

Machine Learning Analysis of Risk Factors for Catheter-Associated Urinary Tract Infections in Stroke Patients and Their Impact on Healthcare Quality

Lu Lu^{1,2}, Xiaojun Li^{2,3}, Donghao Cai^{2,4}, Chuangchuang Mei^{2,4}, Xiaojing Hong^{2,4}, Jiewen Xie^{2,5}

¹Department of Lujingdong Clinic, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China; ²Guangdong Provincial Key Laboratory of Research and Development in Traditional Chinese Medicine, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China; ³Department of Nosocomial Infection, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China; ⁴Department of Laboratory Medicine, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China; ⁵Baiyun Apartment Department, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China

Correspondence: Jiewen Xie, Baiyun Apartment Department, Guangdong Provincial Second Hospital of Traditional Chinese Medicine, Guangzhou, People's Republic of China, Email 80198975@qq.com

Purpose: Catheter-associated urinary tract infections (CAUTIs) represent a common complication among hospitalised stroke patients, exacerbating clinical outcomes and increasing healthcare burdens. This study aims to identify key risk factors for CAUTI in stroke patients using a random forest algorithm, construct and validate a clinical prediction model, and investigate the impact of CAUTIs on healthcare quality.

Methods: Data were retrospectively collected from 7486 stroke patients with indwelling urinary catheters at Guangdong Provincial Second Hospital of Traditional Chinese Medicine between January 2020 and December 2024. Patients were categorised into infection and non-infection groups based on CAUTI occurrence. Healthcare quality indicators including length of stay and hospitalisation costs were compared between groups. Random forest algorithm screening identified variables influencing infection, with a decision tree model constructed. Model performance was assessed using area under Receiver Operating Characteristic curve (AUC), calibration curve, and decision curve analysis (DCA).

Results: Among 7486 patients, 248 developed CAUTI. Hospitalisation duration and costs were significantly higher in the infection group. Random forest identified three potential predictors: age, days of catheter use (DaysCAU), and time from admission to catheter insertion (TTCAU). Multivariate logistic regression confirmed age and DaysCAU as independent risk factors for CAUTI. The constructed nomogram model demonstrated an AUC of 0.794 in the training cohort and 0.778 in the validation cohort. Calibration curves indicated good agreement between predicted and actual values, while DCA confirmed significant clinical net benefit within the 5–20% risk interval.

Conclusion: Secondary CAUTI in stroke patients significantly prolongs hospital stays and increases healthcare costs. The predictive model based on age and catheterisation duration demonstrates favourable discriminatory performance and generalisability. It provides a practical tool for clinicians to identify high-risk patients early and formulate individualised prevention strategies, holding significant implications for enhancing healthcare quality.

Keywords: stroke, catheter-associated urinary tract infection, nomogram, risk prediction, healthcare quality

Introduction

Stroke constitutes a significant global public health issue, ranking among the primary causes of mortality and disability worldwide. The Global Burden of Disease study indicates that stroke is the second leading cause of death and the third leading cause of disability among non-communicable diseases. By 2021, the global stroke patient population had reached 93.8 million, with an annual healthcare burden exceeding US\$890 billion, imposing substantial strain upon national healthcare systems and socio-economic structures.^{1–3} Stroke patients frequently require invasive procedures such as indwelling urinary catheterisation to



manage symptoms including urinary retention, impaired consciousness, or severe mobility limitations, thereby facilitating clinical care. Conversely, these patients often undergo prolonged catheterisation and frequently present with underlying conditions such as advanced age and diabetes mellitus. These factors collectively render stroke patients a high-risk cohort for CAUTI.⁴⁻⁶ The occurrence of CAUTI not only substantially increases the incidence of severe complications such as sepsis but also delays neurological recovery and prolongs rehabilitation periods. It significantly elevates ICU occupancy rates, lengthens hospital stays, and increases healthcare costs, severely compromising patient prognosis and quality of life.⁷ Ultimately, this undermines the therapeutic goal of achieving optimal functional independence in post-stroke patients.

Clinical practice demonstrates⁸ that early identification and timely intervention for high-risk populations can effectively reduce infection rates and improve patient outcomes. However, current clinical management of CAUTI in stroke patients relies predominantly on empirical assessment, lacking precise, quantitative risk prediction tools. This results in insufficiently targeted interventions, delayed implementation, and challenges in achieving personalised prevention. With the growing adoption of precision medicine principles, developing scientifically robust risk prediction models has become a key focus in clinical research. The rapid advancement of machine learning algorithms provides substantial support for this objective. As an efficient machine learning algorithm, random forest constructs multiple decision trees through resampling and integrates their results. It effectively addresses collinearity issues in high-dimensional data and precisely measures variable importance through the mean reduction in Gini index, demonstrating significant advantages in clinical predictive model development. This algorithm has been successfully applied to predict risks of poor functional outcomes and haemorrhagic transformation in stroke patients, providing crucial guidance for clinical decision-making.^{9,10} Baffin charts offer an intuitive graphical representation of complex predictive models, enabling rapid calculation of individual patient risk probabilities through cumulative scoring of predictive indicators. This approach has matured applications in stroke-related bloodstream infections and prognosis assessment, substantially enhancing clinical model utility.^{11,12}

Although existing research has confirmed the adverse impact of CAUTI on stroke patients, systematic studies on screening for risk factors of CAUTI in stroke patients, constructing precise prediction models, and assessing related healthcare quality impacts remain scarce. Consequently, this study collected clinical data from stroke patients at Guangdong Provincial Second Hospital of Traditional Chinese Medicine between 2020 and 2024. Utilising a random forest algorithm, we screened key risk factors for CAUTI, constructed and validated a clinical prediction model, and simultaneously examined the impact of CAUTI on healthcare quality indicators such as patient length of stay and medical costs. The aim is to provide scientific evidence for the early clinical identification of high-risk individuals and the formulation of personalised prevention strategies, thereby reducing CAUTI incidence and enhancing the quality of healthcare services for stroke patients.

Materials and Methods

Patient Data Clinical data were collected from stroke patients admitted to Guangdong Provincial Second Hospital of Traditional Chinese Medicine between January 2020 and December 2024. Patients were screened based on diagnostic inclusion and exclusion criteria.

Participants meeting all the following criteria were enrolled:

1. Confirmed diagnosis of stroke;¹³
2. Age ≥ 18 years;
3. Clinical indication for indwelling urinary catheter placement during hospitalisation;
4. Complete clinical and laboratory data meeting criteria for CAUTI diagnosis¹⁴ and study predictor variables. The diagnostic criteria for CAUTI are urinary tract infections occurring during the period of indwelling catheterisation or within 48 hours of catheter removal. The diagnostic criteria for urinary tract infection are: the presence of urinary irritation symptoms such as frequency, urgency and dysuria, or tenderness in the lower abdomen and percussion pain in the renal region, with or without fever, accompanied by one of the following: a urine analysis showing white blood cells of ≥ 5 per high-power field in males or ≥ 10 per high-power field in females, or the isolation of viable microorganisms from a culture; All patients with no history of asymptomatic bacteriuria who met the diagnostic criteria for CAUTI following catheterisation were included in the study.

Exclusion criteria:

1. Pre-existing urinary tract infection at admission or prior to catheterisation;
2. Catheter indwelling duration <48 hours;
3. Pregnancy or lactation;
4. Participation in other interventional clinical trials potentially affecting infection risk.

The screening identified a total of 7486 eligible cases. Patients who met the aforementioned diagnostic criteria for CAUTI, and in whom pre-admission or admission-related urinary tract infections, as well as urinary tract colonisation secondary to infections in other sites, were excluded, were classified into the infection group; those who did not develop CAUTI during their hospital stay were classified into the non-infection group. The study workflow is detailed in Figure 1. This research was approved by the Ethics Review Committee of Guangdong Provincial Second Hospital of Traditional Chinese Medicine (Ethics Review Opinion No.: Z202404-002-01).

Data Collection

The following demographic and clinical data were collected: age, gender; presence within the preceding 30 days of admission to an intensive care unit, surgical history, antibiotic use, invasive procedures such as mechanical ventilation-(MV) or Peripherally Inserted Central Venous Catheters(PICC); days catheter in use (The number of days prior to infection during which the catheter was in use for CUATI patients, DaysCAU); time to catheterisation after admission (TTCAU); Pre-infection underlying conditions including pneumonia, hypertension, heart failure(HF), diabetes, Chronic

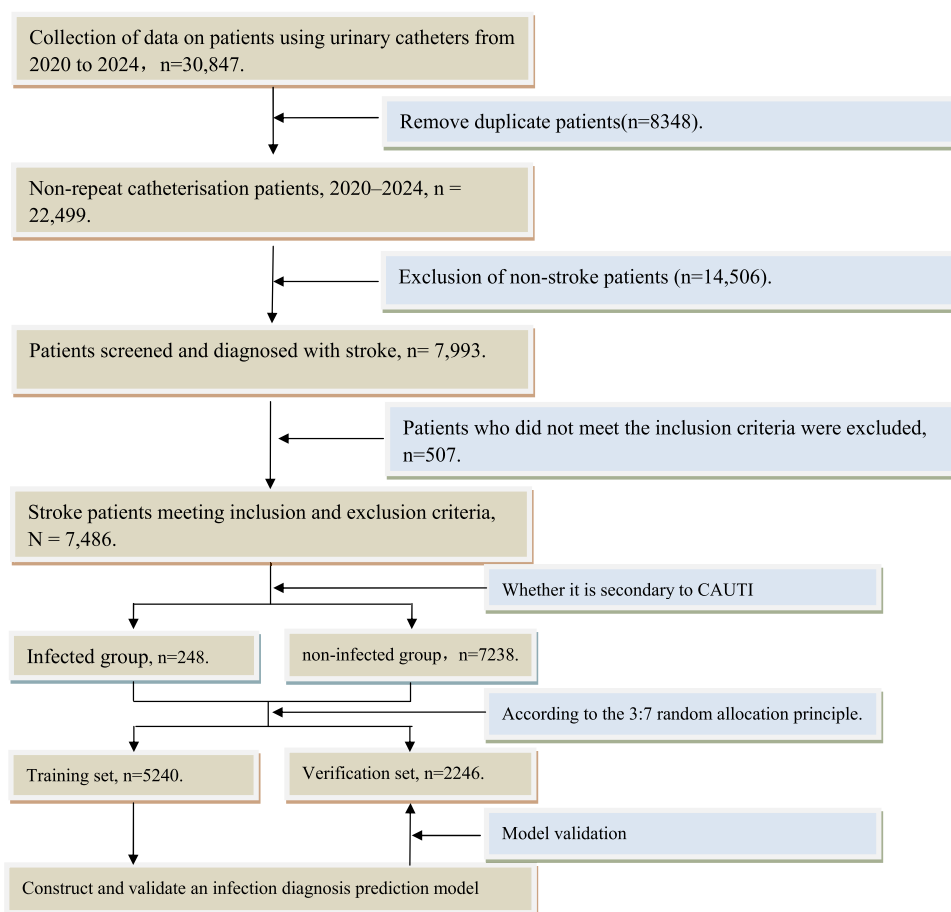


Figure 1 Flowchart of the Secondary Cauti Study in Stroke Patients.

Abbreviation: CAUTI, Catheter-associated urinary tract infections.

kidney insufficiency(CKI), cholecystitis, anaemia, hypoalbuminaemia, electrolyte disturbance(ED), hyperlipidaemia, and thyroid dysfunction. All variables underwent collinearity analysis, with variance inflation factors (VIF) ranging between 1 and 3.5, indicating no high correlation among the analysed independent variables.

Statistical Analysis

The collected dataset was divided into training and validation sets in a 7:3 ratio, with comparisons conducted across variables. Non-normally distributed data were presented using median (interquartile range). In univariate analysis, categorical variables underwent chi-square tests, while continuous variables were assessed using Mann–Whitney *U*-tests. Model performance was assessed using receiver operating characteristic (ROC) curves and calibration curves. Decision curve analysis (DCA) was concurrently employed to determine the net benefit threshold for predictions. Results with *p*-values < 0.05 were considered statistically significant. All statistical analyses were performed using R software (version 4.3.1) and MSTATA software (www.mstata.com). The code can be found in [Supplementary File 1](#).

Results

Impact of Secondary CAUTI on Patient Outcomes

Univariate analyses were conducted on outcomes, hospital costs, and length of stay for both groups, stratified by the occurrence of secondary CAUTI. Results indicated that following CAUTI in stroke patients, the median length of hospital stay increased from 15 days to 27 days, while the median hospital cost rose from ¥34,031.05 to ¥55,535.21. Both differences were statistically significant ($P < 0.001$), whereas there was no significant difference in discharge mortality rates between the two groups ($P = 0.159$) ([Table 1](#)).

Differences in Indicators Between Outcome Groups Within the Training Set

Univariate analysis revealed that within the overall dataset, the following factors constituted risk factors for CAUTI secondary to catheter use in stroke patients: antibiotic administration prior to catheter insertion, prolonged duration of catheter use, extended time from admission to catheter insertion, and underlying conditions such as heart failure, diabetes mellitus, and anaemia. The results from the training set are broadly consistent with the overall dataset. Detailed results are presented in [Tables 2](#) and [3](#).

Variable Selection Using the Random Forest Algorithm

All candidate predictors were incorporated into the equation, with an initial model constructed through 500 resampling iterations. Variable importance was measured by the average reduction in Gini index during node splitting, the overall error rate gradually decreases and stabilises as the number of trees increases. Univariate analysis revealed that age, DaysCAU, and TTCAU were all associated with CAUTI ($P < 0.05$), establishing a basis for candidate predictor selection. The random forest algorithm identified these three variables as the top three potential predictors, confirming their significance. Although TTCAU lacked statistical significance, all three were incorporated into a multivariate logistic regression model for independent risk factor screening to ensure the model's completeness and clinical utility. Ultimately identifying three potential predictors: days of catheter use, time from admission to catheter insertion, and age ([Figure 2](#)).

Table 1 Analysis of Healthcare Quality for Patients with Secondary CAUTI

Characteristic	Total (n = 7486)	Non-Infection Group, (n =7238)	INFECTION GROUP (N =248)	P
Death Discharge, n (%)	584 (8)	571 (8)	13 (5)	0.159
Hospitalisation costs, Median (Q1,Q3)	34,600.79 (20,914.04, 60,654.27)	34,031.05 (20,749.16, 59,014.87)	55,535.21 (34,990.11, 98,423.5)	< 0.001
Length of hospital stay, Median (Q1,Q3)	15.5 (12, 24)	15 (12, 23)	27 (19, 34.25)	< 0.001

Table 2 Patient Demographics and Baseline Characteristics

Characteristic	Total (n = 7486)	Non-Infection Group, (n =7238)	Infection Group (n =248)	P
Age, Median (Q1,Q3)	71 (59, 81)	71 (59, 81)	73 (60, 84)	0.113
Male, n (%)	4480 (60)	4329 (60)	151 (61)	0.784
ICUstay, n (%)	882 (12)	850 (12)	32 (13)	0.648
Surgical, n (%)	1190 (16)	1156 (16)	34 (14)	0.385
Antibiotic, n (%)	5789 (77)	5565 (77)	224 (90)	< 0.001
DaysCAU, Median (Q1,Q3)	12 (6, 18)	12 (5, 18)	25 (15.75, 32)	< 0.001
TTCAU, Median (Q1,Q3)	0 (0, 4.53)	0 (0, 4)	0 (0, 12)	< 0.001
MV, n (%)	961 (13)	925 (13)	36 (15)	0.479
PICC, n (%)	3840 (51)	3697 (51)	143 (58)	0.048
Pneumonia, n (%)	3480 (46)	3358 (46)	122 (49)	0.421
Hypertensive, n (%)	5390 (72)	5204 (72)	186 (75)	0.318
HF, n (%)	366 (5)	345 (5)	21 (8)	0.012
Diabetes, n (%)	2469 (33)	2370 (33)	99 (40)	0.022
CKI, n (%)	711 (9)	687 (9)	24 (10)	0.922
Cholecystitis, n (%)	400 (5)	380 (5)	20 (8)	0.073
Anaemia, n (%)	1043 (14)	995 (14)	48 (19)	0.016
Hypoproteinaemia, n (%)	709 (9)	683 (9)	26 (10)	0.657
Electrolyte disturbance, n (%)	571 (8)	554 (8)	17 (7)	0.730
Hyperlipidaemia, n (%)	289 (4)	280 (4)	9 (4)	0.980
Thyroiddysfunction, n (%)	483 (6)	461 (6)	22 (9)	0.148

Abbreviations: DaysCAU,days catheter in use; TTCAU, time to catheterisation after admission;MV, mechanical ventilation; PICC, peripherally inserted central catheter;HF,heart failure; CKI,Chronic kidney insufficiency.

Table 3 Univariate Analysis Results for Patients with Different Outcomes in the Training Set

Variables	Total (n = 5240)	Non-Infection Group (n = 5058)	Infection Group (n = 182)	P
Age, Median (Q1,Q3)	71 (59, 81)	71 (59, 81)	74 (60, 84)	0.061
Male, n (%)	3136 (60)	3022 (60)	114 (63)	0.481
ICUstay, n (%)	642 (12)	614 (12)	28 (15)	0.231
Surgical, n (%)	838 (16)	810 (16)	28 (15)	0.901
Antibiotic, n (%)	4061 (78)	3897 (77)	164 (90)	<0.001
DaysCAU, Median (Q1,Q3)	13 (6, 18)	12 (5, 18)	25 (16, 33.75)	<0.001
TTCAU, Median (Q1,Q3)	0 (0, 4.58)	0 (0, 4.53)	0 (0, 12.6)	0.033
MV, n (%)	691 (13)	662 (13)	29 (16)	0.316
PICC, n (%)	2710 (52)	2603 (51)	107 (59)	0.062
Pneumonia, n (%)	2441 (47)	2354 (47)	87 (48)	0.795
Hypertensive, n (%)	3762 (72)	3625 (72)	137 (75)	0.328
HF, n (%)	255 (5)	237 (5)	18 (10)	0.002
Diabetes, n (%)	1731 (33)	1655 (33)	76 (42)	0.014
CKI, n (%)	495 (9)	478 (9)	17 (9)	0.961
Cholecystitis, n (%)	270 (5)	256 (5)	14 (8)	0.159
Anaemia, n (%)	744 (14)	706 (14)	38 (21)	0.012
Hypoproteinaemia, n (%)	514 (10)	496 (10)	18 (10)	0.971
Electrolyte disturbance, n (%)	398 (8)	386 (8)	12 (7)	0.706
Hyperlipidaemia, n (%)	207 (4)	199 (4)	8 (4)	0.904
Thyroid dysfunction, n (%)	337 (6)	327 (6)	18 (10)	0.067

Abbreviations: DaysCAU,days catheter in use; TTCAU, time to catheterisation after admission;MV, mechanical ventilation; PICC, peripherally inserted central catheter;HF,heart failure; CKI,Chronic kidney insufficiency.

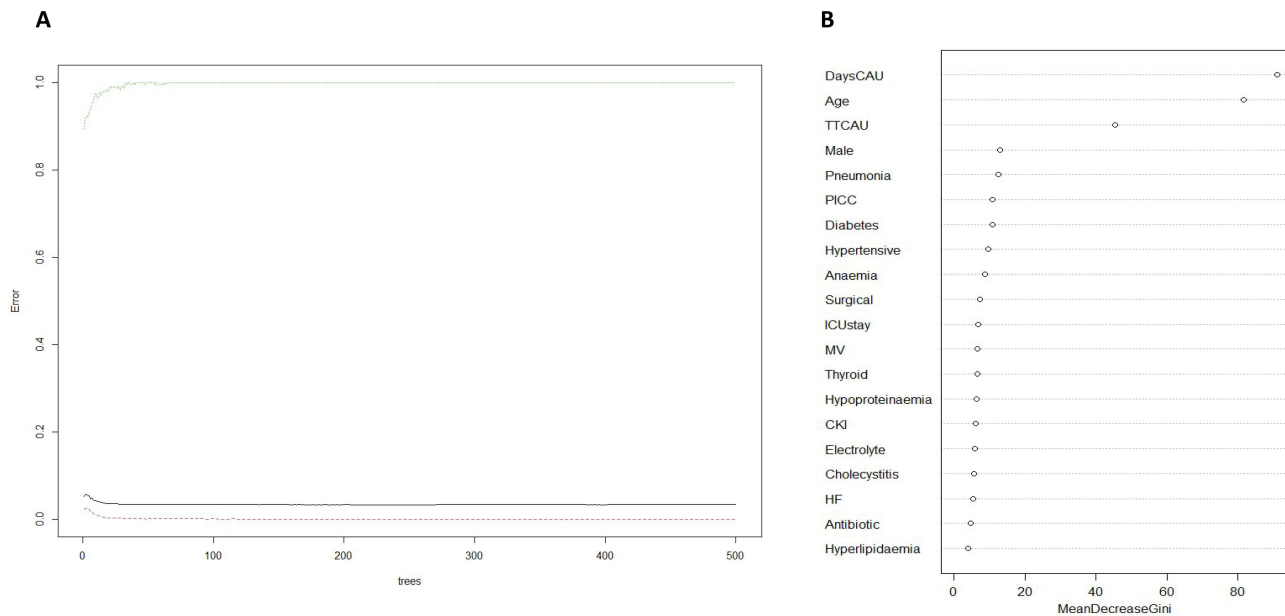


Figure 2 Random Forest Feature Selection Results. **(A)** The relationship between the overall error rate and the number of trees. **(B)** Variable Importance Ranking. Figure **(A)** shows the relationship between the number of decision trees in a random forest model and the model's overall error rate; Figure **(B)** shows the ranking of variable importance based on the mean decrease in Gini.

Abbreviations: DaysCAU, days catheter in use; TTCAU, time to catheterisation after admission; MV, mechanical ventilation; PICC, peripherally inserted central catheter; HF, heart failure; CKI, Chronic kidney insufficiency.

Constructing the Multivariate Logistic Regression Analysis Equation

A linear analysis was conducted on three variables based on whether they were secondary infections. The results indicated that the R-value for the linear analysis of variables such as DaysCAU and age was 0.231, demonstrating a significant linear relationship between the independent and dependent variables. This supports the use of linear variables within the model.

The final logistic regression model comprised three independent predictors (age, DaysCAU, and TTCAU). Age was significantly associated with the probability of outcome occurrence (odds ratio 1.01, 95% confidence interval 1.00 to 1.02, $p=0.029$). The DaysCAU variable demonstrated a more pronounced effect, exhibiting a significant positive correlation (OR=1.08, 95% CI 1.07–1.09, $p<0.001$); conversely, the TTCAU variable failed to achieve statistical significance (OR=1.00, 95% CI 0.98–1.01, $p=0.751$) (Table 4).

Constructing a Nomogram

The final logistic regression model incorporates three independent predictors (age, DaysCAU, and TTCAU) and has been developed into a user-friendly predictive chart. The nomogram indicates that the risk of CAUTI in stroke patients increases with advancing age and prolonged catheterisation duration; earlier catheterisation following admission also elevates the risk score, as illustrated in Figure 3.

Table 4 Results of Multivariate Logistic Regression for Training Cohort

Characteristic	N	Event N	OR	95% CI	p-value
Age	5240	182	1.01	1.00, 1.02	0.029
DaysCAU	5240	182	1.08	1.07, 1.09	<0.001
TTCAU	5240	182	1.00	0.98, 1.01	0.751

Abbreviations: CI, Confidence Interval; OR, Odds Ratio; DaysCAU, days catheter in use; TTCAU, time to catheterisation after admission.

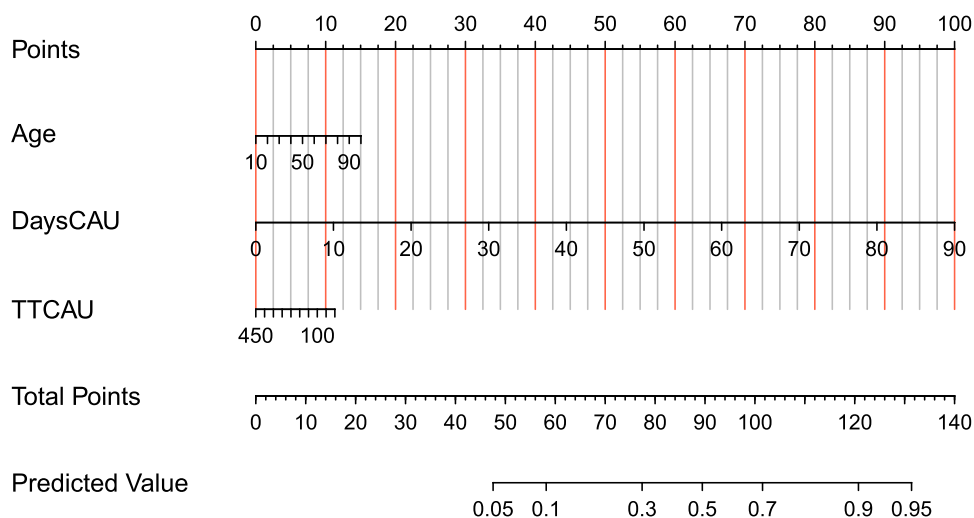


Figure 3 A nomogram model for predicting the risk of CAUTI in stroke patients. On the left-hand side of the model are the scoring scales for each predictor variable; in the centre is the scale for the converted total score (Total Points); and on the right-hand side is the scale for the predicted risk of CAUTI (Predicted Value).

Abbreviations: DaysCAU, days catheter in use; TTCAU, time to catheterisation after admission.

Model Validation

Cross-Validation Results of the Model on the Entire Dataset

To validate the model's robustness, 5-fold cross-validation was performed 400 times on the full dataset. Results indicate an average AUC of 0.763, with a 95% confidence interval of 0.628–0.800 (Figure 4).

ROC for Predictive Models

ROC analysis revealed that the variable “DaysCAU” demonstrated the highest discriminatory efficacy in predicting CAU in catheterised stroke patients, with an AUC of 0.790 (95% confidence interval: 0.758–0.823); The AUC values for the remaining two variables, age and the TTCAU predictive factor, demonstrated moderate discriminatory performance

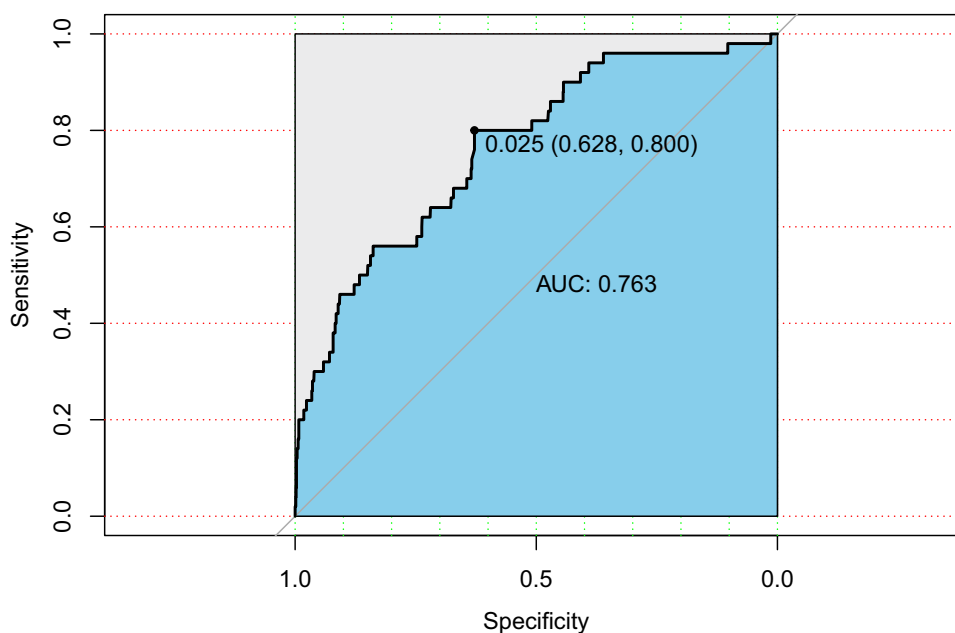


Figure 4 AUC results for the model across 400 rounds of 5-fold cross-validation on the entire dataset. The horizontal axis represents the false positive rate, and the vertical axis represents the true positive rate.

Abbreviation: AUC, Area under the curve.

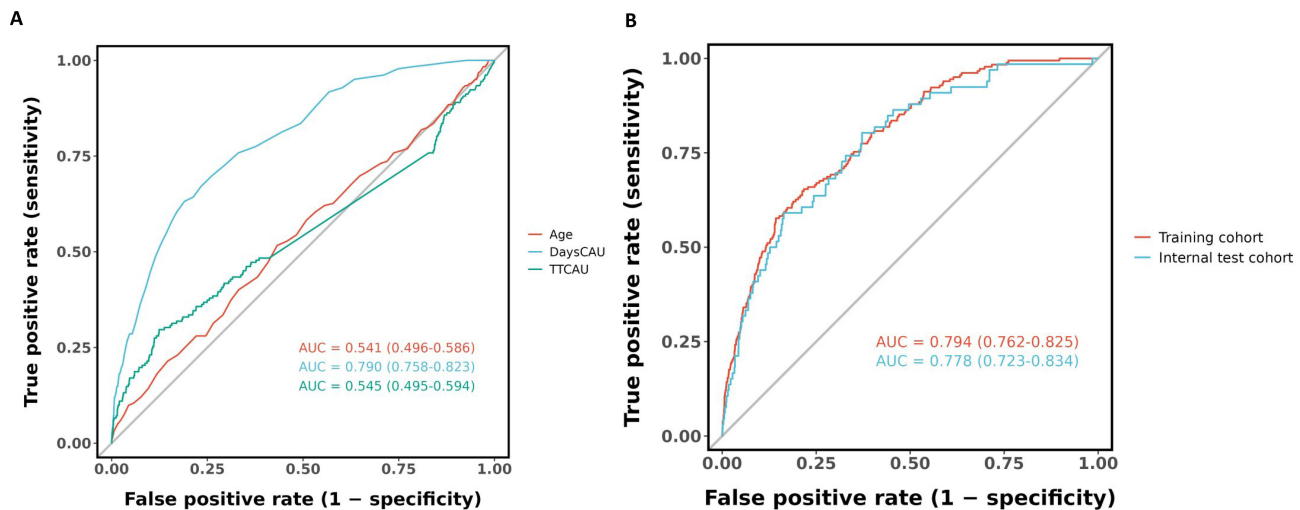


Figure 5 ROC curves for the predictive model. **(A)** ROC curves for individual variables within the training set; **(B)** Model ROC curves across different datasets. The horizontal axis represents the false positive rate, and the vertical axis represents the true positive rate.

Abbreviation: AUC, Area under the curve.

(Figure 5A). Among the model's three variables, optimal cut-off points were identified at: DaysCAU = 20.5 days (sensitivity 83.2%, specificity 81.0%), TTCAU = 10.5 days (sensitivity 69.8%, specificity 86.1%), and age = 73.5 years (sensitivity 51.6%, specificity 56.8%) and age = 73.5 years (sensitivity 51.6%, specificity 56.8%). When the three variables reach their optimal threshold, the corresponding score on the column chart is approximately 45 points, which may serve as a suitable intervention point. Among the AUC values across different cohorts, the training cohort yielded 0.794 (95% confidence interval: 0.762–0.825), while the test set produced 0.778 (95% CI: 0.723–0.834), indicating the model possesses sound discriminatory capability and generalisability (Figures 5B).

Model Calibration Curves

Calibration curves were plotted using 500 resamples. The model yielded Hosmer-Lemeshow test p -values of 0.133 and 0.674 across the two datasets respectively, both exceeding 0.05; The training set Brier score is 0.032, and the test set is 0.028. This indicates no significant divergence between predicted and actual values. The calibration curve results for both the training and test datasets indicate that the curves for both datasets are close to the ideal curve. This demonstrates that the predicted outcomes align with the actual findings, indicating that both datasets exhibit good correlation within the model (Figures 6).

Decision Curve Analysis of the Model

Figure 7 presents the DCA curve associated with this model. Within the DCA, both the training and test sets demonstrate favourable net benefit within the 5–20% risk interval. Within the 5–10% risk range, the decision curve exhibits a broader range of benefits, demonstrating significant net benefit in clinical practice. Within this zone, clinical intervention yields the greatest benefit for high-risk patients, such as elderly individuals with prolonged catheter use. Within this range, 10% of the 100 patients with a 10% probability of stroke who use catheters are highly likely to develop CAUTI.

Discussion

This study, involving 7486 stroke patients undergoing indwelling urinary catheterisation, successfully constructed and validated a nomogram-based predictive model for CAUTI in stroke patients. This was achieved by screening risk factors using a random forest algorithm and integrating them with multivariate logistic regression. The study also demonstrated the significant impact of CAUTI on healthcare quality. The findings not only enrich the clinical evidence for predicting CAUTI risk in stroke patients but also provide a practical tool for precise clinical prevention and control.

The findings of this study indicate that stroke patients who develop CAUTI experience a significant prolongation of hospital stay and a marked increase in hospital costs; however, there was no significant difference in discharge mortality rates between the two groups. This finding aligns with Hutton et al's conclusion¹⁵ that CAUTI, as a common hospital-

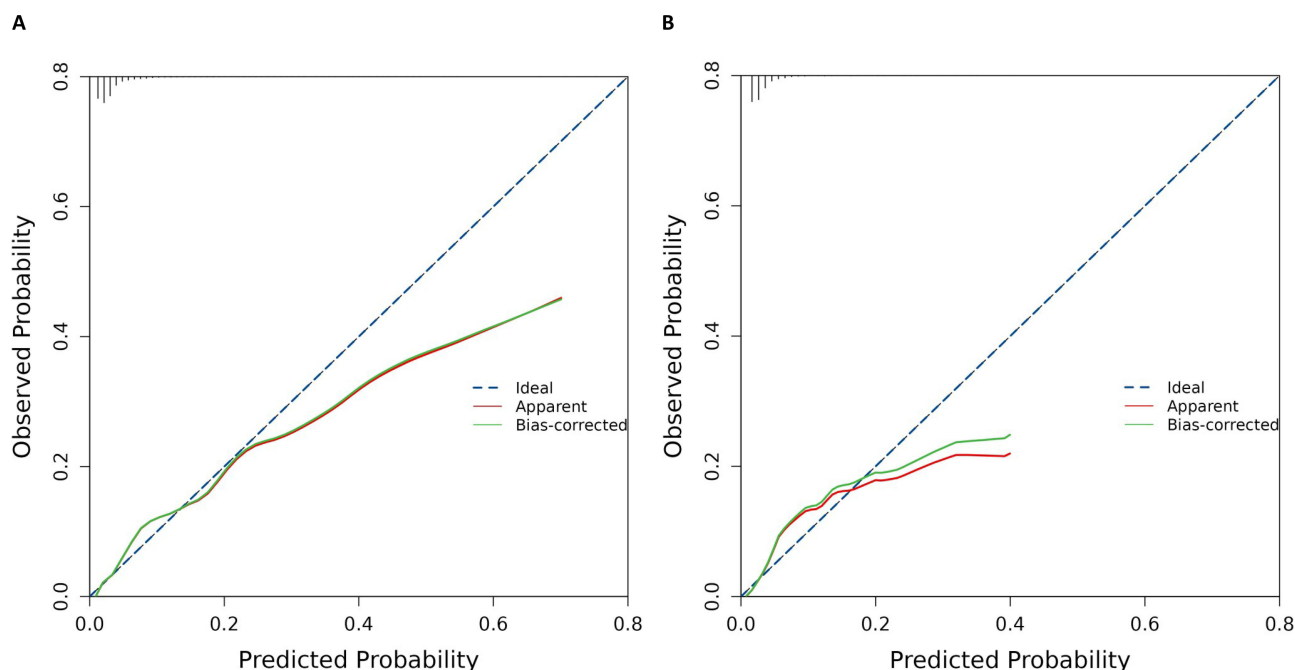


Figure 6 Calibration curves of the model. **(A)** Calibration curve of the model on the training set; **(B)** Calibration curve of the model on the test set. The horizontal axis shows the probability of CAUTI predicted by the model, whilst the vertical axis shows the actual observed probability of CAUTI. The solid line in the figure represents the model's actual calibration curve, whilst the dotted line represents the ideal calibration curve. The closer the two curves fit together, the better the model's calibration performance, and the greater the consistency between the predicted and actual values.

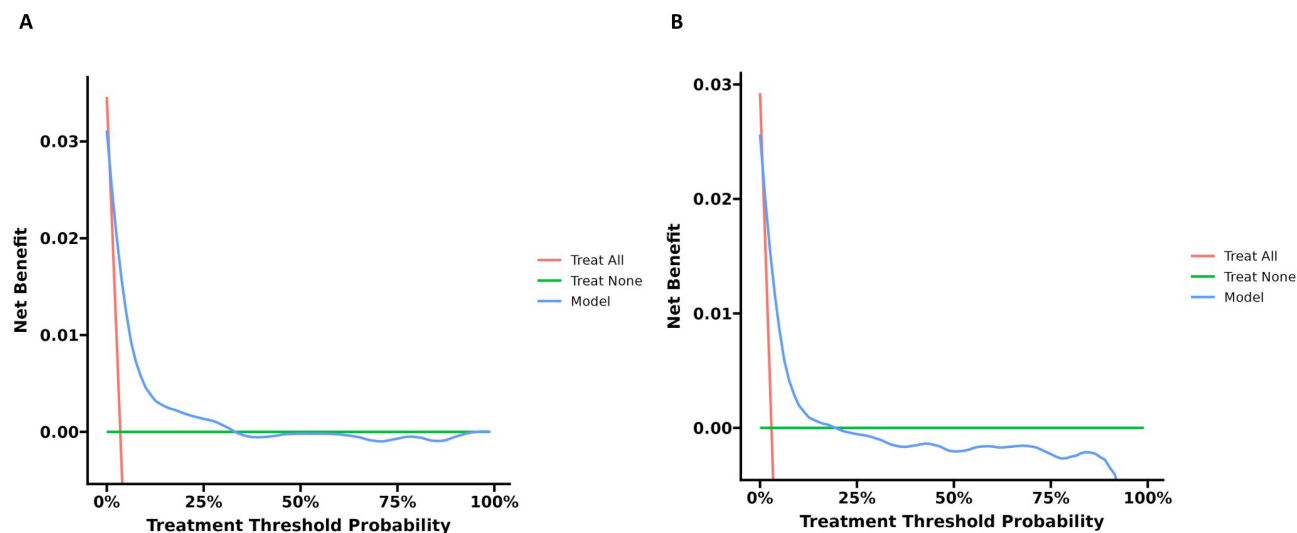


Figure 7 DCA curves for the model. **(A)** DCA curve for the model on the training set; **(B)** DCA curve for the model on the test set. The horizontal axis represents the threshold probability, whilst the vertical axis represents the net benefit rate. The “Model” curve in the figure depicts the net benefit curve of the model constructed in the study; the “Ideal” curve represents the net benefit curve of the ideal model; and the horizontal line indicates the net benefit curve of the no-intervention model.

acquired infection in stroke patients, directly increases healthcare resource consumption by prolonging hospital stays. Concurrently, the local inflammatory response triggered by infection may indirectly impair neurological recovery, reducing rehabilitation efficiency and further exacerbating the healthcare burden.⁷ Moreover, numerous studies^{16–18} indicate a positive correlation between the number of comorbidities in stroke patients and hospital costs. As a significant secondary complication, CAUTI necessitates treatment adjustments that further elevate healthcare expenditure, corroborating the significantly higher medical costs observed in the infection group within this study. Notably, no significant difference in mortality rates was observed between the two groups in this study, potentially attributable to timely clinical

intervention for CAUTI. However, this does not diminish the harm caused by CAUTI—its core impact is more evident in patient recovery processes, quality of life, and the efficiency of healthcare resource allocation. This conclusion provides direct evidence for prioritising CAUTI prevention and control in clinical practice to reduce the healthcare burden.

This study's univariate analysis confirmed that antibiotic use prior to catheter insertion is significantly associated with the risk of CAUTI in stroke patients. Presently, the antibiotic crisis has become a major challenge to global public health. For stroke patients concurrently affected by CAUTI, the therapeutic challenges posed by the antibiotic crisis are particularly pronounced: on the one hand, CAUTI caused by multidrug-resistant organisms exhibits markedly diminished responsiveness to conventional antimicrobial regimens, necessitating the use of higher-grade, broad-spectrum antibiotics. This not only increases treatment failure rates but also readily induces secondary infections, further complicating the patient's clinical condition; On the other hand, stroke patients are predominantly elderly with comorbidities such as diabetes and hypertension, presenting compromised immune function. CAUTI caused by resistant bacteria more readily progresses to sepsis, markedly increasing the risk of poor patient outcomes^{6,11} while prolonging antimicrobial treatment cycles, thereby creating a vicious cycle. Moreover, against the backdrop of the antibiotic crisis, the treatment costs for CAUTI have substantially escalated. The use of high-level broad-spectrum antibiotics, repeated urine cultures and antimicrobial susceptibility testing, alongside the management of infection-related complications, collectively intensify the consumption of healthcare resources. This aligns with the findings observed in this study, which revealed significantly elevated hospitalisation costs for CAUTI patients. It further corroborates, from a therapeutic perspective, the dual negative impact of CAUTI on both healthcare quality and the burden on healthcare systems. Multivariate logistic regression analysis revealed that the duration of urinary catheterisation was the strongest independent risk factor for CAUTI in stroke patients (OR=1.08, 95% CI 1.07–1.09, $P<0.001$). The presence of a urinary catheter may damage the bladder mucosa, impairing its defence mechanisms and providing a pathway for bacterial entry.^{19,20} As an invasive procedure, indwelling catheterisation disrupts the natural barrier of the urethral mucosa. Prolonged catheterisation significantly increases the risk of bacterial colonisation—opportunistic pathogens around the urethra may ascend through the catheter's outer wall or lumen into the bladder, forming biofilms on the catheter surface and subsequently triggering infection^{21,22}. Furthermore, studies indicate that prolonged catheter use exacerbates damage to the glycosaminoglycan layer of urinary epithelial cell walls, rendering them more susceptible to bacterial invasion and infection, thereby increasing the likelihood of infection.^{23,24} A prospective randomised trial by Kordis et al confirmed²⁵ that reducing catheterisation duration from 4 days to 24 hours in patients undergoing pelvic organ prolapse surgery not only failed to increase urinary retention risk but also significantly lowered infection rates and shortened hospital stays. For stroke patients, who frequently present with neurogenic bladder and mobility restrictions, the need for indwelling catheterisation is more urgent and duration tends to be prolonged. Consequently, catheterisation duration emerges as a core predictor of CAUTI risk in this population.²⁶ Therefore, clinical practice should emphasise daily reassessment of catheter necessity in stroke patients. Where clinical condition permits, early catheter removal or transition to intermittent catheterisation should be pursued to mitigate infection risk.

This study confirms that age is an independent risk factor for CAUTI (OR=1.01, 95% CI 1.00–1.02, $P=0.029$), a finding consistent with multiple clinical studies:^{7,27,28} Dunn et al's retrospective cohort study found that elderly patients aged 80 years and above undergoing intravesical botulinum toxin A injections exhibited a significantly higher incidence of urinary tract infections compared to younger cohorts;²⁷ Ajjawi et al's analysis of total hip replacement patients similarly indicated advanced age as a significant perioperative risk factor for urinary tract infections.²⁸ For stroke patients, the physiological decline associated with advanced age compounds post-stroke neurological dysfunction, facilitating bacterial colonisation and proliferation within the urethra and bladder. Furthermore, elderly stroke patients frequently present with comorbidities such as diabetes mellitus and hypertension, which further compromise the body's anti-infective capacity and indirectly elevate the risk of CAUTI.^{29,30} This finding suggests that clinical practice should establish more stringent catheter care protocols for elderly stroke patients, emphasising foundational preventive measures such as meticulous urethral orifice cleansing and regular drainage bag replacement, alongside close monitoring of infection-related indicators.

Although the random forest algorithm identified the duration from hospital admission to catheter use as a potential predictor, multivariate logistic regression analysis revealed it did not achieve statistical significance (OR=1.00, 95% CI

0.98–1.01, $P=0.751$). This indicates that the risk of CAUTI in stroke patients is more dependent on the absolute duration of catheterisation rather than the timing of insertion. This may be explained by the fact that catheterisation decisions in stroke patients are often driven by urgent clinical needs, such as impaired consciousness preventing autonomous voiding. The duration from hospital admission to catheter primarily correlates with disease severity, whereas the core drivers of infection remain bacterial colonisation and biofilm formation during catheter retention. This finding informs clinical catheterisation decisions: when urgent catheterisation is required in stroke patients, excessive focus on whether delayed placement reduces infection risk is unnecessary. Instead, emphasis should be placed on post-placement care quality and assessment of appropriate timing for catheter removal.

This study employs a nomogram prediction model constructed based on factors such as age and duration of catheter use. Compared to traditional empirical assessments, this model offers significant advantages: firstly, it is built upon a large sample dataset and validated through training and validation sets, yielding higher reliability; Secondly, the use of the random forest algorithm for variable selection effectively mitigates multicollinearity issues in high-dimensional data, ensuring the accuracy of risk factor identification. Thirdly, the nomogram format is intuitive and straightforward. Clinicians can rapidly calculate a patient's probability of CAUTI occurrence by simply summing the scores corresponding to age and catheterisation duration, eliminating the need for complex statistical calculations and facilitating bedside application. Against the backdrop of increasing multidrug-resistant pathogens in clinical settings, this model assists clinicians in precisely identifying high-risk individuals. This enables prioritisation of targeted prevention measures, such as enhanced aseptic techniques and antimicrobial-coated catheters, thereby avoiding unnecessary interventions for low-risk patients while effectively reducing infection risks for high-risk individuals. This approach aligns with the core principles of precision medicine.

This study has several limitations: firstly, it employed a single-centre retrospective design with data sourced from Guangdong Provincial Second Hospital of Traditional Chinese Medicine. Clinical characteristics and diagnostic protocols may differ from other medical centres, potentially limiting the model's external generalisability. Secondly, retrospective studies inherently carry limitations, with potential under-recording or bias in certain clinical data, and the influence of unmeasured confounding factors cannot be entirely excluded. Thirdly, the model did not incorporate potential predictive indicators such as biomarkers, suggesting scope for enhancing predictive efficacy. Future research may validate the model's external validity through multicentre studies with expanded sample sources, optimising the model according to regional clinical characteristics. Supplementing the collection of potential predictors—including biomarkers and nursing quality—and integrating machine learning algorithms could further refine the model's predictive accuracy.

In summary, CAUTI in stroke patients significantly prolongs hospital stays, increases healthcare costs, and severely compromises treatment quality. Age and duration of catheterisation represent core independent risk factors for CAUTI. The nomogram prediction model developed based on these two indicators demonstrates excellent discriminatory ability, calibration, and clinical net benefit. It assists clinicians in early identification of high-risk patients and implementation of targeted prevention strategies. This study provides a scientific tool and clinical basis for the precise prevention and control of CAUTI in stroke patients. It contributes to reducing infection rates, optimising healthcare resource allocation, and ultimately improving stroke patients' prognosis and quality of life.

Code Availability

The statistical software “R- version 4.3.1” was used. The code can be found in [Supplementary File 1](#).

Abbreviations

CAUTI, Catheter-associated urinary tract infection; AUC, Area under the curve; ROC, Receiver Operating Characteristic curve; DCA, Decision curve analysis; DaysCAU, days catheter in use; TTCAU, time to catheterisation after admission; MV, mechanical ventilation; PICC, peripherally inserted central catheter; HF, heart failure; CKI, Chronic kidney insufficiency; ED, electrolyte disturbance; VIF, variance inflation factors.

Data Sharing Statement

Data are available upon reasonable request to the corresponding author.

Ethics and Consent to Participate Section

This retrospective study employed anonymised patient data from 7486 participants, achieved by removing personally identifiable information during statistical analysis. Conducted in accordance with the Declaration of Helsinki, the research was approved by the Ethics Committee of Guangdong Provincial Second Hospital of Traditional Chinese Medicine, which granted exemption from informed consent requirements as mandated by relevant legislation (Ethics Committee Approval Number: Z202404-002-01).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Guangzhou Municipal Health and Health Science and Technology Project (20251A010038).

Disclosure

All the authors declare that they have no conflicts of interest in this work.

References

1. GBD 2023 Disease and Injury and Risk Factor Collaborators. Burden of 375 diseases and injuries, risk-attributable burden of 88 risk factors, and healthy life expectancy in 204 countries and territories, including 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. *Lancet*. 2025;406(10513):1873–1922. doi:10.1016/S0140-6736(25)01637-X
2. GBD 2021 Stroke Risk Factor Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol*. 2024;23(10):973–1003. doi:10.1016/S1474-4422(24)00369-7
3. Feigin VL, Brainin M, Norrving B, et al. World stroke organization: global stroke fact sheet 2025. *Int J Stroke*. 2025;20(2):132–144. doi:10.1177/17474930241308142
4. Umemura T, Ohta H, Yokota A, Yarimizu S, Nishizawa S. Urinary retention associated with stroke. *J UOEH*. 2016;38(4):263–269. doi:10.7888/juoeh.38.263
5. Jitpratoom P, Boonyasiri A. Determinants of urinary tract infection in hospitalized patients with acute ischemic stroke. *BMC Neurol*. 2023;23(1):251. doi:10.1186/s12883-023-03296-2
6. Bogason E, Morrison K, Zalatimo O, et al. Urinary tract infections in hospitalized ischemic stroke patients: source and impact on outcome. *Cureus*. 2017;9(2):e1014. doi:10.7759/cureus.1014
7. Magill SS, Edwards JR, Bamberg W, et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med*. 2014;370(13):1198–1208. doi:10.1056/NEJMoa1306801
8. Reynolds SS, Sova CD, Lewis SS, et al. Sustained reduction in catheter-associated urinary tract infections using multi-faceted strategies led by champions: a quality improvement initiative. *Infect Control Hosp Epidemiol*. 2022;43(7):925–929. doi:10.1017/ice.2021.135
9. Fernandez-Lozano C, Hervella P, Mato-Abad V, et al. Random forest-based prediction of stroke outcome. *Sci Rep*. 2021;11(1):10071. doi:10.1038/s41598-021-89434-7
10. Wang J, Gong X, Chen H, et al. Causative classification of ischemic stroke by the machine learning algorithm random forests. *Front Aging Neurosci*. 2022;14:788637. doi:10.3389/fnagi.2022.788637
11. Zhang L, Li X, Cai D, et al. Clinical prediction of secondary bloodstream infections in patients with cerebral infarction: a nomogram-driven risk assessment model based on LASSO regression. *Infect Drug Resist*. 2025;18:3677–3687. doi:10.2147/IDR.S529528
12. Li X, Liang Z, Zhang A, et al. Analysis of prognostic risk factors and risk management measures for patients with ischemic stroke and bloodstream infection based on machine learning. *Front Cell Infect Microbiol*. 2025;15:1715309. doi:10.3389/fcimb.2025.1715309
13. Chinese Society of Neurology, Chinese Stroke Society. Diagnostic criteria of cerebrovascular diseases in China (version 2019). *Chin J Neurol*. 2019;52(9):710–715. doi:10.3760/cma.j.issn.1006-7876.2019.09.003
14. National Health Commission. Standard for identifying cases of healthcare associated infections: general principles. *Chin J Infect Control*. 2025;24(8):1175–1177. doi:10.12138/j.issn.1671-9638.20257385
15. Hutton DW, Krein SL, Saint S, et al. Economic evaluation of a catheter-associated urinary tract infection prevention program in nursing homes. *J Am Geriatr Soc*. 2018;66(4):742–747. doi:10.1111/jgs.15316
16. Chen W, Li D. Comorbidity and outcomes among hospitalized patients with stroke: a nationwide inpatient analysis. *Front Neurol*. 2023;14:1217404. doi:10.3389/fneur.2023.1217404
17. Zhang J, Song S, Zhao Y, Ma G, Jin Y, Zheng ZJ. Economic burden of comorbid chronic conditions among survivors of stroke in China: 10-year longitudinal study. *BMC Health Serv Res*. 2021;21(1):978. doi:10.1186/s12913-021-07010-1
18. Shaikh N, Mohammed A, Seddiq M, Kidwai S, Shahzad D, Mahmoud MM. The effect of the charlson comorbidity index on in-hospital complications, hospital length of stay, mortality, and readmissions among patients hospitalized for acute stroke. *Cureus*. 2024;16(5):e60112. doi:10.7759/cureus.60112

19. Bammigatti C, Doradla S, Belgode HN, Kumar H, Swaminathan RP. Healthcare associated infections in a resource limited setting. *J Clin Diagn Res.* 2017;11(1):OC01–OC04. doi:10.7860/jcdr/2017/23076.9150
20. Firoozeh N, Agah E, Bauer ZA, Olusanya A, Seifi A. Catheter-associated urinary tract infection in neurological intensive care units: a narrative review. *Neurohospitalist.* 2022;12(3):484–497. doi:10.1177/19418744221075888
21. Nasrollahian S, Halaji M, Hosseini A, et al. Genetic diversity, carbapenem resistance genes, and biofilm formation in UPEC isolated from patients with catheter-associated urinary tract infection in North of Iran. *Int J Clin Pract.* 2022;2022:9520362. doi:10.1155/2022/9520362
22. Letica-Kriegel AS, Salmasian H, Vawdrey DK, et al. Identifying the risk factors for catheter-associated urinary tract infections: a large cross-sectional study of six hospitals. *BMJ Open.* 2019;9(2):e022137. doi:10.1136/bmjopen-2018-022137
23. Kordis T, Kofol A, Blaganje M. Shortening indwelling catheterization after vaginal surgery for pelvic organ prolapse: results from a prospective randomized trial. *J Clin Med.* 2025;14(23):8295. doi:10.3390/jcm14238295
24. Retelski J, Richardson T, Mahabaleshwarkar R, Gohs FX, Spencer MD. Retrospective analysis of catheter-acquired urinary tract infection: relationship to stroke diagnosis. *Clin Nurse Spec.* 2017;31(4):E11–E16. doi:10.1097/NUR.0000000000000307
25. Manack A, Motsko SP, Haag-Molkenteller C, et al. Epidemiology and healthcare utilization of neurogenic bladder patients in a US claims database. *Neurol Urodyn.* 2011;30(3):395–401. doi:10.1002/nau.21003
26. Vasudeva P, Madersbacher H. Factors implicated in pathogenesis of urinary tract infections in neurogenic bladders: some revered, few forgotten, others ignored. *Neurol Urodyn.* 2014;33(1):95–100. doi:10.1002/nau.22378
27. Dunn FE, Clarke B, Newark A, et al. Intradetrusor onabotulinumtoxin safety and efficacy in patients 80 years and older: a retrospective cohort study. *Int Urogynecol J.* 2025. doi:10.1007/s00192-025-06484-5
28. Ajjawi I, Katsnelson B, Grauer JN. Correlation of sex with adverse outcomes following total hip arthroplasty: a matched cohort analysis. *J Am Acad Orthop Surg Glob Res Rev.* 2025;9(12):e2500104. doi:10.5435/JAAOSGlobal-D-25-00104
29. Abulhasan YB, Alabdulraheem N, Schiller I, et al. Health care-associated infections after subarachnoid hemorrhage. *World Neurosurg.* 2018;115:e393–e403. doi:10.1016/j.wneu.2018.04.061
30. Hagerty T, Kertesz L, Schmidt JM, et al. Risk factors for catheter-associated urinary tract infections in critically ill patients with subarachnoid hemorrhage. *J Neurosci Nurs.* 2015;47(1):51–54. doi:10.1097/jnn.0000000000000111

International Journal of General Medicine

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>

Dovepress
Taylor & Francis Group