

Prediction Models for Postoperative Deep Vein Thrombosis in Elderly Hip Fracture Patients: A Systematic Review

Shuqing Yang, Hongxia Cheng*

Department of Orthopedics, Taihe Hospital, Hubei University of Medicine, Shiyan 442000, Hubei, China

*Author to whom correspondence should be addressed.

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: *Objective:* To systematically evaluate prediction models for postoperative deep vein thrombosis (DVT) in elderly hip fracture patients and assess their methodological quality and predictive performance. *Methods:* Following PRISMA guidelines, we searched eight databases (PubMed, Embase, Cochrane Library, Web of Science, CINAHL, CNKI, Wanfang, VIP) from inception to May 2025. Studies developing or validating DVT prediction models in elderly hip fracture patients were included. Two reviewers independently screened studies, extracted data, and assessed risk of bias and applicability using the PROBAST tool. *Results:* Eleven studies were included, all conducted in China between 2021 and 2025. Sample sizes ranged from 101 to 504 patients (total n = 3,286). Models incorporated 3 to 9 predictors, with D-dimer, age, and time from injury to surgery being most common. All 11 studies (100%) were rated as high risk of bias, primarily due to small sample sizes, lack of validation, and inadequate missing data handling. Applicability concerns were low in 8 studies (72.7%). AUC values ranged from 0.648 to 0.967, with 10 studies (90.9%) reporting AUC > 0.7. Meta-analysis identified time from injury to surgery (OR = 4.63, 95% CI: 2.58–6.68), age (OR = 1.99), D-dimer (OR = 1.51), and Caprini score (OR = 1.75) as significant predictors. *Conclusion:* Current DVT prediction models for elderly hip fracture patients demonstrate acceptable discrimination but are limited by high risk of bias and lack of external validation. Prospective, multicenter studies with rigorous validation are needed to develop clinically applicable models.

Keywords: Hip fracture; Deep vein thrombosis; Prediction model; Risk assessment; Systematic review

Online publication: Dec 31, 2025

1. Introduction

With global population aging, hip fractures in the elderly have become a major public health challenge, associated with significant morbidity and mortality^[1]. Deep vein thrombosis (DVT) is one of the most serious postoperative complications in this population. Due to surgical trauma, prolonged immobilization, and hypercoagulable states, DVT incidence ranges from 8% to 34.9% despite standard thromboprophylaxis^[2]. Untreated DVT can progress to

fatal pulmonary embolism and chronic post-thrombotic syndrome, significantly compromising patient outcomes.

Early risk stratification and targeted prophylaxis are critical for optimizing outcomes. However, widely-used generic prediction models such as the Caprini, Padua, and Wells scores were not specifically designed for elderly hip fracture patients and may lack adequate sensitivity or specificity. Consequently, several prediction models tailored specifically for this population have been recently developed, but they demonstrate considerable heterogeneity in methodology and performance.

Systematic evaluation of these models using standardized methodological assessment tools remains limited. This systematic review aims to comprehensively evaluate DVT prediction models for elderly hip fracture patients using the Prediction model Risk of Bias Assessment Tool (PROBAST), providing evidence-based guidance for clinical decision-making and DVT prevention in this high-risk population.

2. Methods

2.1. Study design and registration

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement.

2.2. Eligibility criteria

Studies were included if they developed or validated prediction models for postoperative DVT risk in elderly patients aged ≥ 60 years with hip fractures who underwent hip surgery. Eligible study designs included case-control, cohort, or retrospective studies that reported model performance metrics such as AUC, C-statistic, or calibration measures. Studies were excluded if they only analyzed risk factors without developing a prediction model, included non-hip fractures, or were duplicate publications.

2.3. Information sources and search strategy

A comprehensive literature search was conducted across eight electronic databases (PubMed, Embase, Web of Science, Cochrane Library, CINAHL, CNKI, Wanfang, and VIP) from inception to May 31, 2025. Search terms included “hip fracture”, “deep vein thrombosis”, “DVT”, “prediction model”, “risk model”, and “nomogram”, combined using Boolean operators. No language restrictions were applied. Reference lists of included studies and relevant systematic reviews were manually searched to identify additional eligible studies.

2.4. Study selection and data extraction

Two reviewers independently screened titles and abstracts, followed by full-text assessment of potentially relevant studies. Disagreements were resolved through discussion or consultation with a third reviewer. Data extraction was performed using a standardized form, capturing study characteristics (author, year, country, design, sample size, DVT events), model characteristics (modeling method, number and type of predictors, presentation format), model performance (AUC, calibration), outcome definition (DVT diagnostic criteria), and missing data handling methods.

2.5. Risk of bias and applicability assessment

Two reviewers independently assessed methodological quality using the Prediction model Risk of Bias Assessment Tool (PROBAST), which evaluates risk of bias and applicability across four domains: participants, predictors,

outcome, and statistical analysis. Risk of bias was rated as high if at least one domain was rated as high risk, low if all four domains were rated as low risk, and unclear otherwise. Applicability was assessed across three domains (participants, predictors, outcome) and rated as high concern if at least one domain raised high concerns, low concern if all domains raised low concerns, and unclear otherwise. Disagreements were resolved through discussion or third-party adjudication.

2.6. Data synthesis and statistical analysis

For predictors reported in at least three studies, meta-analysis was conducted using R software (version 4.3.0) with odds ratios and 95% confidence intervals as effect measures. Heterogeneity was assessed using the I^2 statistic and Cochran's Q test. Fixed-effects models were applied when $I^2 < 50\%$ or $p > 0.10$, while random-effects models were used when $I^2 \geq 50\%$ or $p \leq 0.10$. Publication bias was assessed using funnel plots and Egger's test when at least 10 studies were available. Statistical significance was set at $p < 0.05$. Model performance metrics were summarized descriptively due to expected methodological heterogeneity, and meta-analysis of performance metrics was not conducted due to insufficient comparable data.

3. Results

3.1. Study selection

The database search identified 856 potentially relevant articles. After removing 312 duplicates, 544 records underwent title and abstract screening, followed by full-text review of 28 articles. Ultimately, 11 studies met the inclusion criteria and were included in this systematic review ^[3-13].

3.2. Study characteristics

All 11 included studies were published between 2021 and 2025 and conducted in China. Seven studies employed retrospective cohort designs, and four used case-control designs. Sample sizes ranged from 101 to 504 patients, with a total of 3,286 participants. All studies focused on elderly patients aged ≥ 60 years who underwent hip surgery for hip fractures. Modeling methods included logistic regression ($n = 9$), LASSO regression ($n = 1$), and mixed methods ($n = 1$). The number of predictors in final models ranged from 3 to 9 variables. DVT was diagnosed by lower limb venous Doppler ultrasonography in all studies. Only one study (9.1%) reported internal validation, and none performed external validation.

3.3. Model performance

The area under the receiver operating characteristic curve (AUC) for development cohorts ranged from 0.648 to 0.967, with 10 studies (90.9%) reporting AUC values > 0.7 , indicating acceptable to good discrimination. Nine studies (81.8%) assessed calibration using the Hosmer-Lemeshow goodness-of-fit test, with all demonstrating adequate calibration ($p > 0.05$).

3.4. Risk of bias and applicability

According to PROBAST assessment, all 11 studies (100%) were rated as high risk of bias. The most common source of bias was in the Analysis domain (100% high risk), primarily due to small sample sizes relative to the number of candidate predictors, failure to report missing data handling, and lack of validation. Overall applicability was judged as low concern in 8 studies (72.7%) and unclear concern in 3 studies (27.3%), suggesting

that the predictors and outcomes assessed are clinically relevant to the target population.

3.5. Meta-analysis of common predictors

Eight predictor variables reported in at least three studies were eligible for meta-analysis. Random-effects or fixed-effects models were applied based on heterogeneity assessment. The pooled effect estimates for common predictors are presented in **Table 1**. Time from injury to surgery emerged as the strongest predictor (OR = 4.63, 95% CI: 2.58–6.68, $p < 0.001$) with negligible heterogeneity ($I^2 = 0.0\%$). Other significant predictors included age (OR = 1.99, 95% CI: 1.06–2.92, $p < 0.001$), Caprini score (OR = 1.75, 95% CI: 0.74–2.75, $p < 0.001$), D-dimer (OR = 1.51, 95% CI: 1.09–1.94, $p < 0.001$), duration of bed rest (OR = 1.66, 95% CI: 0.89–2.42, $p < 0.001$), time from injury to admission (OR = 1.16, 95% CI: 0.93–1.39, $p < 0.001$), and systemic immune-inflammation index (SII) (OR = 1.02, 95% CI: 0.99–1.04, $p < 0.001$). Diabetes mellitus showed a non-significant trend (OR = 5.12, 95% CI: -0.58–10.83, $p = 0.078$). As shown in **Table 1**, substantial heterogeneity ($I^2 \geq 50\%$) was observed for Caprini score, SII, time from injury to admission, diabetes mellitus, and duration of bed rest, necessitating random-effects models. Low heterogeneity was found for D-dimer, age, and time from injury to surgery.

Table 1. Meta-analysis of common predictors for postoperative DVT in elderly hip fracture patients

| Predictor | OR | 95% CI | Z value | p value | I ² (%) | Heterogeneity p value |
|-------------------------------|------|-------------|---------|---------|--------------------|-----------------------|
| Caprini score | 1.75 | 0.74–2.75 | 3.40 | < 0.001 | 78.3 | < 0.001 |
| D-dimer | 1.51 | 1.09–1.94 | 6.96 | < 0.001 | 49.2 | 0.08 |
| SII | 1.02 | 0.99–1.04 | 72.76 | < 0.001 | 94.0 | < 0.001 |
| Age | 1.99 | 1.06–2.92 | 4.20 | < 0.001 | 0.0 | 0.47 |
| Time from injury to admission | 1.16 | 0.93–1.39 | 9.90 | < 0.001 | 59.2 | 0.06 |
| Time from injury to surgery | 4.63 | 2.58–6.68 | 4.43 | < 0.001 | 0.0 | 0.84 |
| Diabetes mellitus | 5.12 | -0.58–10.83 | 1.76 | 0.078 | 54.0 | 0.07 |
| Duration of bed rest | 1.66 | 0.89–2.42 | 4.26 | < 0.001 | 84.6 | 0.01 |

Abbreviations: OR = odds ratio; CI = confidence interval; SII = systemic immune-inflammation index; DVT = deep vein thrombosis.

Notes: Random-effects models were applied when $I^2 \geq 50\%$ or heterogeneity $p \leq 0.10$; fixed-effects models were used when $I^2 < 50\%$ and heterogeneity $p > 0.10$.

4. Discussion

4.1. DVT prediction models for elderly hip fracture patients remain in early developmental stages

With the global aging population, developing reliable DVT risk prediction models for elderly hip fracture patients is increasingly critical. However, this field remains in its early stages, as all 11 included studies were published within the past four years (2021–2025), conducted exclusively in single-center settings in China with sample sizes of 101–504 patients. This geographic homogeneity limits generalizability to other populations and healthcare settings.

Methodological limitations were evident across studies. Only one study employed advanced techniques such as LASSO regression, while others used traditional logistic regression. Critically, only one study reported internal validation, and none conducted external validation. Without rigorous validation, reported performance metrics may

be overly optimistic and fail to reflect real-world utility. Future studies should adhere to the TRIPOD statement and leverage multicenter collaborations to develop more robust models.

4.2. Key predictors demonstrate clinical relevance

Time from injury to surgery emerged as the strongest predictor ($OR = 4.63$, 95% CI: 2.58–6.68, $p < 0.001$) with negligible heterogeneity ($I^2 = 0.0\%$). Prolonged preoperative delays contribute to DVT risk through venous stasis, inflammatory activation, and muscle atrophy, underscoring the importance of early surgical intervention. Age ($OR = 1.99$) and D-dimer levels ($OR = 1.51$) demonstrated consistent associations with low heterogeneity, reflecting their biological relevance. The Caprini score ($OR = 1.75$) showed clinical utility but with high heterogeneity ($I^2 = 78.3\%$), suggesting variability in scoring protocols.

The systemic immune-inflammation index (SII) demonstrated statistical significance ($OR = 1.02$) but very high heterogeneity ($I^2 = 94.0\%$), limiting its clinical applicability. Diabetes mellitus showed a strong effect size ($OR = 5.12$) but did not reach statistical significance ($p = 0.078$), possibly due to small sample sizes and failure to account for disease severity. Larger studies with stratified analyses are needed to clarify its independent contribution.

4.3. Methodological quality limits clinical translation

PROBAST assessment revealed that all 11 studies had high risk of bias, predominantly in the Analysis domain. Key concerns included inadequate sample sizes (most studies had < 10 events per variable, below the recommended 20), no reporting of missing data handling, and lack of validation. Most studies relied on univariate screening before multivariable modeling, which fails to account for confounding. These deficiencies substantially limit clinical applicability despite promising discrimination performance (AUC > 0.7 in 90.9% of studies). Applicability concerns were low (72.7%), suggesting clinically relevant predictors and outcomes.

4.4. Implications and future directions

While most models demonstrated acceptable discrimination, their high risk of bias precludes immediate clinical implementation. Clinicians should rely on established risk assessment tools while prioritizing modifiable risk factors such as minimizing surgical delay. Future research should prioritize prospective, multicenter studies with adequate sample sizes, standardized predictor definitions, and rigorous validation including calibration and decision curve analysis. Machine learning algorithms should be explored to capture complex predictor interactions. Development of user-friendly implementation tools and real-world impact studies would facilitate clinical translation.

4.5. Limitations

This systematic review is limited by the small number of single-center studies from China, restricting geographic diversity. Retrospective designs preclude causal inference. Substantial heterogeneity in predictor definitions limited comprehensive meta-analyses. Publication bias could not be assessed due to insufficient studies per predictor.

5. Conclusion

Based on the current evidence, this review concludes that existing DVT prediction models for elderly hip fracture patients, while demonstrating acceptable discrimination, are significantly limited by high risk of bias and insufficient external validation. Therefore, the development of clinically robust and widely applicable models

requires future large-scale, multicenter prospective studies employing rigorous methodology and validation.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Downey C, Wang M, Quinlan J, 2019, Changing Trends in the Mortality Rate at 1-Year Post Hip Fracture: A Systematic Review. *World J Orthop*, 10(3): 166–175.
- [2] Wang T, Liu C, Fang S, et al., 2022, Risk Factors for Preoperative Deep Venous Thrombosis in Hip Fracture Patients: A Meta-Analysis. *J Orthop Surg Res*, 17(1): 197.
- [3] Yu Y, 2024, Construction and Evaluation of a Risk Prediction Model for Postoperative Deep Vein Thrombosis in Elderly Patients with Hip Fractures. *Modern Medicine and Health Research Electronic Journal*, 8(16): 19–22.
- [4] Zhao X, Liu H, Ren Y, 2024, Construction of a Risk Prediction Model for Deep Vein Thrombosis After Hip Fracture Surgery in Elderly Patients Based on Caprini Thrombosis Risk Score and D-Dimer. *Anhui Medical and Pharmaceutical Journal*, 28(12): 2452–2457.
- [5] Wang Y, Du X, Wang L, et al., 2024, Construction of a Prediction Model for Lower Extremity Deep Vein Thrombosis in Patients with Hip Fractures Based on Nomogram. *Journal of Guangdong Pharmaceutical University*, 40(02): 144–149.
- [6] Chen X, Leng B, 2021, Construction of a Prediction Model for Venous Thromboembolism After Hip Replacement in Elderly Patients with Femoral Neck Fracture. *Journal of Beihua University (Natural Science)*, 22(05): 653–657.
- [7] Li Y, Tian X, Wang C, et al., 2021, Risk Factor Analysis and Model Prediction of Deep Vein Thrombosis After Treatment in Elderly Patients with Hip Fractures. *Journal of Vascular and Endovascular Surgery*, 7(08): 909–913.
- [8] Xue G, 2024, Risk Factors and Construction of Risk Prediction Model for Lower Extremity Deep Vein Thrombosis in Elderly Patients with Peri-Hip Fractures, thesis, Jinzhou Medical University.
- [9] Chen X, Bo R, Yuan Y, 2025, Nomogram Prediction Model for Deep Vein Thrombosis After Total Hip Arthroplasty. *Journal of Lanzhou University (Medical Sciences)*, 51(07): 46–53.
- [10] Hu G, Li G, Yin M, et al., 2021, Construction of a Nomogram Model for Predicting the Risk of Deep Vein Thrombosis in Elderly Patients with Hip Fractures. *Chinese Journal of Traditional Chinese Medicine Traumatology & Orthopedics*, 29(06): 26–31.
- [11] Bo R, Chen X, Zheng X, et al., 2024, A Nomogram Model to Predict Deep Vein Thrombosis Risk After Surgery in Patients with Hip Fractures. *Indian Journal of Orthopaedics*, 58(2): 151–161.
- [12] Peng J, Wang H, Zhang L, et al., 2021, Construction and Predictive Performance of a Risk Prediction Model for Venous Thromboembolism After Hip Fractures in the Elderly. *Journal of Central South University (Medical Science)*, 46(02): 142–148.
- [13] Chen X, Fan Y, Tu H, et al., 2024, A Nomogram Model Based on the Systemic Immune-Inflammation Index to Predict the Risk of Venous Thromboembolism in Elderly Patients After Hip Fracture: A Retrospective Cohort Study. *Heliyon*, 10(6): e28389.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.