Supplementary Methods and Reproducibility Appendix

Evidence-Calibrated Simulation of Thromboprophylaxis After Femoral Shaft Fracture

Yuzhan Zhang

1 S1. Targeted evidence mapping

Searches were last updated on 22 August 2026. The evidence map was designed to support model definition, falsification, calibration, uncertainty analysis, and prospective-study design rather than to estimate one pooled treatment effect. Searches centered on PubMed-indexed records, publisher pages, major society/guideline sources, and backward/forward checking around direct femoral shaft fracture (FSF) studies.

Core query families included:

1. `("femoral shaft fracture" AND ("deep vein thrombosis" OR "pulmonary embolism" OR VTE OR thromboprophylaxis))`;
2. `(trauma OR "lower limb fracture" OR orthopedic) AND (LMWH OR heparin OR aspirin OR thromboprophylaxis) AND (VTE OR bleeding)`;
3. `(fracture OR trauma) AND (immobility OR mobilization OR ambulation OR weight-bearing) AND (DVT OR VTE)`;
4. `(fracture OR trauma) AND (TEG OR ROTEM OR "thrombin generation" OR hypercoagulability)`;
5. `("value of information" OR EVPI OR EVPPI OR EVSI) AND (healthcare OR decision model)`;
6. `(model validation OR model transparency) AND (health decision model OR simulation)`; and
7. `("fat embolism syndrome" AND long-bone fracture AND pulmonary)`.

Direct FSF evidence was prioritized over adjacent orthopedic-trauma evidence, which in turn was prioritized over external mechanistic evidence. Methodological sources informed VOI methods, model validation terminology, bleeding definitions, and health-state valuation. The final evidence matrix contains 47 records. A single publication can contribute multiple records when it informs distinct estimands; therefore record count is not equivalent to unique publication count.

2 S2. State variables and observability

For patient i on postoperative day t:

- `B_i`: stable baseline thrombotic susceptibility;
- `H_i(t)`: trauma-associated hypercoagulability;
- `M_i(t)`: residual venous-stasis burden;
- `U_i(t)`: assisted-upright mobility exposure conditional on clinician clearance;
- `A_i(t)`: effective prophylactic anticoagulant exposure;

- `D_i(t)`: latent DVT state;
- `S_i`: symptom-capable phenotype if a DVT forms;
- `O_i(t)`: observed/diagnosed DVT state;
- `P_i(t)`: pulmonary embolism (PE) state; and
- `R_i(t)`: major bleeding state.

Disease and observation states are deliberately separated. A prophylaxis policy can react only to clinically available information; it cannot observe an undiagnosed latent DVT. This restriction is enforced in code and tested automatically.

2.1 S2.1 Hypercoagulability

`H_i(t)` uses patient-specific peak amplitude, peak day, and shape. Serial TEG after hip fracture is used only as a weak external temporal anchor [1]. Direct FSF calibration is allowed to move the effective risk trajectory away from the hip-fracture TEG curve.

2.2 S2.2 Mobility and residual stasis

Natural residual stasis follows

$$M_{0i}(t) = f_i + (1 - f_i)e^{-t/\tau_i},$$

where `f_i` is an individual residual floor and `tau_i` a recovery time constant. If clinician-cleared assisted activity is admissible,

$$M_i(t) = \max\{f_i,\ M_{0i}(t)e^{-\rho U_i(t)}\}.$$

`U_i(t)=0` whenever clinician clearance is false. Weight-bearing permission and assisted-upright activity are distinct variables. The model does not assign a universal postoperative day for either activity.

2.3 S2.3 Latent DVT hazard

The daily latent-DVT hazard is

$$\begin{aligned}\lambda_{D,i}(t) = \lambda_0 q(t) \exp\{&\beta_B B_i + \beta_H H_i(t) + \beta_M M_i(t) \\ &+ \beta_{HM} H_i(t) M_i(t) - \beta_A A_i(t) \\ &- \beta_{HA} H_i(t) A_i(t) - \beta_{MA} M_i(t) A_i(t) + \mathbf{z}_i^T \gamma\}.\end{aligned}$$

The baseline multiplier `q(t)` includes an early peak and residual late component. In the reference implementation, `beta_MA=0` because no transportable FSF-specific mobility-by-anticoagulation interaction estimate was identified. This interaction is retained as a future estimand.

3 S3. Observation model and early PE

Each newly formed DVT is assigned a Bernoulli indicator for a model-defined symptom-capable phenotype, `S_i`. Symptom-triggered clinical diagnosis is possible only through this pathway, whereas scheduled duplex surveillance can detect either symptom-capable or clinically silent latent DVT.

This structure was introduced after a uniform daily clinical-ascertainment model failed to reproduce the combination of high surveillance-detected DVT incidence and much lower, later clinically ascertained DVT incidence.

Source-calibration scenarios may initialize an occult preoperative DVT reservoir. Such DVT can contribute to early PE through a decaying early embolization multiplier. This state is disabled when a source explicitly excludes preoperative DVT and is disabled in the policy-duration experiment. Known DVT is not treated as a prophylaxis problem.

The early-reservoir mechanism is a reduced-form explanation rather than a unique biological claim. The discordance between early PE and later clinically recognized DVT can also reflect surveillance, DVT location, embolization timing, or endpoint misclassification. The model retains broad uncertainty around these mechanisms.

4 S4. Aggregate calibration targets

The principal quantitative targets were:

Source	Target population / protocol	Model target
Yang 2022 [2]	Isolated FSF; routine preoperative duplex	Preoperative DVT 26.4%
Ren 2021 [3]	Surgically treated FSF; preoperative DVT excluded; scheduled duplex	Postoperative DVT 15.6%; 85.4% by day 14; mean day 8.1
Tischler 2024 [4]	Isolated closed FSF; NSQIP clinical capture	DVT 18/1346; PE 15/1346; mean DVT day 10.7; mean PE day 6.6; PE by day 6 = 11/15
You 2021 [1]	Hip fracture; serial TEG	Weak external timing anchor for hypercoagulability
Geerts 1996 [5]	Major trauma randomized trial	Weak external major bleeding information

Direct FSF endpoints received greater calibration weight than transported external targets.

4.1 S4.1 Calibration algorithm

A two-stage common-random-number (CRN) procedure was used. Stage 1 sampled broad research priors. Stage 2 generated local perturbations around particles with low calibration discrepancy. A tempered exponential-discrepancy kernel then defined calibration weights. The production checkpoint used 240 broad particles, 360 local particles, and 2,000 simulated patients per source per particle. The retained effective ensemble size was 48. Independent-seed simulation checks resimulated 20 weighted particles with 3,500 patients per source.

The ensemble is not labeled an exact Bayesian posterior because local proposals are not formally importance-corrected and source protocols are nonexchangeable. The terms `calibration ensemble` and `pseudo-posterior` are used deliberately.

5 S5. Treatment-effect uncertainty and hazard rebasing

The aggregate FSF calibration cannot jointly identify untreated baseline thrombotic hazard and prophylaxis efficacy because source cohorts contain heterogeneous, generally active prophylaxis. Let `beta_A` denote the modeled log-hazard reduction associated with prophylactic exposure. The reference implementation uses a deliberately broad transported distribution anchored to randomized lower-limb-immobilization evidence [6]:

- historical reference coefficient: `beta_A_ref=0.95`;
- transported center: approximately 0.80;
- transported SD: approximately 0.32 after uncertainty inflation; and
- truncation bounds: 0.10 to 1.60.

When a new `beta_A` is drawn, the untreated baseline is rebased to preserve the historical on-treatment calibration:

$$\lambda_{0,new} e^{-\beta_A} = \lambda_{0,ref} e^{-\beta_{A,ref}}.$$

This device propagates treatment-effect uncertainty without implying that an FSF-specific no-prophylaxis contrast has been identified. It is not itself a causal estimator.

6 S6. Decision layer

For each simulated patient,

$$L_i = w_D I(DVT_i \text{ without } PE) + w_P I(PE_i) + w_B I(MB_i) + w_T T_i.$$

PE supersedes DVT in the thrombotic component to avoid double counting. Reference weights are `w_D=0.14`, `w_P=0.19`, `w_B=0.32`, and `w_T=0.008/365` per prophylaxis day. They are abstract research-index weights, not QALYs or universal patient utilities. Prespecified preference scenarios vary thrombotic and bleeding weights.

The following major structural outcomes are omitted from the reference index: fatal PE, distal/proximal DVT severity and propagation, recurrent VTE, PTS, CTEPH, and differentiated critical-site/fatal bleeding. Therefore an apparently low-loss duration is not clinically interpretable.

6.1 S6.1 Finite-policy EVPI

For fixed policies `K={7,14,21,28,35 days}` and uncertain parameter vector `theta`,

$$EVPI = \min_{k \in K} E[L(k,\theta)] - E[\min_{k \in K} L(k,\theta)].$$

The internally designed adaptive scaffold is excluded from the primary EVPI so that an unvalidated adaptive rule cannot define the value of additional research.

6.2 S6.2 EVPPI screens

Single-parameter screening conditions on quantile bins; group screening clusters standardized parameter blocks and calculates conditional optima. These are coarse screens rather than preferred regression-based conditional-expectation estimators [7]. A particle bootstrap evaluates the stability of rankings.

7 S7. Structural EVSI framework

The 72 decision particles are treated as an equally weighted discrete approximation to current uncertainty. A hypothetical study observes selected standardized parameter coordinates. For parameter p, design d, and scenario sample size N,

$$Y_p = z_p(\theta_r) + \epsilon_p, \qquad \epsilon_p \sim N(0, \sigma^2_{p,d}(N)),$$

with

$$\sigma_{p,d}(N) = \frac{1}{s_{p,d}\sqrt{N/300}}.$$

The `s_{p,d}` coefficients are explicit information-strength assumptions, not empirical assay precision. Bayesian reweighting over particles is followed by policy selection using posterior expected loss. Production estimates use 100,000 outer future-study simulations per design/sample-size cell.

Five study designs are compared:

- D0: exact treatment exposure, routine clinical VTE ascertainment, standardized major bleeding;
- D1: D0 plus scheduled bilateral duplex;
- D2: D1 plus longitudinal actual mobility measured separately from prescribed weight-bearing;
- D3: D2 plus repeated coagulation phenotype; and
- D4: D3 plus a causally informative comparison of prophylaxis-duration strategies.

D0-D3 are not permitted to update `beta_A`. D4 represents either randomization where genuine clinical equipoise exists or a prespecified target-trial emulation with defensible identification assumptions [9]. Sample sizes 300, 600, and 1200 are information scenarios and must not be interpreted as sample-size recommendations.

8 S8. Clinical safety and interpretation boundary

The assisted-upright branch applies only to patients individually assessed and cleared by the treating clinician, with a clinically stable intraoperative and postoperative course, no mobilization contraindication, and fixation judged adequate for the specific activity. Patients with hemodynamic instability, neurologic compromise, active bleeding, wound/fixation concerns, or other clinically relevant contraindications are outside that branch. The model cannot override a clinician's restriction.

The decision layer compares hypothetical population policies only. It cannot issue a patient-level instruction. Known DVT remains outside the prophylaxis-stopping cohort and should not be represented as prophylactic dosing.

9 S9. Verification and validation claim boundary

Twenty-two automated tests verify clinical-state invariants, latent-state observability, mobility clearance, treatment-effect prior bounds, hazard rebasing, decision properties, and the separation of observational information from causal treatment-effect learning. Following ISPOR-SMDM terminology, the study distinguishes code verification and internal structural checks from external and predictive validation. It provides code verification, structural plausibility checks, reproduction

of calibration targets, and independent-seed simulation checks [8]; it does not provide external or predictive validation.

The reproducibility archive contains source code, tests, frozen configuration files, calibration particles, policy outputs, VOI/EVSI outputs, evidence matrices, figures, and seeds. No individual-level clinical dataset is distributed or required to reproduce the reported simulations. The public v1.1.0 reproducibility release is archived in Zenodo (doi:10.5281/zenodo.22055590) and mirrored as the GitHub v1.1.0 release.

References

[1] D. You, L. Skeith, R. Korley, et al., Identification of hypercoagulability with thrombelastography in patients with hip fracture receiving thromboprophylaxis, Canadian Journal of Surgery 64 (3) (2021) E324–E329. doi:10.1503/cjs.021019.

[2] W. Yang, Q. Wei, H. Wang, et al., Preoperative incidence and risk factors of deep venous thrombosis in patients with isolated femoral shaft fracture, BMC Surgery 22 (2022) 83. doi:10.1186/s12893-022-01534-x.

[3] Z. Ren, Y. Yuan, W. Qi, Y. Li, P. Wang, The incidence and risk factors of deep venous thrombosis in lower extremities following surgically treated femoral shaft fracture: a retrospective case-control study, Journal of Orthopaedic Surgery and Research 16 (2021) 446. doi:10.1186/s13018-021-02595-z.

[4] E. H. Tischler, J. R. McDermott, A. J. Wolfert, B. Krasnyanskiy, I. Ibrahim, A. N. Malik, J. M. Gross, N. Suneja, Predictors of 30-day mortality, unplanned related readmission and reoperation among isolated closed femoral shaft fractures, Journal of Orthopaedics 55 (2024) 91–96. doi:10.1016/j.jor.2023.10.020.

[5] W. H. Geerts, R. M. Jay, K. I. Code, E. Chen, J. P. Szalai, E. A. Saibil, P. A. Hamilton, A comparison of low-dose heparin with low-molecular-weight heparin as prophylaxis against venous thromboembolism after major trauma, New England Journal of Medicine 335 (1996) 701–707. doi:10.1056/NEJM199609053351003.

[6] A. A. G. Zee, K. van Lieshout, M. van der Heide, L. Janssen, H. M. J. Janzing, Low molecular weight heparin for prevention of venous thromboembolism in patients with lower-limb immobilization, Cochrane Database of Systematic Reviews 8 (2017) CD006681. doi:10.1002/14651858.CD006681.pub4.

[7] A. Heath, I. Manolopoulou, G. Baio, A review of methods for analysis of the expected value of information, Medical Decision Making 37 (2017) 747–758. doi:10.1177/0272989X17697692.

[8] D. M. Eddy, W. Hollingworth, J. J. Caro, J. Tsevat, K. M. McDonald, J. B. Wong, Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7, Medical Decision Making 32 (5) (2012) 733–743. doi:10.1177/0272989X12454579.

[9] M. A. Hernán, J. M. Robins, Using big data to emulate a target trial when a randomized trial is not available, American Journal of Epidemiology 183 (8) (2016) 758–764. https://doi.org/10.1093/aje/kwv254.