BMJ Open

Refinement and validation of a comprehensive clinical diagnostic model (GAMAD) based on gender, age, multitarget circulating tumour DNA methylation signature and commonly used serological biomarkers for early detection of hepatocellular carcinoma: a multicentre, prospective observational study protocol

Tian Yang,1,2 Nanya Wang,1 Fengmei Wang,3 Hongmei Liu,4 Feng Shen,2 Guoyue Lv

ABSTRACT

Introduction Prompt detection of hepatocellular carcinoma (HCC) in patients with chronic liver diseases is critical for enhancing prognosis. Existing imaging techniques and serum markers fall short of clinical needs. This study aims to establish a non-invasive diagnostic model for early HCC detection in the Chinese population.

Methods and analysis This prospective, multicentre, observational study will enrol 2000 participants, including HCC patients, those with chronic liver diseases (hepatitis, cirrhosis and benign liver space-occupying lesions), and healthy individuals. The study will collect demographic data and blood samples, which will be used to test α-fetoprotein (AFP), des-γ-carboxy-prothrombin (DCP) and circulating tumour DNA (ctDNA) methylation. The GAMAD (Gender+Age+Methylation+AFP+DCP) model involving gender, age, ctDNA methylation signature, AFP and DCP will be developed and blindly validated in training and validation sets (1400 and 600 cases, respectively). Primary endpoints include sensitivity, specificity and accuracy (receiver operating characteristic curves; area under the curve value) of GAMAD for HCC and/or high-risk HCC groups. Secondary endpoints involve comparing GAMAD with the established GALAD (Gender+Age+AFP-L3+AFP+DCP) model and each blood index (AFP, DCP and methylation signature) to evaluate: (1) GAMAD’s clinical utility for HCC patients in all stages according to different staging systems; (2) GAMAD’s discrimination ability for patients in various subgroups, including liver cirrhosis (LC) related HCC and LC, hepatitis B virus (HBV) related HCC and HBV, hepatitis C virus (HCV) related HCC and HCV, and non-alcoholic fatty liver disease (NAFLD) related HCC and NAFLD.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study is designed to establish a novel diagnostic model for early detection of hepatocellular carcinoma (HCC), especially among patients with chronic liver diseases in China.
⇒ Detection of multitarget circulating tumour DNA methylation in blood provides the possibility to build a high-performance diagnostic model called GAMAD (Gender+Age+Methylation+AFP+DCP) for HCC.
⇒ The absence of participants from hospitals in other areas of China may affect the generalisability of the GAMAD model to the entire Chinese population.

INTRODUCTION

Primary hepatocellular carcinoma (HCC) is the sixth most commonly diagnosed cancer and the third leading cause of cancer death.1 In China, primary HCC ranks fifth in terms of cancer incidence and the second in terms...
of cancer-related mortality. China accounts for 45.3% of global HCC cases and 47.1% of HCC deaths.2 HCC has a poor prognosis worldwide, with a 5-year survival rate ranging 15%–19% in North America, and only 12.5% in China.3 Due to the lack of distinctive clinical signs in early-stage HCC, most patients are diagnosed at advanced stages. Therefore, early detection of HCC is crucial for improving treatment outcomes and patient prognosis.

In clinical practice, blood-based noninvasive detection methods are often employed to improve the detection rate of HCC. α-Fetoprotein (AFP) is the most commonly used serological test, but its limited sensitivity for early HCC diminishes its value.4–7 Des-γ-carboxy-thrombin (DCP), on the other hand, shows greater sensitivity and specificity than AFP in identifying cirrhosis, chronic hepatitis and HCC. A study indicated that serum DCP exhibited good diagnostic performance for AFP-negative HCC patients, with an area under the curve (AUC) of 0.804, while the combination of AFP and DCP achieved a high sensitivity of 95.1% and specificity of 83.3%, with an AUC of 0.89.8

Despite these advances, existing serum markers fail to meet clinical needs for the diagnosis of high-risk groups and early-stage HCC. Researchers have therefore incorporated demographic characteristics with multiple serum markers to develop prediction models for clinical diagnosis, such as the GALAD (Gender+Age+AFP-L3 (the Lens culinaris agglutinin A-reactive fraction of AFP)+AFP+DCP) model, which is widely used in clinical practice.9 This model based on gender, age and three serological biomarkers of AFP, DCP and AFP-L3 demonstrates good performance. A study showed that the sensitivity and specificity of this model in diagnosing early-stage HCC were 80.2% and 89.7% in UK cohort, 82.1% and 81.6% in Japan cohort, and this model could classify HCC correctly from chronic liver disease with AUC over 0.9 in UK, Japan and Germany cohorts.10 However, the GALAD model exhibited lower performance (AUC=0.891) for discriminating HCC from chronic liver disease in Chinese patients compared with the original UK cohort (AUC=0.97). Moreover, while hepatitis B virus (HBV) infection is the leading aetiology of HCC in China, this GALAD model exhibited poorer performance in patients with HBV aetiology than those with hepatitis C virus (HCV). Meanwhile, the GALAD-C model, which was constructed by optimising the parameters of the GALAD, and the GAAP model, which was constructed on only four variables without AFP-L3, outperformed the GALAD model in diagnostic accuracy.11 In addition, compared with AFP and DCP, AFP-L3 contributed less to the diagnosis of HCC, and several studies found that the inclusion of AFP-L3 did not improve diagnostic performance.11–13 Therefore, the combination of demographic information with multiple serum markers to construct diagnostic models can improve the proportion of early HCC diagnosis compared with traditional clinical methods. Selection of more accurate and stable diagnostic markers may further boost the performance of diagnostic models.

Detection of cell-free DNA (cfDNA) markers using peripheral blood samples has gained significant attention in tumour liquid biopsy in recent years.14–15 cfDNA contains circulating tumour DNA (ctDNA), which is the DNA fragment derived from tumour cells in the peripheral blood of tumour patients and carries genetic information relating to the tumour, thus reflecting tumour burden and providing a large amount of information concerning tumour biology.16–21 ctDNA detection shows considerable clinical practicability in the early detection, disease surveillance and prognosis prediction in HCC.16,22 Furthermore, epigenetic alterations such as DNA methylation play an important role in regulating gene activity in both normal and cancer cells, while cancer cells exhibit hypomethylation at the genome-wide level and hypermethylation at CpG islands with high tumour specificity.23 In HCC, aberrant methylation of tumour suppressor genes plays an important role in hepatocarcinogenesis.24 Chan et al found that hypermethylated RASSF1A sequences were detected in 93% of HCC patients compared with 8% of healthy controls.25 Lu et al constructed a methylation predictive model B which combined four candidate genes (APC, COX2, RASSF1A and miR-203), and demonstrated diagnostic potential by distinguishing HBV-related HCC from healthy controls with AUC of 0.855, sensitivity of 83.3% and specificity of 83.0%.26 Another study established a panel by combining hypermethylated and hypomethylated CpG sites based on genome-wide methylation sequencing of cfDNA and validated that this panel could distinguish HCC from cirrhosis with AUC of 0.956.27 Therefore, ctDNA-based methylation detection shows immense potential for the diagnosis of HCC.28

Based on the heterogeneity and regional characteristics of HCC, mainly due to the difference in aetiology of HCC patients in different countries and regions, it is necessary to establish novel applicable diagnostic methods for early detection of HCC especially based on liver disease patients in China. The aim of this study is to develop and optimise an HCC-discriminating classifier GAMAD based on demographics (Gender/Age), multigene blood ctDNA Methylation testing and commonly used serum biomarkers (AFP/DCP). Through large-scale discovery and validation of performance of the diagnostic model in HCC patients, those with chronic liver diseases, and healthy populations, this study is expected to develop a non-invasive, convenient and high-performance method for early detection of HCC.


test
the enrolled participants will be collected for methylation detection, and 4 mL of whole blood of the enrolled participants will be collected for AFP, AFP-L3 and DCP detection. The dataset will be divided into a training and a validation cohort. In the training set with a total of 1400 cases, 420 are HCC patients, 840 are high-risk HCC patients (hepatitis, cirrhosis), and 140 are benign liver SOLs and healthy individuals with approximately half of them being healthy individuals. An early diagnostic model (GAMAD) will be constructed using gender, age, AFP, DCP and multitarget blood ctDNA methylation testing as its core. The diagnostic performance of the model for HCC will be evaluated. The constructed model will be validated using an independent blind-testing set. This validation set will comprise 600 blood samples taken from 180 HCC patients, 360 high-risk HCC patients (hepatitis, cirrhosis) and 60 benign liver SOLs and healthy individuals with approximately half of them being healthy individuals. The technicians performing the AFP, AFP-L3, DCP and ctDNA methylation tests will be blinded to the diagnosis of the participants. After locking the database, the blind will be opened to verify the performance of GAMAD and compare it with GALAD (an established model) and each blood index (AFP, DCP and ctDNA methylation signature). The classifier's accuracy and utility will be further assessed in the entire cohort (training and validation cohort) to compare its clinical significance with GALAD for HCC patients in early and other stages. An overview of the trial is shown in figure 1. Recommendations for developing and validating a prediction model (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) checklist are available as online supplemental file 1.

**Study sites and recruitment procedures**

This clinical trial involves three hospitals in China. Physicians will screen patients based on clinical data to assess whether they meet the inclusion and exclusion criteria. Subjects who are eligible for and who agree to participate in the study are notified and recruited prior to the start of trial.

**Study population and inclusion/exclusion criteria**

This trial will prospectively enrol 2000 participants with liver diseases or healthy people including: 600 patients in the HCC group, 1400 patients in the chronic liver diseases (cirrhosis, hepatitis and benign liver SOLs) and healthy individuals. Patients with chronic liver diseases will be followed up regularly during the study to ensure none of these patients have evidence of HCC. Patients are eligible if they fit the following inclusion and exclusion criteria.

**Criteria for inclusion**

1. Aged above 18.
2. A definite diagnosis of hepatitis or cirrhosis or HCC.
3. Patients who have recently been diagnosed with HCC and have not received any treatment (local or systematic).
4. Able to provide sufficient and qualified blood samples for testing.
5. Participants who would be willing to complete the detection of tumour markers (AFP, DCP and AFP-L3), and/or serum markers of viral hepatitis, and/or concentration of HBV DNA and/or HCV RNA, and/or imaging tests of the liver, and/or transabdominal ultrasound, etc. according to diagnostic criteria and study design.
6. Voluntarily participating in this trial with a written informed consent form.

**Criteria for exclusion**

1. Obstructive jaundice.
2. HCC of China liver cancer stage IV or Barcelona Clinic liver cancer (BCLC) stage D.
3. Medical history of taking warfarin, vitamin K or anticoagulant drugs treatment within 1 month.
4. With other known malignant tumours or multiple primary tumours.
5. Patients with autoimmune diseases, genetic diseases, mental diseases/disabilities and other diseases considered unsuitable for the study.
6. Patients with Child-Pugh C liver function.
7. Combined hepatocellular and intrahepatic cholangiocarcinoma.
8. Decompensated chronic liver diseases.
9. During pregnancy or lactation.
10. Recipient of blood transfusion within 3 months prior to study blood draw.
11. Insufficient qualified blood sample for testing.

**Study procedures**

**Screening period**

During the screening period, participants' demographic data, current medical history, family history, medication history and participation in relevant clinical trials will be collected to determine their eligibility for inclusion. Eligible participants will be informed about the research design, the type and quantity of biological samples collected, security and confidentiality measures and asked to sign the informed consent form. A detailed schedule of screening, measurement, and clinical data collection is presented in [table 1](#).

**Clinical data collection and blood biomarkers detection**

Clinical data, including demographic data, medical history data and/or serum markers of viral hepatitis (HBV markers, HCV antibodies), and/or concentration of HBV DNA and/or HCV RNA, and/or imaging tests of the liver (MRI and/or CT), and/or transabdominal ultrasound, etc. will be collected based on the diagnostic criteria and study design. Peripheral blood samples will be collected from each participant for measuring AFP, DCP, AFP-L3 and ctDNA methylation. Blood samples will be separated by centrifugation, and serum and plasma will be immediately frozen.

---

Figure 1  Flowchart of study design. AFP, α-fetoprotein; AFP-L3, the Lens culinaris agglutinin A-reactive fraction of AFP; AUC, area under the curve; cHCC-ICC, combined hepatocellular and intrahepatic cholangiocarcinoma; ctDNA, circulating tumour DNA; DCP, des-γ-carboxy-prothrombin; GALAD, Gender+Age+AFP-L3+AFP+DCP; GAMAD, Gender+Age+Methylation+AFP+DCP; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; ROC, receiver operating characteristic; SOLs, space-occupying lesion.
events related to serum sample collection are expected to be minimal.

For patients with HCC, additional information such as pathological stage, tumour diameter and number, lymph node metastasis, distant metastasis, vascular invasion and physical status score of the tumour will be collected for hierarchical analysis.

Sample size calculation
The enrolled cases are generally divided into two groups: HCC group and control group (including cirrhosis, hepatitis, benign SOLs and healthy individuals). The sensitivity of GAMAD model for the diagnosis of HCC in the training set is estimated to be 70%, and specificity is 90%. The HCC and the control group would be allocated according to 3:7 (α=0.05, tolerance error=10%). The Power and Sample Size software is used to calculate the sample size and the total number of training set evaluated is 1137 cases. Considering the drop-off of subjects and the occurrence of potential sample disqualification, the probability of which is about 20%, 1400 cases will be enrolled in the training set. Therefore, a total of about 2000 participants (training set and validation set with a ratio of 7:3) will be recruited.

Recruitment of participants
Recruitment will take place at the outpatient department of The First Hospital of Jilin University, the Eastern Hepatobiliary Surgery Hospital, Naval Medical University and Tianjin Third Central Hospital. Outpatients will be recruited via posters in the hospital and via Internet advertisement.

Patient and public involvement
All aspects of this study (development of the research question, study design and conduct of the trial, interpretation of results and editing of the final manuscript for publication) are taking place independently of patients and public involvement. The results will be disseminated to participants by their physicians.

Statistical aspects
Log transformation will be performed for variables with right-skewed distributions. Univariate and multivariate binary logistic regression analyses will be used to further analyse the efficacy of each parameter in identifying patients with chronic liver diseases and HCC. GAMAD model (Gender+Age+Methylation+AFP+DCP) will be constructed using rms package based on the independent predictors from the multivariate logistic regression analyses. A multi-parameter diagnostic model (GAMAD) of HCC, with quantitative function, based on gender, age, logit value of polygene methylation transformed by logistic regression, quantitative detection value of AFP and DCP will be developed. Receiver operating characteristic (ROC) curves will be applied to calculate the best cut-off points and to determine the AUC for AFP, DCP and ctDNA methylation alone or all as predictors. The AUC value, sensitivity and specificity will be used to

Table 1 Schedule of screening, measurement and clinical data collection

<table>
<thead>
<tr>
<th>Inspection items</th>
<th>Study period</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Screening</td>
<td>Measurement and clinical data collection</td>
</tr>
<tr>
<td>Initial assessment of patients</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Signing the informed consent form</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Demographic data collection</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Medical history data collection</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Routine blood test</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Liver function</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Blood coagulation function</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Blood lipid</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Detection of AFP, DCP and AFP-L3</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Viral hepatitis type B testing</td>
<td>Serum markers of HBV</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Concentration of HBV DNA</td>
<td>x</td>
</tr>
<tr>
<td>Viral hepatitis type C testing</td>
<td>Serum markers of HCV</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>Concentration of HCV DNA</td>
<td>x</td>
</tr>
<tr>
<td>Imaging tests of the liver</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Transabdominal ultrasound</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Sample collection for blood polygene methylation testing</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Detection of blood polygene methylation</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

AFP, α-fetoprotein; AFP-L3, the Lens culinaris agglutinin A-reactive fraction of AFP; DCP, des-γ-carboxy-prothrombin; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus.

at −80°C until testing. The sample storage facilities and conditions will be standardised. The measurements of AFP, DCP and AFP-L3 will be performed in serum samples using a magnetic particle-based chemiluminescence immunoassay on a full-automatic fluorescence analyser (MQ 60 plus, Hotgen Biotech, Beijing) with defined protocols and in the same testing centre. Then, separated plasma samples will be processed for ctDNA methylation detection with the Multi-Gene Methylation Detection Kit for Human Liver Cancer (PCR-Fluorescence Probing) (Cat No.: IVD21QP010024CE, Singlera Genomics (Shanghai) Ltd, China). Briefly, the main steps include DNA extraction, bisulfite conversion, and PCR which are carried out according to the instructions provided by the kit. Finally, the risk scores will be determined by the kit followed by the corresponding manufacturer’s protocols. The technicians performing the laboratory tests will be blinded to the diagnosis of the participants. Adverse
GAMAD model score will be obtained according to the optimal model. Statistical analysis will be performed by the SPSS software V.22.0 and R (V.4.13). The AUC, sensitivity and specificity will be used to report diagnostic accuracy. We will compare the diagnostic performance of GAMAD with the GALAD scoring model, the multi-gene methylation detection technology for HCC, AFP and DCP for diagnosis of HCC.

**Outcome measure**

The primary outcome measure is the sensitivity, specificity and accuracy (ROC curves; AUC value) of the diagnostic model (GAMAD) for HCC (in early and each stage). The secondary outcome measures involve comparing GAMAD with the established GALAD model and each blood index (AFP, DCP and methylation signature) to evaluate: (1) GAMAD's clinical utility and significance for HCC patients in early and other stages (according to different staging systems: tumour, node, metastases staging, the Milan criteria and the BCLC staging criteria); (2) GAMAD’s discrimination ability for patients in various subgroups including liver cirrhosis (LC) related HCC and LC, HBV related HCC and HBV, HCV related HCC and HCV, and non-alcoholic fatty liver disease (NAFLD) related HCC and NAFLD.

**DISCUSSION**

The aim of this study is to establish a diagnostic model referred to as ‘GAMAD’ for early detection of HCC among the Chinese population. The model encompasses gender, age, multitarget blood ctDNA methylation testing, AFP and DCP with the objective of enhancing the sensitivity, specificity, and accuracy of HCC diagnosis at each stage.

In a previous study, we constructed a nomogram model including AFP, DCP, age and sex to predict the risk of HCC in patients with chronic hepatitis B infection. This model showed strong predictive performance in the training set (AUC, 0.941 (95% CI 0.929 to 0.952)) and was successfully validated in the validation set (AUC, 0.931 (95% CI 0.909 to 0.953)). In this study, we incorporate ctDNA methylation as an additional factor because aberrant methylations of DNA have proven to be a key driver in the development of HCC. This suggests that blood markers based on DNA methylation could serve as an effective complement to the diagnosis of HCC. By combining objective clinical and serological factors with multitarget ctDNA methylation signature, our comprehensive diagnostic model (GAMAD) has the potential to benefit patients with chronic liver disease in China.

To ensure that our study reflects real-world clinical practice, we will exclude patients with chronic liver disease who develop HCC during follow-up. Furthermore, our focus will extend beyond just enrolled patients with LC, as the occurrence of HCC is related to chronic liver disease, not limited to the cirrhosis stage. Patients with decompensated liver diseases were also excluded from this study, as their clinical needs primarily revolve around improving the life signs and functions of organs, reducing the reinjury of tissues, or considering liver transplantation, rather than HCC diagnosis.

This study is conducted in three hospitals located in northern, northeastern and eastern China, with the exclusion of hospitals from other regions. As a result, the study may not comprehensively represent the entire Chinese population. The exclusion of hospitals from other regions may introduce regional bias, thereby impacting the generalisability of the GAMAD model to the entire Chinese population. Therefore, further studies that encompass participants from diverse regions are necessary to validate the clinical applicability of the GAMAD model. Nonetheless, we maintain that the inclusion of participants from various regional backgrounds in our study still renders the GAMAD model theoretically reasonable and applicable to the Chinese population.

**Trial status**

The study was conceived and designed in November 2022. Enrolment began in 2023 and is expected to end in May 2024. At the time of manuscript preparation, more than 500 subjects had been enrolled. Enrolment in this study was ongoing at the time of manuscript submission.

**ETHICS AND DISSEMINATION**

This trial has been approved by the Medical Ethics Committees of the First Hospital of Jilin University (#22K073-001), the Eastern Hepatobiliary Surgery Hospital, Naval Medical University (#EHBHXY2023-H0003-P001) and Tianjin Third Central Hospital (#IRB2023-007-001). All participants in the trial will provide written informed consent. Results of this study will be disseminated in peer-reviewed scientific journals and at conferences nationally and internationally.

**Author affiliations**

1Department of Hepatobiliary and Pancreatic Surgery, The First Hospital of Jilin University, Changchun, China
2Department of Hepatic Surgery, Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China
3Department of Gastroenterology and Hepatology, Tianjin Third Central Hospital, Tianjin, China
4Department of Medical Affairs, Singlera Genomics (Shanghai) Ltd, Shanghai, China

**Acknowledgements**

We thank all subjects for their voluntary participation in this study and physicians at all recruiting sites.

**Contributors**

YL and TY conceived the idea and secured funding. NW was involved in the drafting of the manuscript. FS was involved in the study concept and design and in the revision of the manuscript. FW and HL provided intellectual input concerning study design and statistical analysis. All authors were involved in drafting or critically revising this work, and in final approval of the version to be published.

**Funding**

This study was supported by the National Natural Science Foundation of China (No. 81 972 726 and 82 273 074 for TY; 82 241 223 and U20A20360 for GL), Dawn Project Foundation of Shanghai (No. 215G63 for TY), Shanghai Health and Hygiene Discipline Leader Project (No. 2022D001 for TY), Shanghai Outstanding
Academic Leader Programme (No. 23XD1424900 for TY) and Singlera Genomics (grant number not provided).

Disclaimer The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests HL reported being an employee of Singlera Genomics (Shanghai). Multi-Gene Methylation Detection Kit for Human Liver Cancer (PCR- Fluorescence Probing) used in this trial was supplied by Singlera Genomics (Shanghai). On manuscript submission, all authors completed the author disclosure form.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

REFERENCES


5. Marrero JA, El-Serag HB. Alpha-Fetoprotein should be included in the hepatocellular carcinoma surveillance guidelines of the American Association for the study of liver diseases. HEPATOLOGY 2011;53:1060–1.


TRIPOD Checklist: Prediction Model Development and Validation

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item</th>
<th>Checklist Item</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>1</td>
<td>D;V</td>
<td>P1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
<td>D;V</td>
<td>P2</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background and objectives</td>
<td>3a</td>
<td>D;V</td>
<td>P4, P5</td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source of data</td>
<td>4a</td>
<td>D;V</td>
<td>P6, P7</td>
</tr>
<tr>
<td>Participants</td>
<td>5a</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Outcome</td>
<td>6a</td>
<td>D;V</td>
<td>P11</td>
</tr>
<tr>
<td>Predictors</td>
<td>7a</td>
<td>D;V</td>
<td>P7</td>
</tr>
<tr>
<td>Sample size</td>
<td>8</td>
<td>D;V</td>
<td>P10</td>
</tr>
<tr>
<td>Missing data</td>
<td>9</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Statistical analysis methods</td>
<td>10a</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>10b</td>
<td>D;V</td>
<td>P11</td>
</tr>
<tr>
<td></td>
<td>10c</td>
<td>V</td>
<td>P11</td>
</tr>
<tr>
<td></td>
<td>10d</td>
<td>D;V</td>
<td>P11</td>
</tr>
<tr>
<td></td>
<td>10e</td>
<td>V</td>
<td>P11</td>
</tr>
<tr>
<td>Risk groups</td>
<td>11</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Development vs. validation</td>
<td>12</td>
<td>V</td>
<td>P6, P7, Figure 1</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>13a</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Model development</td>
<td>14a</td>
<td>D; V</td>
<td>N/A</td>
</tr>
<tr>
<td>Model specification</td>
<td>15a</td>
<td>D</td>
<td>N/A</td>
</tr>
<tr>
<td>Model-updating</td>
<td>17</td>
<td>V</td>
<td>N/A</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>18</td>
<td>D;V</td>
<td>P13</td>
</tr>
<tr>
<td>Interpretation</td>
<td>19a</td>
<td>V</td>
<td>P12</td>
</tr>
<tr>
<td>Implications</td>
<td>20</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Other information</td>
<td>21</td>
<td>D;V</td>
<td>N/A</td>
</tr>
<tr>
<td>Funding</td>
<td>22</td>
<td>D;V</td>
<td>P14</td>
</tr>
</tbody>
</table>

*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.