# Diagnosis of neurosyphilis in HIV-negative patients with syphilis: development, validation, and clinical utility of a suite of machine learning models

Huachun Zou,[a,i] Zhen Lu,[a,i] Wenjia Weng,[b,i] Ligang Yang,[c,i] Luoyao Yang,[a,i] Xinying Leng,[c] Junfeng Wang,[d,e] Yi-Fan Lin,[a] Jiaxin Wu,[c] Leiwen Fu,[a] Xiaohui Zhang,[c] Yuwei Li,[a] Liuyuan Wang,[c] Xinsheng Wu,[a] Xinyi Zhou,[a] Tian Tian,[a] Lixia Huang,[c] Christina M. Marra,[f,*] Bin Yang,[c,**] Tian-Ci Yang,[g,h,***] and Wujian Ke[c,****]

[a]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, 518107, China
[b]Department of Dermatology, Beijing Youan Hospital, Capital Medical University, Beijing, 100012, China
[c]Department of STD Clinic, Dermatology Hospital of Southern Medical University, Guangzhou, 510091, China
[d]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
[e]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands
[f]Department of Neurology, University of Washington, Seattle, WA, 98104, USA
[g]Center of Clinical Laboratory, Zhongshan Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, 361004, China
[h]Institute of Infectious Disease, School of Medicine, Xiamen University, Xiamen, 361004, China

## Summary

**Background** The ability to accurately identify the absolute risk of neurosyphilis diagnosis for patients with syphilis would allow preventative and therapeutic interventions to be delivered to patients at high-risk, sparing patients at low-risk from unnecessary care. We aimed to develop, validate, and evaluate the clinical utility of simplified clinical diagnostic models for neurosyphilis diagnosis in HIV-negative patients with syphilis.

**Methods** We searched PubMed, China National Knowledge Infrastructure and UpToDate for publications about neurosyphilis diagnostic guidelines in English or Chinese from database inception until March 15, 2023. We developed and validated machine learning models with a uniform set of predictors based on six authoritative diagnostic guidelines across four continents to predict neurosyphilis using routinely collected data from real-world clinical practice in China and the United States (through the Dermatology Hospital of Southern Medical University in Guangzhou [659 recruited between August 2012 and March 2022, treated as Development cohort], the Beijing Youan Hospital of Capital Medical University in Beijng [480 recruited between December 2013 and April 2021, treated as External cohort 1], the Zhongshan Hospital of Xiamen University in Xiamen [493 recruited between November 2005 and November 2021, treated as External cohort 2] from China, and University of Washington School of Medicine in Seattle [16 recruited between September 2002 and April 2014, treated as External cohort 3] from United States). We included all these patients with syphilis into our analysis, and no patients were further excluded. We trained eXtreme gradient boosting (XGBoost) models to predict the diagnostic outcome of neurosyphilis according to each diagnostic guideline in two scenarios, respectively. Model performance was measured through both internal and external validation in terms of discrimination and calibration, and clinical utility was evaluated using decision curve analysis.

**Findings** The final simplified clinical diagnostic models included neurological symptoms, cerebrospinal fluid (CSF) protein, CSF white blood cell, and CSF venereal disease research laboratory test/rapid plasma reagin. The models showed good calibration with rescaled Brier score of 0.99 (95% CI 0.98–1.00) and excellent discrimination (the minimum value of area under the receiver operating characteristic curve, 0.84; 95% CI 0.81–0.88) when externally validated. Decision curve analysis demonstrated that the models were useful across a range of neurosyphilis

*Corresponding author. Harborview Medical Center, Box 359775, 325 9th Avenue, Seattle, WA 98104, USA.
**Corresponding author. No.2 Lujing Road, Yuexiu district, Dermatology Hospital of Southern Medical University, Guangzhou, 510091, China.
***Corresponding author. No.1, Houbin Road, Siming district, Director of Institute of Infectious Disease, Medical College of Xiamen University/ Director of Clinical Laboratory, Zhongshan Hospital, Medical College of Xiamen University, Xiamen, 361004, China.
****Corresponding author. No.2 Lujing Road, Yuexiu district, Dermatology Hospital of Southern Medical University, Guangzhou, 510091, China.
    *E-mail addresses:* cmarra@uw.edu (C.M. Marra), yangbin101@hotmail.com (B. Yang), yangtianci@xmu.edu.cn (T.-C. Yang), kewujianstauch@163.com (W. Ke).
[i]Huachun Zou, Zhen Lu, Wenjia Weng, Ligang Yang and Luoyao Yang contributed equally to this work.
Translation: For the Chinese language translation of the Summary, see the Supplementary Materials section.

probability thresholds between 0.33 and 0.66 compared to the alternatives of managing all patients with syphilis as if they do or do not have neurosyphilis.

**Interpretation** The simplified clinical diagnostic models comprised of readily available data show good performance, are generalisable across clinical settings, and have clinical utility over a broad range of probability thresholds. The models with a uniform set of predictors can simplify the sophisticated clinical diagnosis of neurosyphilis, and guide decisions on delivery of neurosyphilis health-care, ultimately, support accurate diagnosis and necessary treatment.

**Keywords:** Neurosyphilis; Syphilis; Clinical diagnostic model; Risk-differentiated; Diagnosis

---

**Research in context**

**Evidence before this study**
We searched PubMed, China National Knowledge Infrastructure (CNKI) and UpToDate (available at: https://www.uptodate.com/) for publications about neurosyphilis diagnostic guidelines in English or Chinese from database inception until March 15, 2023, using the following search terms ("neurosyphilis" OR "syphilis") AND ("syphilis and treatment" OR "neurosyphilis and treatment"). Reference of relevant articles and reviews were also screened for additional publications. 26 publications in English or Chinese about neurosyphilis diagnostic guidelines were identified, from which six authoritative diagnostic guidelines worldwide were selected. As there is no universal gold clinical standard globally to diagnose neurosyphilis so far, current diagnostic guidelines are replete with various distinct essential prerequisites that may be not readily available across different clinical settings and population from different regions in the world, especially for those from resource-limited regions. In addition, no prediction modelling studies on diagnosis of neurosyphilis in HIV-negative patients with syphilis were identified to April 21, 2023. Previous studies of neurosyphilis risk primarily focused on significant predictors of neurosyphilis diagnosis.

**Added value of this study**
The final simplified clinical diagnostic models developed in this study accurately predict the individualised risk of

diagnostic outcome of neurosyphilis in HIV-negative patients with syphilis, with good calibration and excellent discrimination. The machine learning models were developed and validated using routinely collected data from real-world clinical practice in China and the United States, with a uniform set of predictors based on six authoritative diagnostic guidelines across four continents. Their performance was maintained when internally and externally evaluated. The models have clinical utility across a broad range of probability thresholds, to assist in shared decision-making for personalised and risk-differentiated care. We developed a free online app to make these models widely available and easily applicable in practical use.

**Implications of all the available evidence**
Our models with a uniform set of predictors can simplify the sophisticated clinical diagnosis of neurosyphilis, and facilitate shared decision-making at the individual level and risk-differentiated care at a health service level, ultimately, supporting more personalised care for patients with syphilis. To promote translation into real-world clinical care, these models have been translated into an online clinical risk app allowing clinicians to calculate individualised risks of neurosyphilis according to six authoritative diagnostic guidelines, to guide decisions of risk-differentiated approaches to care and treatment.

---

## Introduction

Neurosyphilis is a clinically serious disease caused by the infection of central nervous system (CNS) by *Treponema pallidum* subspecies *pallidum* (hereafter, *T. pallidum*).[1] After initial infection, *T. pallidum* disseminates within days[2] and invades CNS in approximately 30% of patients with untreated primary and secondary syphilis.[3] If the organism is not cleared from the CNS, complicated syphilis, including early or late neurosyphilis, oucular or otic syphilis may ensue.[4] Early neurosyphilis includes asymptomatic neurosyphilis, syphilitic meningitis, and meningovascular syphilis,[1] which occurs weeks to month to the first few years after *T. pallidum* initial infection.[5] Late neurosyphilis includes general paresis and tabes dorsalis,[1] which occur years to decades after initial infection.[5]

There are few population based epidemiological data on neurosyphilis. In the pre-antibiotic era, about 30% of individuals with untreated syphilis developed neurosyphilis, of which 30% were asymptomatic neurosyphilis.[6] In another study, 9.50% of untreated early syphilis developed late neurosyphilis.[7] More recent data showed that the prevalence of confirmed or suspected neurosyphilis among primary, secondary, and early latent syphilis in the United States from 2009 to 2015 was 0.84%.[8] The incidence rate of neurosyphilis in Netherlands from 1999 to 2010 was 0.47 cases per 100,000 adults.[9] In British Columbia, Canada, the reported incidence of neurosyphilis increased 26.67-fold from 0.03 cases per 100,000 adults in 1992 to 0.80 cases per 100,000 adults in 2012.[10] In Guangdong province of China, the reported incidence of late neurosyphilis increased 1.48-fold from 0.21 cases per 100,000 adults in 2009 to 0.31 cases per 100,000 adults in 2014.[11] Between May 2013 and May 2020, neurosyphilis was found to have a prevalence of 3.10% in a tertiary university hospital located in Southern Italy.[12] Meanwhile, Dombrowski et al. reported 68 cases of possible neurosyphilis among 573 syphilis cases in King County, WA, from 3rd January 2012 to 30th September 2013.[13] The true burden of neurosyphilis worldwide is likely underestimated due to underreporting and lack of recognition.[14]

There is no one test that can rule in or rule out the diagnosis of neurosyphilis,[15] but several diagnostic guidelines are proposed.[16–23] These guidelines may not be applicable across different clinical settings and population from different regions in the world. For example, a national cross-sectional study of 398 hospitals located in 116 cities in China showed that only 154 (38.69%) hospitals could perform neurosyphilis diagnostic laboratory tests [i.e., venereal disease research laboratory (VDRL)/rapid plasma reagin tests (RPR)/toluidine red unheated serum tests (TRUST), or *T. pallidum* particle agglutination or haemagglutination tests (TPPA or TPHA)].[24] Moreover, although cerebrospinal fluid (CSF) VDRL is currently considered as the definitive standard diagnosis test of confirming neurosyphilis, it requires specialized glass plates and a light microscope in practice, which might be difficult to meet in resource-limited hospitals, and its whole testing process is time-consuming and cumbersome, which restricts its availability further in real-world clinical practice.

Ascertaining a diagnosis of neurosyphilis typically necessitates a confluence of neurological or neuropsychiatric symptoms and signs, laboratory assessments of both blood and cerebrospinal fluid, and in some instances, imaging evaluations. In light of the complexity involved in arriving at a diagnosis, the involvement of skilled clinicians as well as specialized testing equipment and reagents is essential. The diagnostic challenges associated with neurosyphilis in various countries or regions frequently result in a scarcity of reliable epidemiological data. In this context, our objective is to leverage the potential of machine learning to fashion a suite of neurosyphilis diagnostic models that are highly simplified and practical, thereby enabling epidemiological inquiries employing extensive clinical data drawn from the real world. These models have been constructed based on six guidelines spanning four continents, drawing upon multicentre clinical data derived from patients who were clinically suspected of neurosyphilis, and incorporating minimal predictors of neurosyphilis that are applicable regardless of the specific diagnostic criteria employed. Our overarching aim is to develop an accurate and user-friendly diagnostic model and app for neurosyphilis that is suitable for deployment in diverse geographical regions.

## Methods
### Identification of neurosyphilis diagnostic guidelines
We searched PubMed, China National Knowledge Infrastructure (CNKI) and UpToDate (available at: https://www.uptodate.com/) for publications about neurosyphilis diagnostic guidelines in English or Chinese from database inception until March 15, 2023, using the following search terms ("neurosyphilis" OR "syphilis") AND ("syphilis and treatment" OR "neurosyphilis and treatment"). Reference of relevant articles and reviews were also screened for additional publications. 26 publications[16–23,25–42] in English or Chinese about neurosyphilis diagnostic guidelines were identified. The details for 26 publications are given in Supplementary Table S4, from which six diagnostic guidelines (China 2020,[18] selected as a representative of Asia; Europe 2020,[23] selected as a representative of Europe; Australia 2022,[21] selected as a representative of Oceania; and UpToDate 2020,[19] US CDC 2018 (Case Definitions),[16] US 2021 (Treatment Guidelines)[17] selected as representatives of America) were identified (Supplementary Table S1); we were unable to identify a guideline from Africa. The detailed considerations are given in Supplementary Appendix S1.

Various guidelines have different number of diagnostic classifications of neurosyphilis (Supplementary Table S1), which posed challenges for our modeling analysis that aimed to simplify the real-world diagnosis of neurosyphilis and were designed for binary classification. However, we intended to use information of differential risk indications from different number of diagnostic classifications of the guidelines. In this regard, we transformed six national and international guidelines into binary diagnostic outcome variables for considering two possible scenarios. Specifically, Scenario 1 encoded non-neurosyphilis as "0" (i.e., non-neurosyphilis) while verified/probable/possible neurosyphilis as "1" (i.e., neurosyphilis); Scenario 2 encoded

non-neurosyphilis and probable/possible neurosyphilis as "0" (i.e., non-neurosyphilis), while verified neuro-syphilis as "1" (i.e., neurosyphilis). These rules were also depicted in Supplementary Table S2. The detailed considerations of these transformations are given in Supplementary Appendix S1.

### Choice of candidate predictors

Based on consensus of five syphilis experts from both China and the US (CMM, TCY, LY, BY and WK) and systematic review of published studies of neurosyphilis diagnosis, the following candidate predictors were evaluated for inclusion in machine learning models: reactivity of serum non-treponemal test (TRUST and RPR), CSF protein concentration, CSF white blood cell concentration (WBC), CSF treponemal test reactivity (CSF TPPA and CSF fluorescent treponemal antibody-absorption [FTA-ABS]), CSF non-treponemal test reactivity (CSF VDRL, CSF TRUST, and CSF RPR) and neurological symptoms or signs (neurologic, otologic, or ocular symptoms or signs consistent with neuro-syphilis). The detailed considerations are given in Supplementary Appendix S1. Notably, age and sex were not selected for inclusion in the models, due to the minor contributions to the clinical diagnosis of neurosyphilis.

### Development and validation patient cohorts

We obtained participant-level data of the confirmed HIV-negative patients with syphilis, through the Dermatology Hospital of Southern Medical University in Guangzhou (recruited between August 2012 and March 2022, treated as Development cohort), the Beijing Youan Hospital of Capital Medical University in Beijng (recruited between December 2013 and April 2021, treated as External cohort 1), the Zhongshan Hospital of Xiamen University in Xiamen (recruited between November 2005 and November 2021, treated as External cohort 2) from China, and University of Washington School of Medicine in Seattle (recruited between September 2002 and April 2014, treated as External cohort 3) from United States. The deidentified patient data were obtained from four hospitals to form a large, multi-centre cohort. We included all these patients with syphilis into our analysis, and no patients were further excluded.

This study was a secondary analysis of the retrospective data. Our study is a retrospective, multicohort, observational study. Consent and Research Ethics Board approvals were not required for the use of deidentified data. Further details for diagnosis of syphilis are given in Supplementary Appendix S1.

### Statistical analysis and modelling

Age, CSF protein and CSF WBC were treated as continuous variables and described using medians with inter-quartile ranges (IQR) due to non-normality of values,

while the remaining variables were treated as categorical variables and described as counts and proportions.

We trained eXtreme gradient boosting (XGBoost)[43,44] models (Supplementary Appendix S1) to predict the diagnostic outcome of neurosyphilis using the full available features in Development cohort according to each diagnostic guideline in two scenarios, respectively. The XGBoost algorithm is a scalable decision tree-based boosting algorithm that is an ideal candidate for nonlinear, sparse, and class-imbalanced classification data.[44] The XGBoost models output a continuous prob-ability specifying the likelihood of classification to neu-rosyphilis for each patient, which was assessed at the class decision threshold of 0.50. Tenfold cross-validation was used to evaluate performance of each model, avoid any overfitting/underfitting, ensure robustness of models and minimize bias. Specifically, an inner tenfold cross-validation was applied to tune the hyper-parameters with a random gird search, set to maximize the area under the receiver operating characteristic curve (AUROC). The two steps of tenfold cross-validation constituted the double tenfold cross-validation in our study to minimize bias in perfor-mance evaluation (Fig. 1), which has been successfully applied in the design of another study.[45]

To allow for interpretation of our models' pre-dictions, we assessed feature importance, respectively, using the Shapley values[46–48] to identify a feature's rela-tive contribution to uncover key features. Based on the rankings of feature importance from the models in two scenarios, respectively, we selected a panel of consensus-based key features (i.e., a consensus reached by comprehensive considerations on the rankings of feature importance in all models developed from the diagnostic guidelines and ready availability and acces-sibility in real-world clinical practice) as a panel of key drivers for clinical diagnosis of neurosyphilis. We retrained our model using this subset of features (i.e., the panel of consensus-based key features), and arrived at simplified clinical diagnostic models through three external validation cohorts to identify patients with neurosyphilis. An online browser-accessible version of the final simplified clinical diagnostic models was also made available for external use.

We report the model performance in terms of discrimination and calibration. For the overall discrim-inatory ability of models, we reported standard diag-nostic accuracy estimates[49,50] (AUROC along with 95% confidence interval (CI), accuracy along with 95% CI, precision, recall and F1 measures) to evaluate perfor-mance of models in the presence of class imbalance.[51] Calibration was assessed graphically using a flexible calibration curve (Supplementary Appendix S1). To avoid instability we stratified the calibration curve in quintiles. In addition, to evaluate the calibration accu-racy, the overall calibration performance of models was also evaluated using Brier score.[52] We used the 2.5 and
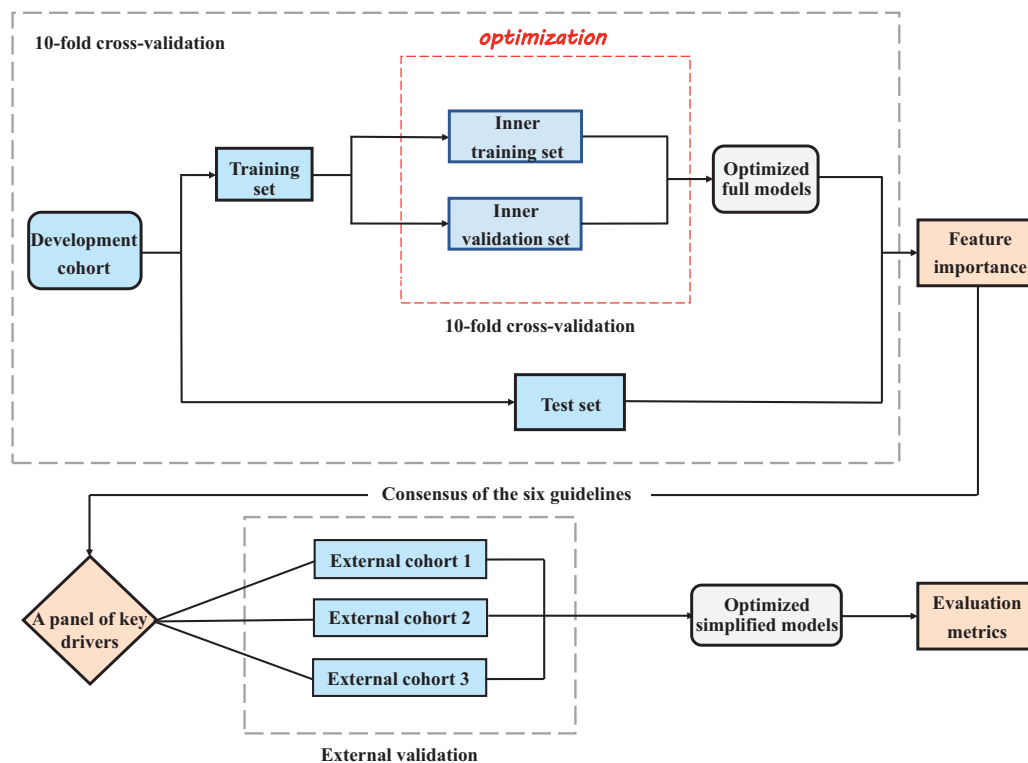
*Fig. 1:* **Schematic overview of the framework of this study.** Tenfold cross-validation was used to evaluate performance of each model, avoid any overfitting/underfitting, ensure robustness of models and minimize bias. Specifically, an inner tenfold cross-validation was applied to tune the hyperparameters with a random gird search, set to maximize the area under the receiver operating characteristic curve (AUROC). The two steps of tenfold cross-validation constituted the double tenfold cross-validation in our study to minimize bias in performance evaluation. We trained XGBoost models to predict the diagnostic outcome of neurosyphilis using the full available features in Development cohort (Table 1) according to each diagnostic guideline in two scenarios, respectively. To allow for interpretation of our models' predictions, we assessed feature importance, respectively, using the Shapley values to identify a feature's relative contribution to uncover key features. Based on the rankings of feature importance from the models in two scenarios, respectively, we selected a panel of consensus-based key features (i.e., a consensus reached by comprehensive considerations on the rankings of feature importance in all models developed from the diagnostic guidelines and ready availability and accessibility in real-world clinical practice) as a panel of key drivers for clinical diagnosis of neurosyphilis. We retrained our model using this subset of features (i.e., the panel of consensus-based key features), and arrived at simplified clinical diagnostic models through three external validation cohorts to identify patients with neurosyphilis. An online browser-accessible version of the final simplified clinical diagnostic models was also made available for external use (https://zhen-lu.shinyapps.io/Machine-learning-based-diagnosis-for-neurosyphilis/).

97.5 percentiles from 200 bootstrap samples as the limits of the 95% confidence intervals for the rescaled Brier score. A glossary of terms of statistics and machine learning used in this study could be found in the Supplementary Table S5, which defined all the metrics and provided the calculation formula for each.

### Decision curve analysis
In addition to evaluating the predictive performance in terms of discrimination and calibration for the machine learning models, we assessed the potential clinical utility of the models (i.e., the net benefit of the models) by using a decision curve analysis.[53] The net benefit of the models incorporating the trade-offs between true-positives and false-positives for a wide range of clinical probability thresholds is considered by the decision curve

analysis.[53,54] Thus, the decision curve analysis could consider the benefits and harms of using a model for clinical decision making, which allows decisions on management of patients with syphilis with variable neurosyphilis risk probabilities.[55] Further details for decision curve analysis are given in Supplementary Appendix S1.

All statistical analyses and modelling were performed using R software version 4.2.1 (R Core Team, Vienna, Austria, available at: https://www.R-project.org). We followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement.[55]

### Role of the funding source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing

## Results

### Characteristics of cohorts

Overall, we included 659, 480, 493 and 16 HIV-negative patients with syphilis in Development cohort, External cohort 1, 2, and 3 (Table 1). An overview of characteristics of all four cohorts and diagnostic outcomes by guideline are summarized in Table 1. In addition, the absolute numbers with percentages of patients with and without neurosyphilis in each cohort are also indicated in Table 1.

In Scenario 1, the transformed binary classifications of all four cohorts by six diagnostic guidelines are shown in Supplementary Figure S1. The class-imbalanced classifications (i.e., the number of neurosyphilis versus the number of patients without neurosyphilis) were much more common in Development cohort, External cohort 1 and 3 compared with External cohort 2. Supplementary Figure S2 shows transformed classifications of all four cohorts by six diagnostic guidelines in Scenario 2. The original classifications of all four cohorts by six diagnostic guidelines are shown in Table 1 and Supplementary Figure S3.

### Feature importance

Based on assessment of the rank order of variable importance (Fig. 2) in all models developed from six diagnostic guidelines and ready availability and accessibility in real-world clinical practice, the final panel of three key variables that distinguished patients with neurosyphilis from those without in Scenario 1 were neurological symptoms, CSF protein, and CSF WBC. For Scenario 2, neurological symptoms, CSF VDRL, and CSF protein were selected (Fig. 2). However, as shown in Table 1, in External cohort 1 and 2 only CSF RPR was available. Considering the lack of availability of CSF VDRL in resource-limited regions, CSF RPR has similar performance to CSF VDRL, and is easier to perform with readily available commercial test kits, we made a new variable named CSF VDRL/RPR, which was CSF VDRL or CSF RPR. Hence, the final panel of variables in Scenario 2 were neurological symptoms, CSF VDRL/RPR, and CSF protein. Detailed considerations of the final panel of features are shown in Supplementary Appendix S1.

### Performance of simplified models

In Scenario 1, overall predictive performance of the simplified models in the internal and external validation datasets are shown in Fig. 3. When internally validated, the simplified US 2021 model had the best value of AUROC (0.98, 95% CI 0.94–1.00), accuracy (0.96, 95% CI 0.88–0.99), precision (1.00), recall (0.79), and F1 measure (0.88), among simplified models. The simplified NT Australia 2022 model had the minimum value of AUROC

(0.94, 0.88–0.98). Furthermore, when externally validated, the simplified US CDC 2018 model had the best value of AUROC from 0.99 to 1.00, accuracy from 0.98 to 1.00, recall all of 1.00, and F1 measure from 0.91 to 1.00, among simplified models. All simplified models demonstrated excellent predictive performance when validated internally and externally. For calibration performance of six simplified models, the calibration curves (Supplementary Figure S5) and rescaled Brier scores (Supplementary Table S3) showed good agreement between predicted and observed risks overall of diagnostic performance, with the minimum value of 0.34 (0.22–0.79) [the simplified NT Australia 2022 model in External cohort 3], and the maximum value of 0.99 (0.98–1.00) [the simplified Europe 2020 model in External cohort 3].

Overall predictive performance of six simplified models in Scenario 2 are shown in Supplementary Figure S4. When internally validated, the simplified China 2020 model had the minimum value of AUROC (0.91, 0.81–0.98), accuracy (0.91, 0.81–0.97), recall (0.33), and F1 measure (0.50) among six simplified models. When externally validated, the simplified UpToDate 2020 model in External cohort 1 had the minimum value of AUROC (0.84, 0.81–0.88) among six simplified models in all three external validation datasets. All six simplified models illustrated excellent predictive performance and calibration (Supplementary Figure S6) when validated internally and externally.

### Clinical utility of models

For almost all six simplified models in both the internal and external validation datasets, in Scenario 1, using the models could yield net benefits for probability thresholds between 0.14 and 0.66 (Fig. 4, Supplementary Figures S7, S9, and S11), and it offered net benefits for probability thresholds between 0.33 and 0.66 (Fig. 4, Supplementary Figures S8, S10, and S12) in Scenario 2.

The overall decision curve analysis demonstrated the net benefits of all simplified models in identifying patients with syphilis who actually have neurosyphilis, compared with two reference strategies (i.e., manage all as if those patients with syphilis will or will not have a diagnostic outcome of neurosyphilis ['treat all' or 'treat none']), for probability thresholds between 0.33 and 0.60.

## Discussion

In this study, we used machine learning to develop and externally validate six simplified neurosyphilis diagnostic models based on six guidelines across four continents using routinely collected data from real-world clinical practice in China and the United States. All simplified clinical diagnostic models accurately predicted the diagnosis of neurosyphilis in demographically diverse HIV-negative patients with syphilis. They had good calibration and excellent discrimination over a broad range of clinical probability thresholds.

| Features | Development cohort (N = 659) | External cohort 1 (N = 480) | External cohort 2 (N = 493) | External cohort 3 (N = 16) |
|---|---|---|---|---|
| **Demographical information** | | | | |
| Age | 41 (29, 54) | 40 (30, 52) | 53 (44, 63) | NC |
| Sex | | | | |
| Female | 341 (51.70%) | 277 (57.70%) | 174 (35.30%) | NC |
| Male | 318 (48.30%) | 203 (42.30%) | 319 (64.70%) | NC |
| **Laboratory results** | | | | |
| Serum non-treponemal test[a] | | | | |
| Negative | 54 (8.19%) | 174 (36.25%) | 91 (18.46%) | 0 |
| Positive | 605 (91.81%) | 306 (63.75%) | 402 (81.54%) | 16 (100.00%) |
| CSF protein (g/L)[b] | 25.96 (17.20, 38.35) | 18.83 (13.42, 30.00) | 45.80 (31.84, 69.80) | 42.00 (36.75, 55.75) |
| CSF WBC (/mL)[b] | 2.00 (1.00, 4.00) | 5.00 (2.00, 9.00) | 6.00 (2.00, 21.00) | 3.00 (1.50, 24.80) |
| CSF treponemal test | | | | |
| CSF TPPA | | | | |
| Negative | 391 (59.33%) | 198 (41.25%) | 254 (51.52%) | 6 (37.50%) |
| Positive | 268 (40.67%) | 282 (58.75%) | 239 (48.48%) | 10 (62.50%) |
| CSF FTA-ABS-IgG | | | | |
| Negative | 425 (64.49%) | 221 (46.04%) | NA[d] | 8 (50.00%) |
| Positive | 234 (35.51%) | 259 (53.96%) | NA[d] | 8 (50.00%) |
| CSF non-treponemal test | | | | |
| CSF VDRL | | | | |
| Negative | 546 (82.85%) | NA[d] | NA[d] | 10 (62.50%) |
| Positive | 113 (17.15%) | NA[d] | NA[d] | 6 (37.50%) |
| CSF TRUST | | | | |
| Negative | 562 (85.28%) | NA[d] | NA[d] | NA[d] |
| Positive | 97 (14.72%) | NA[d] | NA[d] | NA[d] |
| CSF RPR | | | | |
| Negative | NA[d] | 407 (84.79%) | 346 (70.18%) | 13 (81.25%) |
| Positive | NA[d] | 73 (15.21%) | 147 (29.82%) | 3 (18.75%) |
| **Vital sign** | | | | |
| Neurological symptoms | | | | |
| Negative | 366 (55.54%) | 335 (69.79%) | 180 (36.51%) | 11 (68.75%) |
| Positive | 293 (44.46%) | 145 (30.21%) | 313 (63.49%) | 5 (31.25%) |
| **Observed diagnostic outcomes by guidelines[c]** | | | | |
| **China 2020** | | | | |
| Verified neurosyphilis | 94 (14.26%) | 64 (13.33%) | 168 (34.08%) | 3 (18.75%) |
| Probable neurosyphilis | 8 (1.21%) | 6 (1.25%) | 34 (6.90%) | 1 (6.25%) |
| Non-neurosyphilis | 557 (84.52%) | 410 (85.42%) | 291 (59.03%) | 12 (75.00%) |
| **Europe 2020** | | | | |
| Verified neurosyphilis | 136 (20.64%) | 93 (19.38%) | 239 (48.48%) | 4 (25.00%) |
| Non-neurosyphilis | 523 (79.36%) | 387 (80.62%) | 254 (51.52%) | 12 (75.00%) |
| **NT Australia 20,222** | | | | |
| Verified neurosyphilis | 92 (13.96%) | 41 (8.54%) | 141 (28.60%) | 2 (12.50%) |
| Probable neurosyphilis | 92 (13.96%) | 60 (12.50%) | 87 (17.65%) | 3 (18.75%) |
| Non-neurosyphilis | 475 (72.08%) | 379 (78.96%) | 265 (53.75%) | 11 (68.75%) |
| **UpToDate 2020** | | | | |
| Verified neurosyphilis | 376 (57.06%) | 279 (58.12%) | 413 (83.77%) | 9 (56.25%) |
| Non-neurosyphilis | 283 (42.94%) | 201 (41.88%) | 80 (16.23%) | 7 (43.75%) |
| **US CDC 2018** | | | | |
| Verified neurosyphilis | 92 (13.96%) | 41 (8.54%) | 141 (28.60%) | 2 (12.50%) |
| Probable neurosyphilis | 30 (4.55%) | 32 (6.67%) | 61 (12.37%) | 2 (12.50%) |
| Possible neurosyphilis | 140 (21.24%) | 33 (6.88%) | 61 (12.37%) | 1 (6.25%) |
| Non-neurosyphilis | 397 (60.24%) | 374 (77.92%) | 230 (46.65%) | 11 (68.75%) |

(Table 1 continues on next page)

| Features | Development cohort (N = 659) | External cohort 1 (N = 480) | External cohort 2 (N = 493) | External cohort 3 (N = 16) |
|---|---|---|---|---|
| (Continued from previous page) | | | | |
| **US 2021** | | | | |
| Verified neurosyphilis | 92 (13.96%) | 49 (10.21%) | 141 (28.60%) | 2 (12.50%) |
| Probable neurosyphilis | 44 (6.68%) | 44 (9.17%) | 98 (19.88%) | 2 (12.50%) |
| Non-neurosyphilis | 523 (79.36%) | 387 (80.62%) | 254 (51.52%) | 12 (75.00%) |

Abbreviations: WBC, White blood cell; CSF, Cerebrospinal fluid; TPPA, Treponema pallidum particle agglutination; FTA-ABS-IgG, Fluorescent treponemal antibody-absorbed immunoglobulin G; VDRL, Venereal disease research laboratory test; TRUST, Toluidine red unheated serum test; RPR, Rapid plasma reagin. [a]Serum non-treponemal test represents serum TRUST or serum RPR, instead of CSF TRUST or CSF RPR. [b]CSF protein and CSF WBC were treated as continuous variables and described using medians with interquartile ranges (IQR) due to non-normality of values, while the remaining variables were treated as categorical variables and described as counts and proportions. [c]The count of diagnostic outcomes by guidelines corresponds to that in Supplementary Figure S3. NC represents not complete data were available. [d]NA, no applicable data, which represents the whole patients in the specific cohort did not have the specific CSF variable in Table 1. For example, the whole patients (N = 659) in Development cohort did not have available values of CSF RPR, while they all had the values of CSF protein, CSF WBC, CSF TPPA, CSF FTA-ABS-IgG, CSF VDRL, and CSF TRUST. Other cohorts followed this suit.

*Table 1:* Candidate features to be incorporated into the modeling analysis and diagnostic outcomes by guidelines.
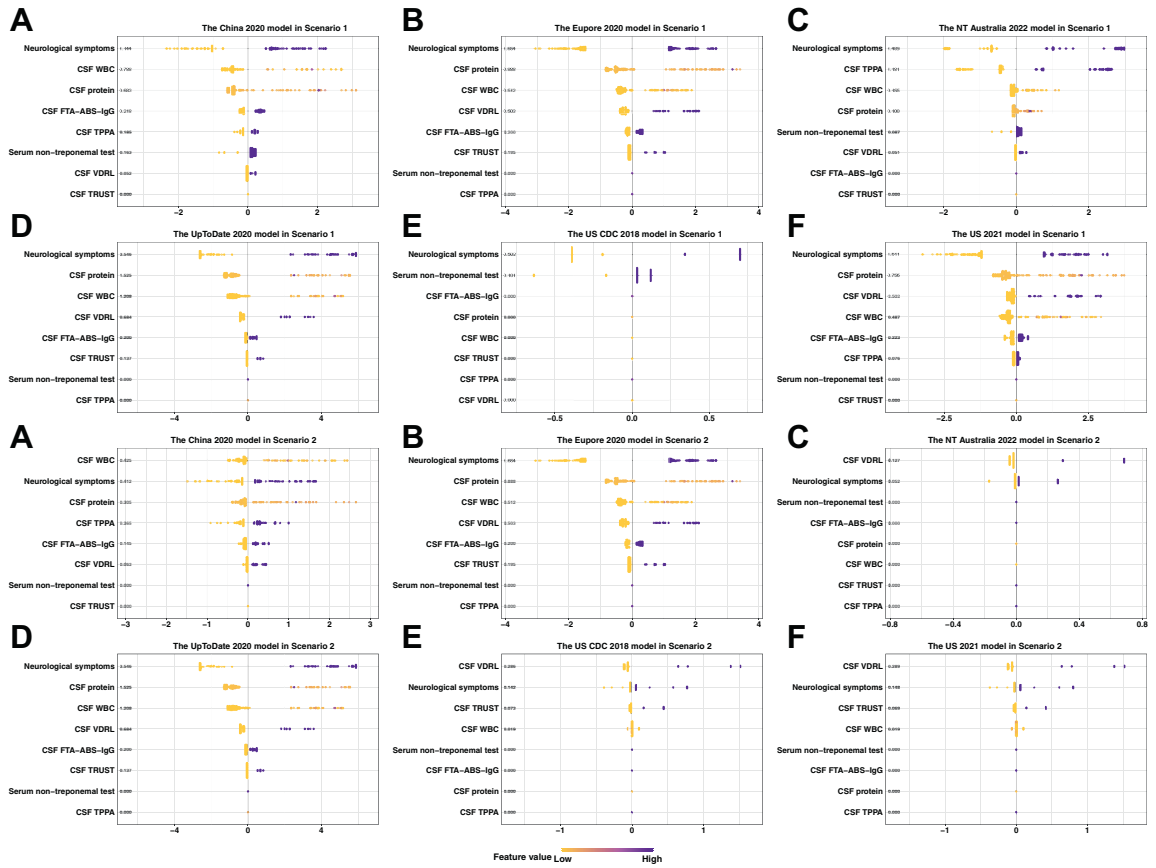


*Fig. 2:* **The Shapley values for each of candidate features in Scenario 1 and 2.** To allow for interpretation of our models' predictions, we assessed feature importance, respectively, using the Shapley values to identify a feature's relative contribution to uncover key features. Based on the rankings of feature importance from the models in two scenarios, respectively, we selected a panel of consensus-based key features (i.e., a consensus reached by comprehensive considerations on the rankings of feature importance in all models developed from the diagnostic guidelines and ready availability and accessibility in real-world clinical practice) as a panel of key drivers for clinical diagnosis of neurosyphilis. Each point in the plots is a Shapley (importance) value for a single patient. The color of each point represents the magnitude and direction of the value of that feature for that patient. A feature for a patient with a Shapley value below zero decreases the probability specifying the likelihood of being diagnosed as neurosyphilis, in addition, a higher probability indicates a higher likelihood of neurosyphilis. A in the figure indicates the results for the China 2020 model; B in the figure indicates the results for the Europe 2020 model; C in the figure indicates the results for the NT Australia 2022 model; D in the figure indicates the results for the UpToDate 2020 model; E in the figure indicates the results for the US CDC 2018 model; F in the figure indicates the results for the US 2021 model.
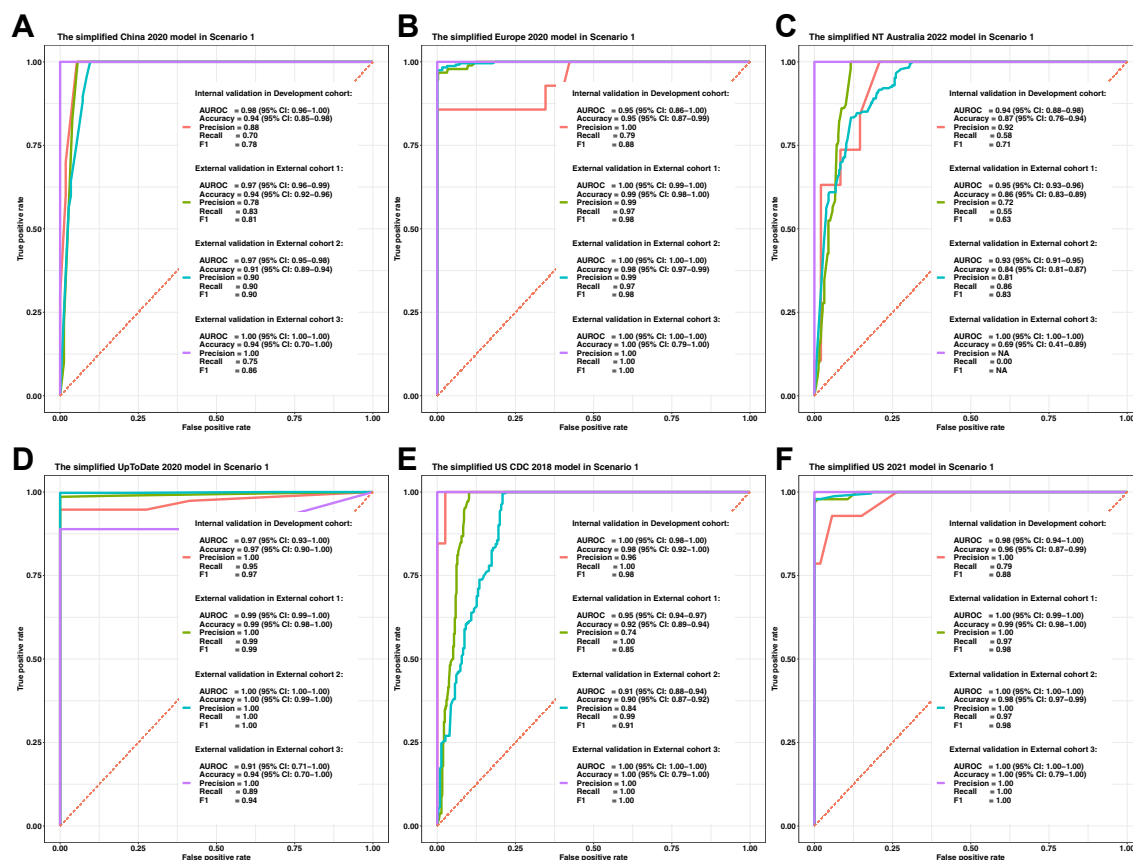
**Fig. 3: Predictive performance of six simplified models in the internal and external validation datasets in Scenario 1.** The lines with different colors in the plots represented the values of area under the receiver operating characteristic curve for each model. Several receiver operating characteristic curves overlapped completely due to the same values of area under the receiver operating characteristic curve. Predictive performance of six simplified models in the internal and external validation datasets in Scenario 2 were shown in Supplementary Figure S4. Abbreviations: AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval; NA, Not applicable. A glossary of terms of statistics and machine learning used in this study could be found in the Supplementary Table S5. A in the figure indicates the results for the China 2020 model; B in the figure indicates the results for the Europe 2020 model; C in the figure indicates the results for the NT Australia 2022 model; D in the figure indicates the results for the UpToDate 2020 model; E in the figure indicates the results for the US CDC 2018 model; F in the figure indicates the results for the US 2021 model.

To the best of our knowledge, this is the first study that applies machine learning techniques to aid in simplifying the diagnosis of neurosyphilis. As there is no universal gold clinical standard globally to diagnose neurosyphilis so far,[15] various distinct diagnostic guidelines have been endorsed and subsequently recommended as silver standard diagnostic criteria worldwide to answer such a medically important question with reasonable accuracy. The societies acknowledge that different diagnostic criteria exist and that the optimal diagnostic criteria may vary depending on the local condition, health resources, and preferences of clinicians and patients. However, current diagnostic guidelines are replete with various distinct essential prerequisites that may be not readily available across different clinical settings and population from different regions in the world, especially for those from resource-limited regions. The machine learning approach used in this study succeeded in identifying a panel of the fewest required key predictors to simplify the diagnosis of neurosyphilis.

We strongly support the principle of seeking to validate and, where possible, update existing models rather than developing new models de novo.[56] However, to date, no prediction modelling studies on diagnosis of neurosyphilis were identified. Previous studies of neurosyphilis risk primarily focused on significant predictors of neurosyphilis diagnosis using logistic regression mainly.[57–63] In the 2004 study of assessing CSF abnormalities, Christina M. Marra and colleagues used logistic regression to define clinical and laboratory features that identified patients with neurosyphilis.[58] Another study assessing CSF abnormalities in HIV-negative patients with neurosyphilis applied a multiple regression with a
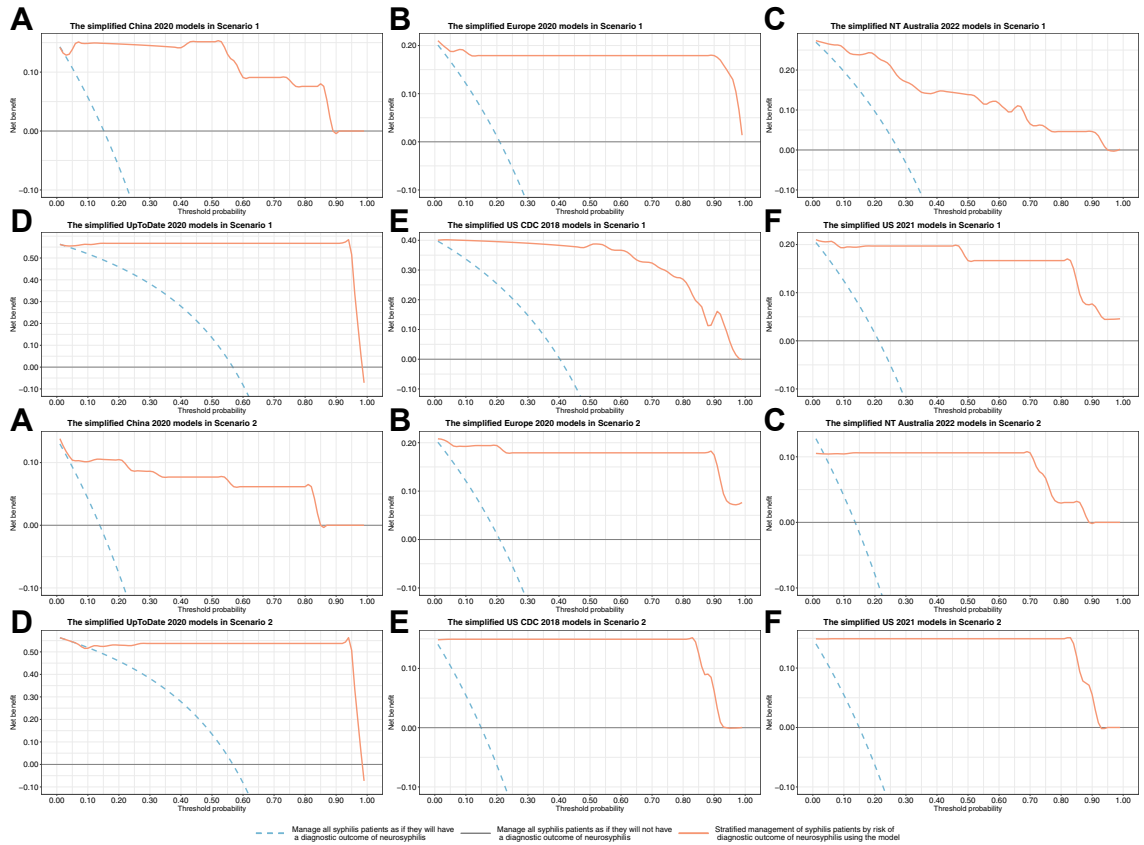
*Fig. 4:* **Decision curve analysis using the simplified models in the internal validation dataset from Development cohort in Scenario 1 and 2 to guide decisions at health-care systems and individual shared decision-making levels.** We assessed the potential clinical utility of the models (i.e., the net benefit of the models) by using a decision curve analysis. The net benefit of the models incorporating the trade-offs between true-positives and false-positives for a wide range of clinical probability thresholds is considered by the decision curve analysis. Thus, the decision curve analysis could consider the benefits and harms of using a model for clinical decision making, which allows decisions on management of patients with syphilis with variable neurosyphilis risk probabilities. A in the figure indicates the results for the China 2020 model; B in the figure indicates the results for the Europe 2020 model; C in the figure indicates the results for the NT Australia 2022 model; D in the figure indicates the results for the UpToDate 2020 model; E in the figure indicates the results for the US CDC 2018 model; F in the figure indicates the results for the US 2021 model.

backward elimination selection procedure to identify contribution of variables to the prediction of neuro-syphilis.[59] A study by Jeannot Dumaresq et al. on clinical prediction and diagnosis of neurosyphilis in HIV-positive patients with early syphilis used multivariable logistic regression to derive odds ratios and respective 95% CI as estimates of the relative risks of the putative predictors.[60] In an observational study in 2017, Yao Xiao et al. also implemented logistic regression model to identify novel predictors of neurosyphilis among HIV-negative patients with syphilis.[61] Similarly, in 2019, Yong Lu et al. conducted a case–control study using multivariable logistic regression to explore diagnostic indictors for the clinical diagnosis of neurosyphilis in HIV-negative patients.[62] In a latest single-centre, retrospective cohort study in 2022 in China, researchers used logistic regression to investigate and evaluate predictors of neurosyphilis among

patients with syphilis with different HIV status.[63] In these studies, the striking gaps between the evaluation of risk factors and the development and validation of diagnostic models were clear.

In our study, the overall performance of the simplified models based on discrimination and calibration was compelling, and this was reflected in the decision curve analyses which demonstrated clinical utility across a range of probabilities to support decision-making of health-care provision for various model generated risk probabilities.[55,64] In the previous seven relevant studies, those logistic regression models have not been evaluated formally regarding clinical utility, which might limit their clinical application. However, acceptability by clinicians and patients as well as shared decision making of clinical diagnostic models are essential prerequisites for translation into clinical application and the models

must align with real-world clinical understanding. Here, in our decision curve analysis, the clinical utility of the simplified models was explored in comparisons to two reference strategies, i.e., manage all as if those patients with syphilis will or will not have an a diagnostic outcome of neurosyphilis ('treat all' or 'treat none'). The overall decision curve analysis demonstrated the net benefits of all simplified models in identifying patients with syphilis who are likely to develop a diagnostic outcome of neurosyphilis, over a range of probability thresholds.

As health resources and clinical practice vary in different settings worldwide, we avoided recommending a universal optimal probability threshold for which patients with syphilis should undergo investigations for neurosyphilis. Instead, we reported the range of probability thresholds ranging from 0.33 to 0.66, showing that the models work well within this range of disease probability.[65] In our case, a lower threshold may be preferred to avoid missing potential cases of neurosyphilis, while a higher threshold may be preferred in situations where false positives could lead to unnecessary treatments or interventions.

To maximize generalizability and promote translation into real-world clinical practice, we developed the neurosyphilis diagnosis Shiny app (Supplementary Figure S13), available at: https://zhen-lu.shinyapps.io/Machine-learning-based-diagnosis-for-neurosyphilis/, allowing clinicians to calculate individualized risks of neurosyphilis according to six authoritative diagnostic guidelines. It is worth noting that, in real-world clinical practice, the models should be treated as a whole, rather than making a diagnosis prediction based on a single feature without considering the rest key features, which means the models need the inputs of the panel of features to run for the discrimination between non-neurosyphilis and neurosyphilis.

Strengths of our study include external validation in independent cohorts from three distinct settings (including External cohort 1 and 2 in China, and External cohort 3 in the United States), and a large number of participants with and without neurosyphilis. At the same time, we acknowledged the issue of not enough samples for External cohort 3 from Washington, due to the difficulty in collecting external data. Regarding the collection of participant-level data of HIV-negative patients with neurosyphilis, however, the diagnostic challenges associated with neurosyphilis in various countries or regions frequently result in a scarcity of reliable epidemiological data. In this study, in light of the real-word complexity of the diagnosis of neurosyphilis, we tried our best to find enough clinical collaborators outside of China to support data for conducting this research, but could not access much more data for our modeling analysis at that time. In this regard, the samples in External cohort 3 from Washington could not be representative of the population outside of China, but we do believe it might be indicative of possibly excellent generalizability of the models in population outside of China, which awaits further validation with much larger sample size of population outside of China in our future study. The generalizability of the models to worldwide population of patients with syphilis still awaits further investigation in the future study with much more data from various source in the world. And in this regard, the diagnostic predictions of the models should be interpreted and treated with cautious when applicable to population of US and countries other than China.

Our simplified models were well calibrated, and require only three variables. The negligible decrease in discrimination from the internal validation to external validation suggests negligible overfitting and provides confidence in the overall robustness of the simplified models. In addition, judging from the results of our modeling analysis in both scenarios, this indeed help improve our capability to simplify the complexity of real-word clinical practice of neurosyphilis diagnosis, but also considering differential risk indications by two possible scenarios to be much more adaptive to the clinical practice worldwide. Hence, the excellent performance of the models in both scenarios supports the robust and unbiased transformation method we applied. Furthermore, regarding the selection and review of the guidelines, within our capabilities, the clinical experts we selected for our study possess significant clinical experience in diagnosing neurosyphilis, and also published high-impact original articles about the neurosyphilis. The clinical experts are from China and US, which could not be representative of the four continents, however, we do believe they represent a high level of expertise on this topic globally. Considering the consensus of five experts from both China and the US (CMM, TCY, LY, BY and WK) and systematic review of the related publications of diagnosis of neurosyphilis, in addition to several statisticians (ZL, HZ, JW, YL), we have sufficient reasons and confidence to believe that the process of selection and review of the guidelines and incorporated features was unbiased and robust. And we followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement,[55] which could increase the confidence in the process of selection and review of the guidelines and features.

Limitations of this study are worth acknowledging. We acknowledged that we initially overlooked well-known databases such as Cochrane reviews and Medline when conducting our literature search for neurosyphilis diagnostic guidelines. However, we have since performed an additional search specifically in these databases and found no additional relevant articles beyond what was already included in our review (Supplementary Table S4), which means the search of literatures for available guidelines may be

comprehensive till now. Although our study involved totally 1648 confirmed patients with syphilis from four settings in China and the United States, the size of population in External cohort 3 was small. Thus, our findings are most generalizable to individuals with syphilis in China. We considered CSF VDRL and CSF RPR as equivalent for the final panel of variables in Scenario 2. The use of CSF RPR as a surrogate for CSF VDRL in the simplified models need further research and would benefit from complete data of the two variables if available.

We did not model neurosyphilis diagnosis in people living with HIV. Actually, constructing neurosyphilis-diagnostic models that could be applicable to the entire population of patients with syphilis (including HIV-negative and HIV-positive patients) was our original intention. However, the diagnostic challenges associated with neurosyphilis in various countries or regions frequently result in a scarcity of reliable epidemiological data, and the true burden of neurosyphilis worldwide is likely underestimated due to underreporting and lack of recognition.[14] In light of the real-word complexity of the diagnosis of neurosyphilis, we tried our best but could not access enough data from HIV-positive patients with neurosyphilis for our modeling analysis. In this regard, we had to turn to developing models for neurosyphilis diagnosis in HIV-negative patients with syphilis, which resulted in that the models are not adapted to HIV-positive patients. Finally, our models could not be used to differentiate types of neurosyphilis (including early or late neurosyphilis). The logic of our modeling analysis follows the chronological order of clinical diagnosis.[56] Subdivided diagnostic outcomes of different types of neurosyphilis may compromise the diagnostic performance of the models. In this study, we did not categorize the diagnostic outcome of diagnosed neurosyphilis into too detailed classifications, which is one of the limitations of our study and awaits future modeling study to target diagnostic outcomes of different types and/or stages of neurosyphilis. In this regard, the models of this study could not be applicable to discrimination among different types and/or stages of neurosyphilis.

In conclusion, we developed simplified, validated, diagnostic models that accurately diagnosed neurosyphilis in HIV-negative patients with syphilis. Both internal and external validation demonstrated the models are transportable across clinical settings. Stratifying patients with syphilis for risk-differentiated decision on the health-care by using the models offers net benefits over a broad range of probability thresholds. Future studies should validate our models in PLWH and individuals with syphilis outside China.

## Contributors
Research idea, study design, and writing: WK, HZ, ZL, WW and LY; data acquisition: LY, XL, LF, YL, TY and JW; data analysis or interpretation: ZL, HZ, WK, XZ and LW; statistical analysis: ZL, XW, XZ, TT, YL and LH; supervision or mentorship: LY, CMM, TY and BY. All authors

have directly accessed and verified the underlying data in this study, and were responsible for the decision to submit the manuscript.

## References
1 Ropper AH. Neurosyphilis. *N Engl J Med*. 2019;381(14):1358–1363.
2 Lukehart SA. Biology of treponemes. *Sex Transm Dis*. 2008;4:647–659.
3 Lukehart SA, Hook EW 3rd, Baker-Zander SA, Collier AC, Critchlow CW, Handsfield HH. Invasion of the central nervous system by Treponema pallidum: implications for diagnosis and treatment. *Ann Intern Med*. 1988;109(11):855–862.
4 Ghanem KG, Ram S, Rice PA. The modern epidemic of syphilis. *N Engl J Med*. 2020;382(9):845–854.
5 Gonzalez H, Koralnik IJ, Marra CM. *Neurosyphilis. Seminars in neurology; 2019*. Thieme Medical Publishers; 2019:448–455.
6 Merritt HH, Adams RO, Solomon HC. *Neurosyphilis*. Neurosyphilis. 1946:443.
7 Clark EG, Danbolt N. The Oslo study of the natural history of untreated syphilis: an epidemiologic investigation based on a restudy of the Boeck-Bruusgaard material a review and appraisal. *J Chronic Dis*. 1955;2(3):311–344.
8 de Voux A, Kidd S, Torrone EA. Reported cases of neurosyphilis among early syphilis cases—United States, 2009 to 2015. *Sex Transm Dis*. 2018;45(1):39.
9 Ouwens IMD, Koedijk FD, Fiolet AT, et al. Neurosyphilis in the mixed urban–rural community of the Netherlands. *Acta Neuropsychiatr*. 2014;26(3):186–192.
10 Lester R, Morshed M, Gilbert M, Hottes T. Syphilis and neurosyphilis increase to historic levels in BC. *BCMJ*. 2013;55(4):204.
11 Tang W, Huang S, Chen L, et al. Late neurosyphilis and tertiary syphilis in Guangdong Province, China: results from a cross-sectional study. *Sci Rep*. 2017;7(1):45339.
12 Pinchera B, Vicecone G, Buonomo AR, et al. Epidemiological and clinical features of syphilis in the 21st century: a seven-year observational retrospective study of outpatients. *Clin Epidemiol Glob Health*. 2022;16:101100.
13 Dombrowski JC, Pedersen R, Marra CM, Kerani RP, Golden MR. Prevalence estimates of complicated syphilis. *Sex Transm Dis*. 2015;42(12):702–704.
14 Jackson DA, McDonald R, Quilter LAS, Weinstock H, Torrone EA. Reported neurologic, ocular, and otic manifestations among syphilis cases—16 States, 2019. *Sex Transm Dis*. 2022;49(10):726–732.
15 Tuddenham S, Ghanem KG. Neurosyphilis: knowledge gaps and controversies. *Sex Transm Dis*. 2018;45(3):147.

16 US CDC. *Case definitions in effect during*; 2019. https://www.cdc.gov/std/statistics/2019/case-definitions.htm.

17 Workowski KA, Bachmann LH, Chan PA, et al. Sexually transmitted infections treatment guidelines, 2021. *MMWR Recomm Rep.* 2021;70(4):1.

18 National Center for STD Control, Chinese Center for Disease Control and Prevention, Venereology Group, Chinese Society of Dermatology, Subcommittee on Venereology, China Dermatologist Association. Guidelines for diagnosis and treatment of syphilis, gonorrhea and genital Chlamydia trachomatis infection (2020). *Chin J Dermatol.* 2020;53(3):168–179.

19 UpToDate. Neurosyphilis. https://www.uptodate.com/contents/neurosyphilis.

20 Kingston M, French P, Higgins S, et al. UK national guidelines on the management of syphilis 2015. *Int J STD AIDS.* 2016;27(6):421–446.

21 Birrell JM, Lowe M, Gunathilake M, Krause VL. Neurosyphilis in the Northern Territory of Australia: a clinical guideline. *Intern Med J.* 2023;53(5):738–744.

22 Ramachandran PS, Baird RW, Markey P, et al. Neurosyphilis: still prevalent and overlooked in an at risk population. *PLoS One.* 2020;15(10):e0238617.

23 Janier M, Unemo M, Dupin N, Tiplica GS, Potocnik M, Patel R. 2020 European guideline on the management of syphilis. *J Eur Acad Dermatol Venereol.* 2021;35(3):574–588.

24 Du FZ, Wang QQ, Zheng ZJ, et al. The challenge of diagnosis and treatment of neurosyphilis in China: results from a nationwide survey. *Sex Health.* 2021;18(4):333–339.

25 UpToDate. *Neurosyphilis*; 2013. https://www.uptodate.com/contents/neurosyphilis.

26 Guidelines for diagnosis and treatment of syphilis, gonorrhea and genital Chlamydia trachomatis infection (2014). *Chin J Dermatol.* 2014;47(5):365–372.

27 French P, Gomberg M, Janier M, Schmidt B, van Voorst Vader P, Young H. IUSTI: 2008 European guidelines on the management of syphilis. *Int J STD AIDS.* 2009;20(5):300–309.

28 Workowski KA, Bolan GA. Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm Rep.* 2015;64(3):1–137.

29 US CDC. *Sexually transmitted disease surveillance*; 2017. https://www.cdc.gov/std/stats.

30 US CDC. *Sexually transmitted disease surveillance*; 2011. http://www.cdc.gov/std/stats.

31 Workowski KA, Berman SM. *Sexually transmitted diseases treatment guidelines, 2010.* 2010.

32 Workowski KA, Berman SM. *Sexually transmitted diseases treatment guidelines, 2006.* 2006.

33 Organization WH. *WHO guidelines for the treatment of Treponema pallidum (syphilis).* 2016.

34 Organization WH. *Guidelines for the management of sexually transmitted infections.* World Health Organization; 2003.

35 Jespers V, Stordeur S, Carville S, et al. Diagnosis and treatment of syphilis: 2019 Belgian National guideline for primary care. *Acta Clin Belg.* 2022;77(1):195–203.

36 Wong T, Fonseca K, Chernesky MA, Garceau R, Levett PN, Serhir B. Canadian Public Health Laboratory Network laboratory guidelines for the diagnosis of neurosyphilis in Canada. *Can J Infect Dis Med Microbiol.* 2015;26(Supplement A):18A–22A.

37 Gaspar PC, Bigolin Á, Alonso Neto JB, Pereira EDDS, Bazzo ML. Brazilian Protocol for sexually transmitted infections 2020: syphilis diagnostic tests. *Rev Soc Bras Med Trop.* 2021;54:e2020630.

38 Freitas FLS, Benzaken AS, Passos MRL, Coelho ICB, Miranda AE. Brazilian protocol for sexually transmitted infections 2020: acquired syphilis. *Rev Soc Bras Med Trop.* 2021;54:e2020616.

39 Janier á, Hegyi V, Dupin N, et al. 2014 European guideline on the management of syphilis. *J Eur Acad Dermatol Venereol.* 2014;28(12):1581–1593.

40 Kingston M, French P, Goh B, et al. UK national guidelines on the management of syphilis 2008. *Int J STD AIDS.* 2008;19(11):729–740.

41 US CDC. *Sexually transmitted disease surveillance*; 2019. https://wwwn.cdc.gov/nndss/case-de!nitions.html.

42 US CDC. *Sexually transmitted disease surveillance*; 2014. https://www.cdc.gov/std/stats/archive/surv-2014-print.PDF.

43 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016.* 2016:785–794.

44 Farooq Z, Rocklov J, Wallin J, et al. Artificial intelligence to predict West Nile virus outbreaks with eco-climatic drivers. *Lancet Reg Health Eur.* 2022;17:100370.

45 Ma B, Yan G, Chai B, Hou X. XGBLC: an improved survival prediction model based on XGBoost. *Bioinformatics.* 2021;38(2):410–418.

46 Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30.

47 Roth AE. *The Shapley value: essays in honor of Lloyd S. Shapley.* Cambridge University Press; 1988.

48 Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2(10):749–760.

49 Simundic AM. Measures of diagnostic accuracy: basic Definitions. *EJIFCC.* 2009;19(4):203–211.

50 Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res.* 2016;18(12):e323.

51 Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol.* 2019;25(6):485–495.

52 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology.* 2010;21(1):128.

53 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res.* 2019;3:18.

54 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565–574.

55 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55–63.

56 Cooray SD, Boyle JA, Soldatos G, et al. Development, validation and clinical utility of a risk prediction model for adverse pregnancy outcomes in women with gestational diabetes: the PeRSonal GDM model. *eClinicalMedicine.* 2022:101637.

57 MacLean S, Luger A. Finding neurosyphilis without the venereal disease research laboratory test. *Sex Transm Dis.* 1996;23(5):392–394.

58 Marra CM, Maxwell CL, Smith SL, et al. Cerebrospinal fluid abnormalities in patients with syphilis: association with clinical and laboratory features. *J Infect Dis.* 2004;189(3):369–376.

59 Liu LL, Zhang HL, Huang SJ, et al. Assessing cerebrospinal fluid abnormalities in neurosyphilis patients without human immunodeficiency virus infection. *Int Immunopharmacol.* 2013;17(4):1120–1124.

60 Dumaresq J, Langevin S, Gagnon S, et al. Clinical prediction and diagnosis of neurosyphilis in HIV-infected patients with early Syphilis. *J Clin Microbiol.* 2013;51(12):4060–4066.

61 Xiao Y, Tong ML, Liu LL, et al. Novel predictors of neurosyphilis among HIV-negative syphilis patients with neurological symptoms: an observational study. *BMC Infect Dis.* 2017;17(1):310.

62 Lu Y, Ke W, Yang L, et al. Clinical prediction and diagnosis of neurosyphilis in HIV-negative patients: a case-control study. *BMC Infect Dis.* 2019;19(1):1017.

63 Yan J, Luo L, Han J, et al. Comparing noninvasive predictors of neurosyphilis among syphilis patients with and without HIV Co-infection based on the real-world diagnostic criteria: a single-center, retrospective cohort study in China. *AIDS Res Hum Retroviruses.* 2022;38(5):406–414.

64 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;102(3):148–158.

65 Wynants L, Van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B. Three myths about risk thresholds for prediction models. *BMC Med.* 2019;17(1):1–7.