Table of Contents

Supplementary Material 1: Diagnosis Codes for Urology included in the dataset

ICD Codes	Description
C60	Malignant neoplasm of penis
C61	Malignant neoplasm of prostate
C62	Malignant neoplasm of testis
C63	Malignant neoplasm of other and unspecified male genital organs
C64	Malignant neoplasm of kidney, except renal pelvis
C65	Malignant neoplasm of renal pelvis
C66	Malignant neoplasm of ureter
C67	Malignant neoplasm of bladder
C68	Malignant neoplasm of other and unspecified urinary organs
N20	Calculus of kidney and ureter
N21	Calculus of lower urinary tract
N22	Calculus of urinary tract in diseases classified elsewhere
N23	Unspecified renal colic
N31	Neuromuscular dysfunction of bladder, not elsewhere classified
N32	Other disorders of bladder
N33	Bladder disorders in diseases classified elsewhere
N34	Urethritis and urethral syndrome
N35	Urethral stricture
N36	Other disorders of urethra
N37	Urethral disorders in diseases classified elsewhere
N39	Other disorders of urinary system
N40	Hyperplasia of prostate
N41	Inflammatory diseases of prostate
N42	Other disorders of prostate
N43	Hydrocele and spermatocele
N44	Torsion of testis
N45	Orchitis and epididymitis
N46	Male infertility
N47	Redundant prepuce, phimosis and paraphimosis
N48	Other disorders of penis
N49	Inflammatory disorders of male genital organs, not elsewhere classified
N50	Other disorders of male genital organs
N51	Disorders of male genital organs in diseases classified elsewhere

Supplementary Material 2: model selection process and Bayesian Zero-inflated negative binomial model definitions

Bayesian analysis has the advantage for reporting and providing feedback to clinicians and healthcare professionals as it offers a much more intuitive interpretation of the parameters by expressing it as probabilities. Credible intervals allow for outcome quality indicators (QIs) to be expressed as a range of expected values based on *n* probability, i.e., 95% credible interval of length of stay (LOS) is between 0 – 9, it means that there is a 95% probability that the LOS is between 0 – 9.

We chose the Bayesian Zero-inflated negative binomial model as it was heavily supported by literature[1,2]. Two distribution candidates were available in modelling zero-inflated distributions such as length of stay (LOS) and hospital acquired complication (HACs) rates, Zero-Inflated Poisson (ZIP) or Zero-Inflated negative binomial (ZINB). We used PSIS-LOOCV[3] (Table 1) as a model selection criteria, and posterior predictive checks were done graphically (Figure 1).[4] As a baseline, we also compared it to linear regression. Overall, ZINB performed the best, followed by ZIP and linear regression.

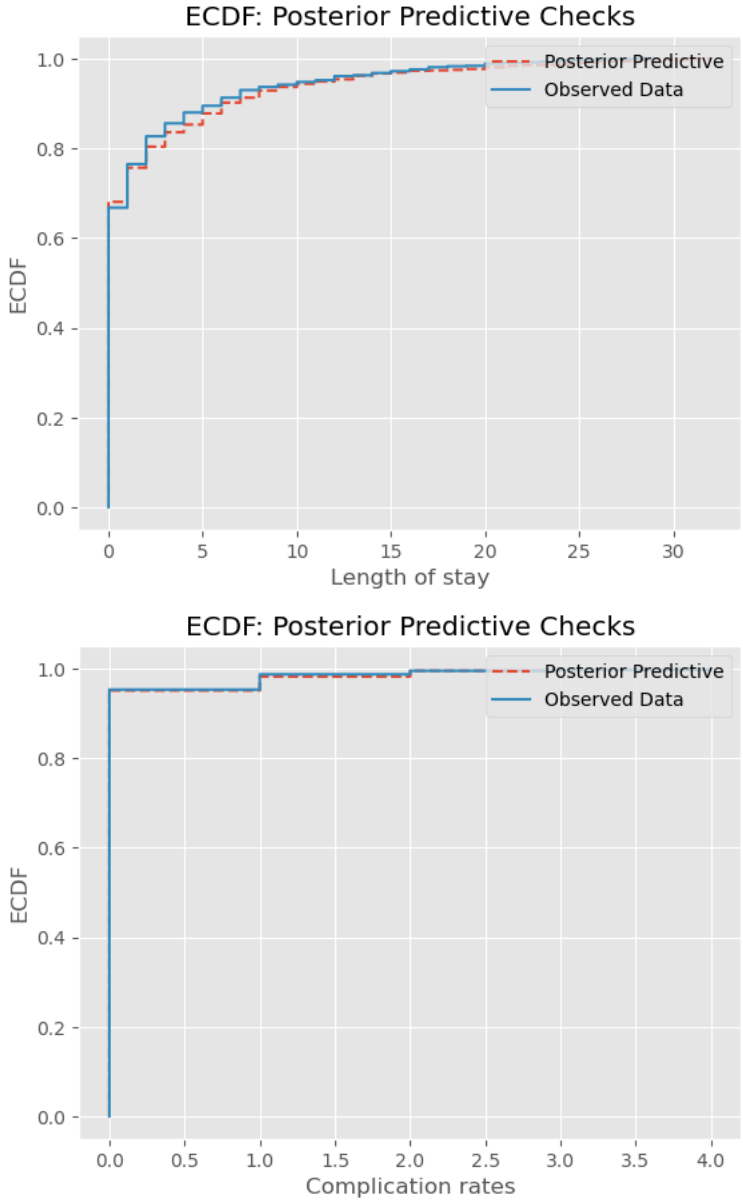


Figure 1: Posterior Predictive Check for both LOS and HACs model

Table 1: Model Performance of Bayesian multiple membership models using PSIS-LOOCV

Models	Length of Stay	HACs Rate
	PSIS-LOO-CV#	
Bayesian MMM* + Linear Regression	954055.14	113110.1
Bayesian MMM + Zero Inflated Poisson Regression	433895.26	67347.14
Bayesian MMM + Zero Inflated Negative Binomial^	**376456.33**	**66879.28**

* MMM: Multiple membership model
^ Best model highlighted in bold.
# PSIS-LOOCV: Pareto smoothed importance sampling leave one out cross validation score.

Model Definition

Let y be the observed count, μ the mean count, and logit the probability of a zero inflation.

Likelihood

The likelihood function $p(y|\mu,\text{logit})$ is based on ZINB:

$$p(y|\mu,\text{logit}) = \text{logit}\delta(y) + (1-\text{logit})\frac{\Gamma(y+1/\omega)}{\Gamma(y+1)\Gamma(1/\omega)}\left(\frac{\mu}{\mu+1/\omega}\right)^{y}\left(\frac{1/\omega}{\mu+1/\omega}\right)^{1/\omega}$$

where $\delta(y)$ is the Dirac delta function and ω is the dispersion parameter.

Linear Predictor and Multiple Membership

We model μ and logit as:

$$\log(\mu) = \mathbf{X\beta} + \mathbf{Zu}$$

$$\text{logit} = \mathbf{X'\beta'} + \mathbf{Z'u'}$$

Here, $\mathbf{u}$ and $\mathbf{u'}$ represent the random effects for ICD groups and comorbidities. $\mathbf{Z}$ and $\mathbf{Z'}$ are design matrices that map these random effects onto the data.

Prior Distribution

For the intercepts and coefficients, we use normal priors centered at 0 and a scale of 1:

$$\beta \sim \mathcal{N}(0,1), \quad \beta' \sim \mathcal{N}(0,1)$$

For the random effects:

$$\mathbf{u} \sim \mathcal{N}(\boldsymbol{\alpha}, \Sigma), \quad \mathbf{u'} \sim \mathcal{N}(\boldsymbol{\alpha'}, \Sigma')$$

Hyperpriors for $\boldsymbol{\alpha}$ and Σ:

$$\boldsymbol{\alpha} \sim \mathcal{N}(0,1), \quad \Sigma \sim \text{Half-Cauchy}(25)$$

$$\boldsymbol{\alpha'} \sim \mathcal{N}(0,1), \quad \Sigma' \sim \text{Half-Cauchy}(25)$$

The choice of Half-Cauchy for Σ and Σ' follows the rationale that it is a weakly informative prior that allows for heavy tails.[5]

Inference Method

We employ mean-field variational Bayes for posterior inference. Given the high dimensionality of our model, traditional MCMC methods are computationally expensive. Variational methods approximate the true posterior with a factorizable distribution, making it computationally efficient.

Sensitivity Analysis

We also tested different weakly informed priors as part of our sensitivity analysis, overall, weakly informative priors of $\mathcal{N}(0,100)$ yield no tangible differences in the final predictive model as well as PSIS-LOOCV results.

Supplementary Material 3: Model Performances

Table 2: Model Performance of LASSO Regression, Random Forest and Bayesian SKIM ZINB using bootstrap.

Quality Indicators	Length of Stay	HACs^ rate
Model	RMSE (95% CI)	RMSE (95% CI)
Linear Regression w/ LASSO (log transformed)	4.42 (4.34 – 4.5)	0.31 (0.30 – 0.32)
Random Forest	2.82 (2.77 – 2.88)	0.28 (0.27 – 0.29)
Bayesian Zero Inflated Negative Binomial with SKIM* Variable Selection	**2.72 (2.65 – 2.77)**	**0.28 (0.27 -0.29)**

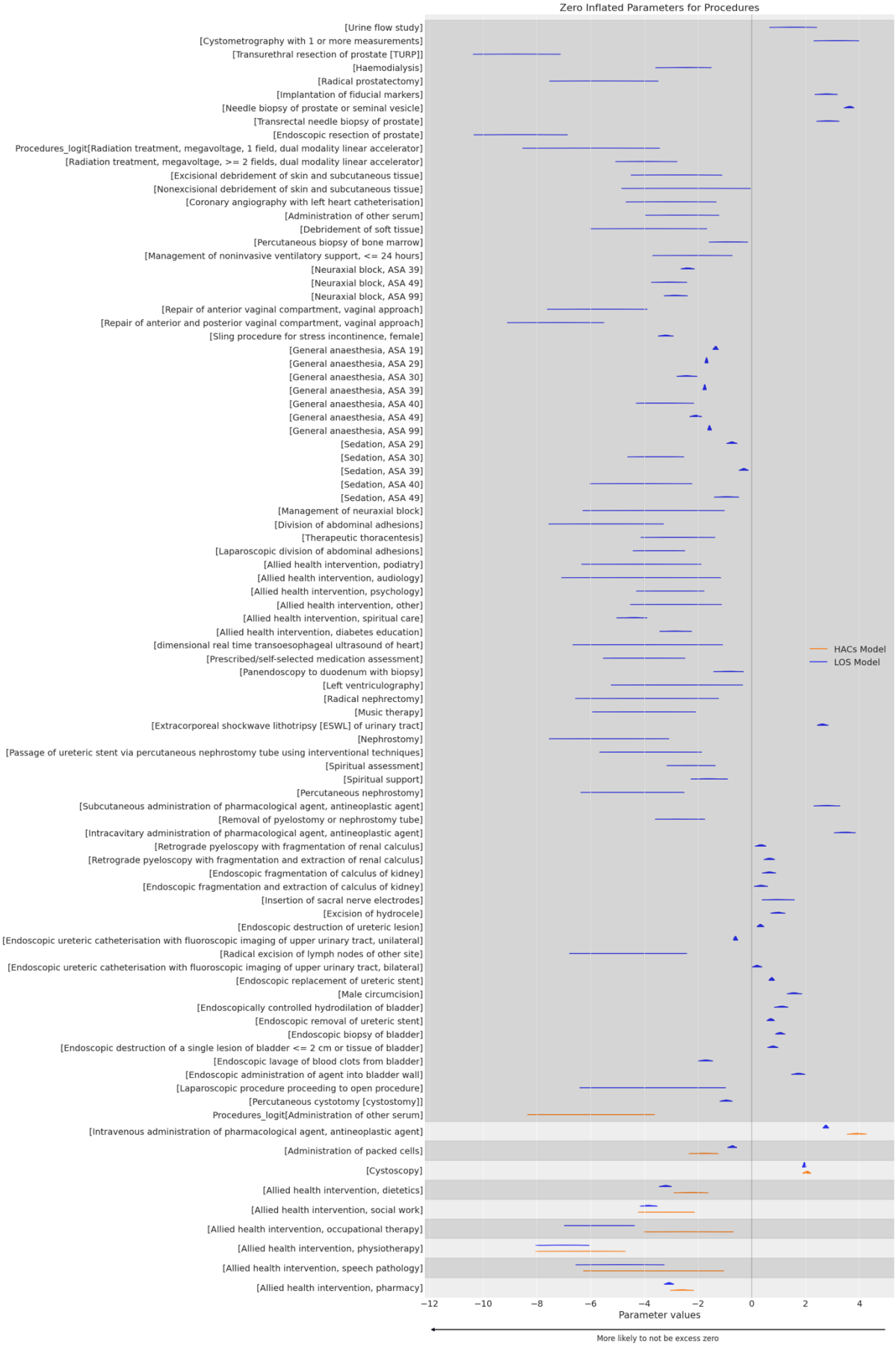
* SKIM: Sparse kernel interaction model
^ HACs: Hospital acquired complications

Table 3: Model Performance of LASSO Regression, Random Forest and Bayesian SKIM ZINB using bootstrap for external dataset.

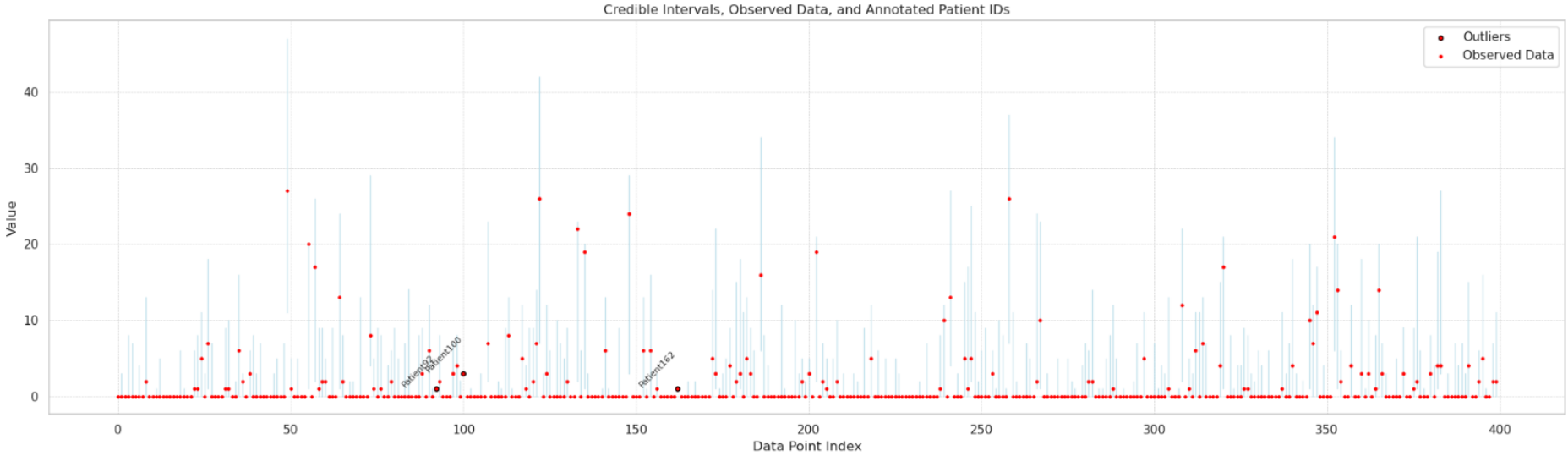
Quality Indicators	Length of Stay	HACs^ rate
Model	RMSE (95% CI)	RMSE (95% CI)
Linear Regression w/ LASSO (log transformed)	3.63 (2.44 - 5.09)	0.24 (0.16 - 0.32)
Random Forest	2.63(1.66 - 3.91)	0.20 (0.14 - 0.26)
Bayesian Zero Inflated Negative Binomial with SKIM* Variable Selection	**1.97 (1.38 - 2.43)**	**0.20 (0.14 -0.25)**

* SKIM: Sparse kernel interaction model
^ HACs: Hospital acquired complications

Supplementary Material 4: Estimates of posterior distributions for procedures related to LOS and HACs


Zero Inflated Parameters for Procedures
[Urine flow study]
[Cystometrography with 1 or more measurements]
[Transurethral resection of prostate [TURP]]
[Haemodialysis]
[Radical prostatectomy]
[Implantation of fiducial markers]
[Needle biopsy of prostate or seminal vesicle]
[Transrectal needle biopsy of prostate]
[Endoscopic resection of prostate]
Procedures_logit[Radiation treatment, megavoltage, 1 field, dual modality linear accelerator]
[Radiation treatment, megavoltage, >= 2 fields, dual modality linear accelerator]
[Excisional debridement of skin and subcutaneous tissue]
[Nonexcisional debridement of skin and subcutaneous tissue]
[Coronary angiography with left heart catheterisation]
[Administration of other serum]
[Debridement of soft tissue]
[Percutaneous biopsy of bone marrow]
[Management of noninvasive ventilatory support, <= 24 hours]
[Neuraxial block, ASA 39]
[Neuraxial block, ASA 49]
[Neuraxial block, ASA 99]
[Repair of anterior vaginal compartment, vaginal approach]
[Repair of anterior and posterior vaginal compartment, vaginal approach]
[Sling procedure for stress incontinence, female]
[General anaesthesia, ASA 19]
[General anaesthesia, ASA 29]
[General anaesthesia, ASA 30]
[General anaesthesia, ASA 39]
[General anaesthesia, ASA 40]
[General anaesthesia, ASA 49]
[General anaesthesia, ASA 99]
[Sedation, ASA 29]
[Sedation, ASA 30]
[Sedation, ASA 39]
[Sedation, ASA 40]
[Sedation, ASA 49]
[Management of neuraxial block]
[Division of abdominal adhesions]
[Therapeutic thoracentesis]
[Laparoscopic division of abdominal adhesions]
[Allied health intervention, podiatry]
[Allied health intervention, audiology]
[Allied health intervention, psychology]
[Allied health intervention, other]
[Allied health intervention, spiritual care]
[Allied health intervention, diabetes education]
[dimensional real time transoesophageal ultrasound of heart]
[Prescribed/self-selected medication assessment]
[Panendoscopy to duodenum with biopsy]
[Left ventriculography]
[Radical nephrectomy]
[Music therapy]
[Extracorporeal shockwave lithotripsy [ESWL] of urinary tract]
[Nephrostomy]
[Passage of ureteric stent via percutaneous nephrostomy tube using interventional techniques]
[Spiritual assessment]
[Spiritual support]
[Percutaneous nephrostomy]
[Subcutaneous administration of pharmacological agent, antineoplastic agent]
[Removal of pyelostomy or nephrostomy tube]
[Intracavitary administration of pharmacological agent, antineoplastic agent]
[Retrograde pyeloscopy with fragmentation of renal calculus]
[Retrograde pyeloscopy with fragmentation and extraction of renal calculus]
[Endoscopic fragmentation of calculus of kidney]
[Endoscopic fragmentation and extraction of calculus of kidney]
[Insertion of sacral nerve electrodes]
[Excision of hydrocele]
[Endoscopic destruction of ureteric lesion]
[Endoscopic ureteric catheterisation with fluoroscopic imaging of upper urinary tract, unilateral]
[Radical excision of lymph nodes of other site]
[Endoscopic ureteric catheterisation with fluoroscopic imaging of upper urinary tract, bilateral]
[Endoscopic replacement of ureteric stent]
[Male circumcision]
[Endoscopically controlled hydrodilation of bladder]
[Endoscopic removal of ureteric stent]
[Endoscopic biopsy of bladder]
[Endoscopic destruction of a single lesion of bladder <= 2 cm or tissue of bladder]
[Endoscopic lavage of blood clots from bladder]
[Endoscopic administration of agent into bladder wall]
[Laparoscopic procedure proceeding to open procedure]
[Percutaneous cystotomy [cystostomy]]
Procedures_logit[Administration of other serum]
[Intravenous administration of pharmacological agent, antineoplastic agent]
[Administration of packed cells]
[Cystoscopy]
[Allied health intervention, dietetics]
[Allied health intervention, social work]
[Allied health intervention, occupational therapy]
[Allied health intervention, physiotherapy]
[Allied health intervention, speech pathology]
[Allied health intervention, pharmacy]
HACs Model
LOS Model
−12
−10
−8
−6
−4
−2
0
2
4
Parameter values
More likely to not be excess zero

Supplementary Material 5: Sample Report



Supplementary Material 5: Shows the credible intervals for predictions of length of stay the first 400 patients for the test set. Patients who are outliers (Patient 92, Patient 100, and Patient 162) are highlighted in black circle and annotated.

Supplementary Material 6: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	D;V	Describe eligibility criteria for participants.	6
	5c	D;V	Give details of treatments received, if relevant.	6
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	6
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	n/a
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	n/a
Sample size	8	D;V	Explain how the study size was arrived at.	10
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	6
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	6
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	7, SM
	10c	V	For validation, describe how the predictions were calculated.	8, SM
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8, SM
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	n/a
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	n/a
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	8
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	11
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	11
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	9
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	17, SM
	15b	D	Explain how to the use the prediction model.	20
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	15
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	n/a
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	22
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	22
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	22
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	SM
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	1

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.