**Table S1. Checklist of TRIPOD statement.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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. | Title |
| Abstract | 2 | D;V | Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | Abstract |
| **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. | Introduction Para1,2 |
| 3b | D;V | Specify the objectives, including whether the study describes the development or validation of the model or both. | Introduction Para 3 |
| **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. | Methods-data source Para 2,3 |
| 4b | D;V | Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.  | Methods-data source Para 2,3 |
| 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. | Methods-data source Para 2,3 and Table S2 |
| 5b | D;V | Describe eligibility criteria for participants.  | Methods-data source Para 2,3 and Fig 1 |
| 5c | D;V | Give details of treatments received, if relevant.  | Table 1 |
| Outcome | 6a | D;V | Clearly define the outcome that is predicted by the prediction model, including how and when assessed.  | Methods-data source Para 2; Methods- Endpoint and Variables Para 1 |
| 6b | D;V | Report any actions to blind assessment of the outcome to be predicted.  | NA; Collected from SEER public database |
| Predictors | 7a | D;V | Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured. | Methods- Endpoint and Variables Para 2 and Table 1 |
| 7b | D;V | Report any actions to blind assessment of predictors for the outcome and other predictors.  | NA; Collected from SEER public database |
| Sample size | 8 | D;V | Explain how the study size was arrived at. | Methods-data source Para 2,3 and Fig 1 |
| 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.  | Methods- Establishment of the Prediction Model Para 1 |
| Statistical analysis methods | 10a | D | Describe how predictors were handled in the analyses.  | Methods- Establishment of the Prediction Model Para 1,4 |
| 10b | D | Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation. | Methods-Establishment of the Prediction Model Para 2-4 |
| 10c | V | For validation, describe how the predictions were calculated.  | Methods- Evaluation of the Prediction Model |
| 10d | D;V | Specify all measures used to assess model performance and, if relevant, to compare multiple models.  | Methods- Evaluation of the Prediction Model |
| 10e | V | Describe any model updating (e.g., recalibration) arising from the validation, if done. | NA |
| Risk groups | 11 | D;V | Provide details on how risk groups were created, if done.  | Methods- Interpretability |
| Development vs. validation | 12 | V | For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.  | Methods-data source Para 3 |
| **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.  | Figure 1 |
| 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.  | Results Para 1, Table 1 |
| 13c | V | For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).  | Table S3 |
| Model development  | 14a | D | Specify the number of participants and outcome events in each analysis.  | Results Para 1,Figure 1  |
| 14b | D | If done, report the unadjusted association between each candidate predictor and outcome. | Results Para 2-4, Table 1 |
| 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). | Results- Random Forest Prediction Website |
| 15b | D | Explain how to use the prediction model. | Results- Random Forest Prediction Website |
| Model performance | 16 | D;V | Report performance measures (with CIs) for the prediction model. | Results-model evaluation; Figure 2 |
| Model-updating | 17 | V | If done, report the results from any model updating (i.e., model specification, model performance). | NA |
| **Discussion** |
| Limitations | 18 | D;V | Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).  | Discussion, Para 7,8 |
| Interpretation | 19a | V | For validation, discuss the results with reference to performance in the development data, and any other validation data.  | Discussion, Para 5 |
| 19b | D;V | Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.  | Discussion,Para 1-4 |
| Implications | 20 | D;V | Discuss the potential clinical use of the model and implications for future research.  | Discussion,Para 5-6 |
| **Other information** |
| Supplementary information | 21 | D;V | Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.  | Results- Random Forest Prediction Website; TableS2-4 |
| Funding | 22 | D;V | Give the source of funding and the role of the funders for the present study.  | NA |

\*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.