecast was generated from the full-sample ARIMA(0,1,1) fit. As shown in Figure 4.4, the forecast follows a gentle, near-linear upward trajectory from 7,178,641 attendances in 2026-27 Q2 to 7,717,253 by 2031-32 Q1, a 7.5% increase over five years, driven almost entirely by the estimated per-quarter drift term rather than by autoregressive dynamics (the model has no AR component). Forecast uncertainty widens substantially over the horizon: the 95% prediction interval spans ±905,004 attendances in 2026-27 Q2 but ±5,333,674 by 2031-32 Q1, a more than five-fold expansion reflecting the compounding uncertainty inherent to a non-stationary, differenced-model

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

forecast projected twenty steps ahead. This widening interval is not a modelling weakness; it is an honest statistical statement that a single quarterly ARIMA model, however well-specified, cannot deliver a precise point forecast five years out, and its reporting is treated in this study as an ethical requirement of transparent forecasting practice (Pham, 2025) rather than an optional technical detail.

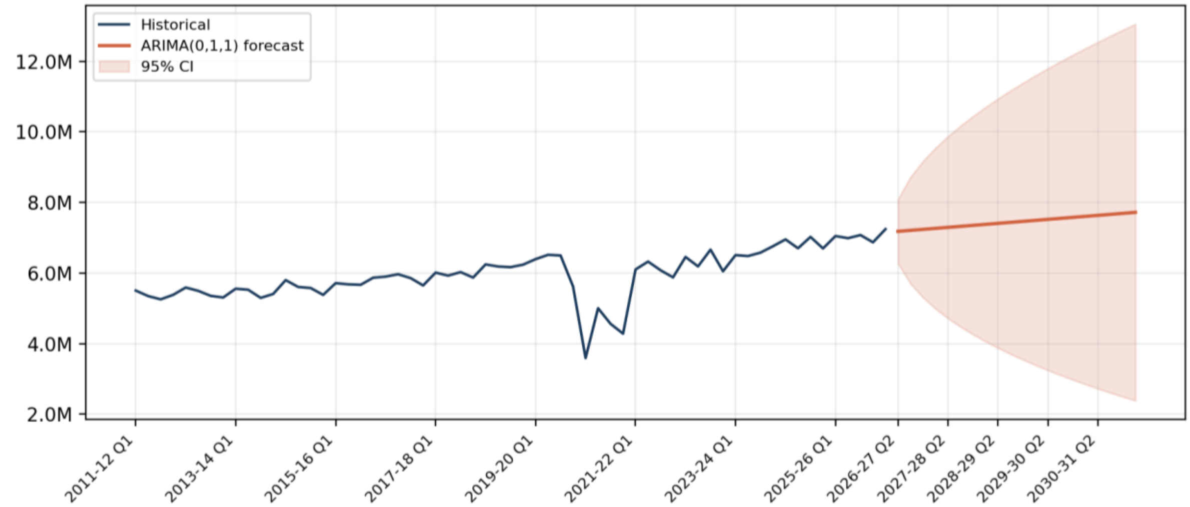

Figure 4.4. ARIMA(0,1,1) Five-Year Forecast of Quarterly A&E Attendances (2026-27 Q2 – 2031-32 Q1)

Shows the point forecast and its 95% prediction interval, which widens from ±905,000 to ±5.33 million attendances over the five-year horizon. Full forecast table in Appendix B.

4.6 Holt’s Linear Exponential Smoothing Forecast

As a benchmark, Holt’s LES was fitted to the same full 61-quarter series, with smoothing parameters optimised by minimising in-sample squared error: α = 0.739 (level smoothing) and β = 0.054 (trend smoothing). The low β indicates the estimated trend adjusts only slowly to recent changes in growth rate, yet the model still extrapolates a steeper linear trend (41,782 attendances per quarter) than the ARIMA drift term (28,348 per quarter), because Holt’s trend estimate is more heavily weighted toward the strong post-pandemic recovery growth visible in the second half of the series. As shown in Figure 4.5, this produces a forecast rising from 7,209,314 in 2026-27 Q2 to 8,003,175 by 2031-32 Q1, an 11.0% increase over five years, materially steeper than the ARIMA projection.

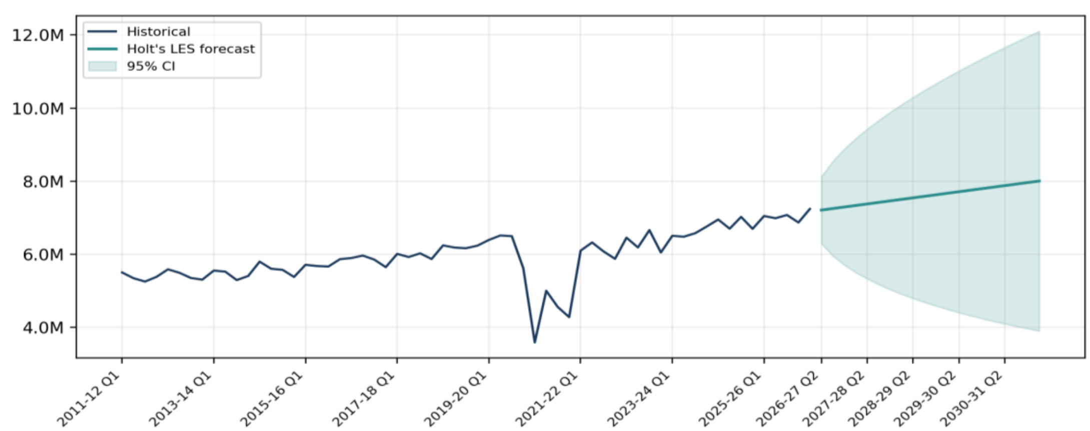

Figure 4.5. Holt’s Linear Exponential Smoothing Five-Year Forecast (2026-27 Q2 – 2031-32 Q1)

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

Shows the benchmark forecast and its 95% prediction interval, extrapolating a steeper linear trend than the selected ARIMA model.

4.7 Comparative Forecast and Model Selection

Table 4 summarises backtest accuracy for all six candidate specifications, evaluated by refitting each model on the first 53 quarters and forecasting the withheld final eight quarters (2024-25 Q3 – 2026-27 Q1).

Table 4: Backtest Model Comparison (Eight-Quarter Holdout, 2024-25 Q3 – 2026-27 Q1)

Reports out-of-sample accuracy (RMSE, MAE, MAPE) and information criteria (AIC, BIC) for all six candidate specifications, together with the model-selection decision for each.

Model	AIC	BIC	RMSE	MAE	MAPE	Decision
ARIMA(0,1,1)	1364.6	1368.5	154,563	127,861	1.86%	Selected
Holt's LES	n/a[a]	n/a[a]	198,757	151,591	2.22%	Competitive
ARIMA(1,1,0)	1365.7	1369.6	169,068	129,271	1.89%	Rejected
ARIMA(1,1,1)	1360.0	1365.9	524,144	466,647	6.63%	Rejected
ARIMA(2,1,1)	1362.0	1369.8	511,054	453,911	6.45%	Rejected
ARIMA(2,1,2)	1347.1[b]	1356.9[b]	290,024	252,499	3.59%	Rejected — overfit

Note. ARIMA(0,1,1) achieves the best out-of-sample accuracy on all three error metrics. [a] Holt's LES AIC/BIC derive from a different (ETS) likelihood basis and are not directly comparable; comparison relies on RMSE/MAE/MAPE (Hyndman & Athanasopoulos, 2021). [b] ARIMA(2,1,2) achieves the lowest in-sample AIC/BIC but the second-worst out-of-sample RMSE/MAE/MAPE among all candidates, a clear case of in-sample overfitting.

ARIMA(0,1,1) achieves the best out-of-sample accuracy on all three error metrics (RMSE, MAE, MAPE) among the six candidates (Table 4), and is materially simpler (one parameter) than the AIC/BIC-preferred ARIMA(2,1,2) (four parameters), which performs worse on every backtest metric despite its superior in-sample fit — direct evidence that the additional autoregressive and moving-average terms are fitting sample-specific noise rather than genuine structure. ARIMA(0,1,1) is therefore selected as the definitive forecasting model. Figure 4.6 overlays both the selected ARIMA forecast and the Holt's LES forecast against the full historical series; the two diverge by 285,922 attendances (3.7% of the ARIMA forecast level) at the end of the five-year horizon, illustrating that forecasting-technique selection alone carries a quarter-million-attendance planning consequence by 2031-32, a divergence of the same order of magnitude as a moderate winter surge.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

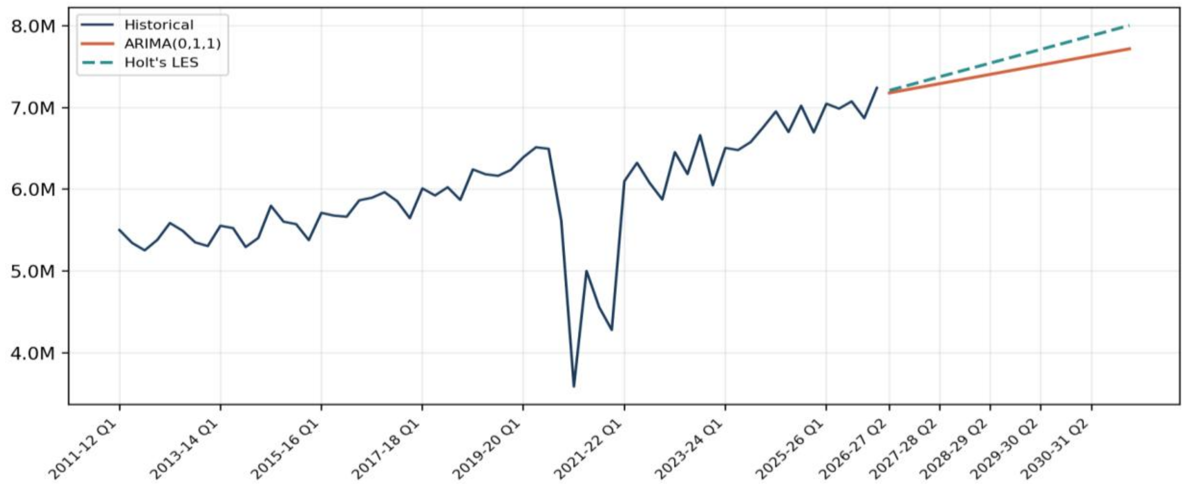

Figure 4.6. Comparative Forecast — ARIMA(0,1,1) versus Holt's LES (2026-27 Q2 – 2031-32 Q1)

Overlays both five-year forecasts against the full historical series, showing a 285,922-attendance divergence between the two techniques by the end of the horizon.

V. DISCUSSION: IMPLICATIONS FOR NHS OPERATIONAL AND POLICY DECISION-MAKING

The ARIMA(0,1,1) forecast corridor, 7.18 million to 7.72 million quarterly attendances through 2031-32 (Figure 4.4), carries direct operational and fiscal implications for NHS capacity planning. At the operational level, a system already recording bed occupancy above the 85% safety threshold and four-hour performance around 73-75% (Health Foundation, 2025) is being asked to absorb a further 7.5% rise in quarterly attendance demand over five years on the ARIMA forecast, or an 11.0% rise on the Holt's LES forecast, a divergence of 285,922 attendances by 2031-32 Q1 that is itself large enough to represent the difference between a manageable winter and a declared critical incident at national scale. This mirrors the broader forecasting literature's finding that technique selection is not a neutral technical choice but one with material downstream consequences (Kontopoulou et al., 2023).

The widening 95% prediction interval from ±905,000 attendances one quarter ahead to ±5.33 million five years ahead, is the most operationally consequential finding of this study. It implies that a single ARIMA point forecast is a defensible basis for near-term (one-to-four quarter) staffing and bed-capacity decisions but becomes progressively less reliable as a basis for capital planning, workforce commissioning, or estate expansion decisions with multi-year lead times. NHS planning bodies that anchor five-year capital plans to a single mean-forecast trajectory, without carrying the corresponding prediction interval into scenario planning, risk a structural misallocation of the kind that suppressed forecast uncertainty is known to produce in other high-stakes forecasting domains; the practical response is not to abandon ARIMA for longer horizons but to pair it with explicit low/central/high scenario bands, consistent with the transparency argument developed in Section 2.4 (Pham, 2025).

At the policy level, the recent national rollout of an AI-based A&E demand-forecasting tool to 50 NHS organisations (GOV.UK, 2025) demonstrates that government has already accepted the operational premise of this study, that better demand forecasting materially improves staffing and bed-space decisions. The present findings suggest two refinements to that direction of travel: first, that a transparent, auditable ARIMA baseline should sit alongside any more complex machine-learning system as a sanity-check and communication tool for non-technical decision-makers, consistent with the interpretability argument advanced by Kontopoulou et al. (2023); and second,

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1722770

that any operational forecasting tool deployed at trust level should surface its own prediction interval rather than a bare point forecast, so that local managers can distinguish a forecast they can safely staff against from one that requires contingency planning.

VI. CONCLUSIONS

This study set out to identify a statistically adequate forecasting model for NHS England quarterly A&E attendance demand, benchmark it against a simpler alternative, and translate the resulting forecast into concrete planning implications. Using 61 quarters of official NHS England data (2011-12 Q1 – 2026-27 Q1), the series was confirmed non-stationary in levels and stationery after first differencing. ACF/PACF analysis identified a single significant lag-1 autocorrelation, and a systematic backtest comparison across six candidate specifications selected ARIMA(0,1,1) as the best-generalising model, outperforming both a trend-extrapolating Holt's LES benchmark and a more heavily parameterised ARIMA(2,1,2) that overfit the training sample. Diagnostic testing confirmed the selected model's residuals are statistically indistinguishable from white noise across all tested lags.

The resulting five-year forecast projects a 7.5% rise in quarterly attendance demand, reaching approximately 7.72 million attendances by 2031-32 Q1, with a 95% prediction interval that widens more than five-fold over the forecast horizon. This uncertainty growth, not the point forecast itself, is the study's central practical finding: it delineates the horizon over which a univariate ARIMA forecast can responsibly inform NHS capacity commitments, and the horizon beyond which scenario-based planning becomes necessary. The 285,922-attendance divergence between the ARIMA and Holt's LES forecasts by the end of the horizon further demonstrates that forecasting-technique selection is itself a decision with material operational consequences, not a neutral analytical formality.

More broadly, the study demonstrates that a Box-Jenkins ARIMA methodology, developed and most commonly applied to commodity and financial time series, transfers cleanly to a healthcare demand series when the same rigour, stationarity testing, systematic model identification, out-of-sample backtesting, and residual diagnostics, is applied without shortcuts. The principal difference is not methodological but interpretive: where a commodity forecast informs a trading or hedging decision, a healthcare demand forecast informs staffing rotas, bed allocation, and ultimately patient safety, which is why this study has treated the prediction interval, and not only the point forecast, as a primary result.

VII. RECOMMENDATIONS

- NHS trust and system-level planning teams should adopt a parsimony-first model selection discipline as demonstrated here, a simpler ARIMA(0,1,1) specification outperformed a more complex ARIMA(2,1,2) on genuine out-of-sample accuracy despite the latter's superior in-sample fit, and any operational forecasting pipeline should include a holdout backtest step before a model is put into production.
- Forecast outputs used for capacity planning should always be presented with their prediction interval, not as a bare point estimate, particularly beyond a one-year horizon where this study found the interval width increases more than five-fold; dashboards and briefing materials for non-technical decision-makers should visualise the interval, not suppress it for presentational simplicity.
- Point forecasts beyond approximately four to eight quarters should be treated as directional and scenario-based rather than as single-number planning commitments; multi-year capital and workforce planning should be built around low/central/high demand scenarios rather than a single ARIMA trajectory.
- Where richer exogenous data is available at trust level (weather, local flu/RSV surveillance, bank holidays, known local population changes), the ARIMA baseline developed here should be extended with exogenous regressors (ARIMAX) or benchmarked against the feature-rich machine-learning approaches shown in the literature to outperform univariate ARIMA when such data exists (Tuominen et al., 2022); this national-level quarterly analysis is best read as a baseline for trust-level and shorter-horizon extensions, not a substitute for them.

- Any national forecasting tool, including the AI-based system currently being rolled out across NHS trusts (GOV.UK, 2025), should retain a transparent, interpretable ARIMA-class baseline as a communication and validation layer, so that operational managers without a data-science background can sanity-check and explain the forecasts they are asked to act on.

REFERENCES

[1] Bouderhem, R. (2024) 'Shaping the future of AI in healthcare through ethics and governance', Humanities and Social Sciences Communications, 11, p. 416.

[2] Box, G.E.P. and Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

[3] Farimani, R.M., et al. (2024) 'Models to predict length of stay in the emergency department: a systematic literature review and appraisal', BMC Emergency Medicine, 24(1), p. 54.

[4] GOV.UK (2025) Faster treatments and support for health workers as AI tackles A&E bottlenecks. Department for Science, Innovation and Technology. Available at: https://www.gov.uk/government/news/faster-treatments-and-support-for-health-workers-as-ai-tackles-ae-bottlenecks

[5] Health Foundation (2025) Did the NHS experience record pressures this winter? Available at: https://www.health.org.uk/reports-and-analysis/analysis/did-the-nhs-experience-record-pressures-this-winter

[6] House of Commons Library (2024) NHS readiness for winter 2024/25. Available at: https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0139/

[7] Hyndman, R.J. and Athanasopoulos, G. (2021) Forecasting: Principles and Practice. 3rd edn. Melbourne: OTexts.

[8] Kolambe, M. and Arora, S. (2024) 'Forecasting the future: a comprehensive review of time series prediction techniques', Journal of Electrical Systems, 20(2s), pp. 575–586.

[9] Kontopoulou, V.I., Panagopoulos, A.D., Kakkos, I. and Matsopoulos, G.K. (2023) 'A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks', Future Internet, 15(8), p. 255.

[10] Morbey, R.A., Todkill, D., Moura, P., Tollinton, L., Charlett, A., Watson, C. and Elliot, A.J. (2025) 'Using machine learning to forecast peak health care service demand in real-time during the 2022–23 winter season: a pilot in England, UK', PLOS ONE, 20(1), e0292829.

[11] NHS England (2026) A&E Attendances and Emergency Admissions: Quarterly Time Series [dataset]. Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/

[12] Pham, T. (2025) 'Ethical and legal considerations in healthcare AI: innovation and policy for safe and fair use', Royal Society Open Science, 12(5), p. 241873.

[13] Poursoltan, L., et al. (2025) 'Prospective comparison of econometric, machine learning, and foundation models for forecasting emergency department boarding patients', npj Health Systems, 2(1), p. 49.

[14] Silva, E., Pereira, M.F., Vieira, J.T., Ferreira-Coimbra, J., Henriques, M. and Rodrigues, N.F. (2023) 'Predicting hospital emergency department visits accurately: a systematic review', International Journal of Health Planning and Management, 38(4), pp. 904–917.

[15] Tuominen, J., Lomio, F., Oksala, N., Palomäki, A., Peltonen, J., Huttunen, H. and Roine, A. (2022) 'Forecasting daily emergency department arrivals using high-dimensional multivariate data: a feature selection approach', BMC Medical Informatics and Decision Making, 22, p. 121.

[16] Vollmer, M.A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S. and Bhatt, S. (2021) 'A unified machine learning approach to time series forecasting applied to demand at emergency departments', BMC Emergency Medicine, 21(1), pp. 1–14.