

## Review

# Prognostic Prediction Models for Ulcerative Colitis: Systematic Review and Meta-Analysis

Zhijun Bu<sup>1\*</sup>, PhD; Yuan Sun<sup>1\*</sup>, PhD; Zeyang Shi<sup>1\*</sup>, MS; Liming Hu<sup>2</sup>, Prof.; Chuanlan Ju<sup>3</sup>, Prof.; Yuan Liu<sup>4</sup>, Prof.; Jing An<sup>4</sup>, Prof.; Huiyi Sun<sup>5</sup>, Prof.; Jianping Liu<sup>1</sup>, Prof.; Zhaolan Liu<sup>1</sup>, Prof.

<sup>1</sup>School of Traditional Chinese Medicine, Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

<sup>2</sup>Department of Spleen, Stomach, Liver & Gallbladder Diseases, Dongfang Hospital, Beijing University of Chinese Medicine, Beijing, China

<sup>3</sup>Department of Public Health, Yantai Municipal Hospital of Traditional Chinese Medicine, Shandong, China

<sup>4</sup>Department of Spleen and Stomach Diseases, Third Affiliated Hospital, Beijing University of Chinese Medicine, Beijing, China

<sup>5</sup>Department of Spleen and Stomach Diseases, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China

\*these authors contributed equally

## Corresponding Author:

Zhaolan Liu, Prof.

School of Traditional Chinese Medicine

Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine

11 Beisanhuan East Road, Chaoyang District

Beijing 100029

China

Phone: 86 17800176627

Email: [17800176627@163.com](mailto:17800176627@163.com), [lzl1019@163.com](mailto:lzl1019@163.com)

## Abstract

**Background:** Ulcerative colitis (UC) is a chronic inflammatory disease with highly variable symptoms and severity. Prognostic models for UC support precision medicine by enabling personalized treatment strategies. However, the quality and clinical utility of these models remain inadequately assessed.

**Objective:** This study aimed to systematically review and critically evaluate the development, performance, and applicability of prognostic prediction models for UC.

**Methods:** To identify prognostic models for UC, a comprehensive search was conducted in PubMed, Embase, the Cochrane Library, Web of Science, SinoMed, China National Knowledge Infrastructure, Wanfang, and VIP Database up to November 2, 2024. Extracted data included study characteristics, model development methods, validation metrics (eg, area under the curve and concordance index). The risk of bias and applicability were evaluated using the Prediction Model Risk of Bias Assessment Tool. A meta-analysis was conducted to assess model performance.

**Results:** A total of 30 studies involving 7452 patients with UC were included, with the largest numbers conducted in China (11/30, 37%) and Japan (4/30, 13%). Most studies were retrospective (22/30, 73%). The primary objectives of the UC prognostic models included predicting therapeutic effects and responses to treatment, particularly to tumor necrosis factor-alpha inhibitors (eg, infliximab and adalimumab), and assessing the risks of surgery, disease progression, or relapse. Logistic regression was the most frequently used method for both predictor selection (6/30, 20%) and model construction (12/30, 40%). Common predictors included age, C-reactive protein, albumin, hemoglobin, disease extent, and Mayo scores. The meta-analysis yielded a pooled area under the curve of 0.84 (95% CI 0.77-0.92). Most studies exhibited a high risk of bias (29/30, 97%), particularly in participant selection and statistical analysis. Applicability concerns were identified in 18 studies (18/30, 60%), primarily due to subgroup-specific designs that limited the generalizability of the findings. External validation data (14/30, 47%) were limited, and only a small number of studies (12/30, 40%) included calibration curves or decision curve analysis.

**Conclusions:** This study demonstrates that prognostic models for UC have some potential in predictive performance and clinical application. However, most models are constrained by high bias risk, insufficient external validation, and limited generalizability due to small sample sizes and subgroup-specific designs. Future research should prioritize multicenter validations, refine model development approaches, and enhance model applicability to support broader clinical implementation.

**Trial Registration:** PROSPERO CRD42024609424; <https://www.crd.york.ac.uk/PROSPERO/view/CRD42024609424>

*J Med Internet Res* 2025;27:e71944; doi: [10.2196/71944](https://doi.org/10.2196/71944)

**Keywords:** ulcerative colitis; prognostic prediction models; Prediction Model Risk of Bias Assessment Tool; PROBAST; logistic regression; systematic review and meta-analysis

## Introduction

Ulcerative colitis (UC) is a chronic inflammatory bowel disease involving the rectum and colon, characterized by a relapsing-remitting course and affecting individuals across all age groups [1,2]. While UC incidence has stabilized in early industrialized regions, it is rising rapidly in newly industrialized countries, especially in Asia and Latin America [3]. A recent global analysis of over 500 population-based studies proposed a 4-stage model of inflammatory bowel disease evolution and classified many countries, including China, Malaysia, and Brazil, as being in the *accelerating incidence* stage, with steadily increasing UC burden [4]. In China, UC has transitioned from being a rare condition to a prevalent one, accounting for up to one-quarter of inpatient beds in gastroenterology and colorectal surgery departments [5]. The underlying mechanisms of UC are complex and involve a combination of genetic susceptibility, epithelial barrier impairments, immune system disturbances, and environmental triggers [6]. The disease impairs quality of life because of its relapsing nature and treatment burden, affecting mental health and social functioning [7].

UC exhibits substantial heterogeneity in symptoms and the severity of inflammation among patients. Consequently, relying solely on symptom-based approaches to guide treatment often leads to suboptimal management [8]. Precision medicine focuses on creating customized treatment strategies based on the unique characteristics of each patient and the progression of their disease, playing a crucial role in enhancing therapeutic effectiveness. By identifying each patient's specific response to treatment, unnecessary medication use can be avoided, side effects minimized, and treatment efficacy optimized [9,10]. Prognostic prediction models are central to precision medicine because they allow clinicians to predict disease progression, treatment response, and survival, thereby enabling more individualized treatment strategies [11]. However, despite the growing number of UC prognostic models, comprehensive evaluations of their

methodological quality, external validity, and real-world clinical utility are still lacking.

This study aimed to systematically review and summarize the existing literature on UC prognostic prediction models; comprehensively assess their performance, external validation, and clinical utility; identify existing research gaps; and provide evidence to guide the future development of UC prognostic prediction models.

## Methods

### Overview

This study was registered in the International Prospective Register of Systematic Reviews under the registration number CRD42024609424. It adheres to the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [12], and the full 2020 PRISMA checklist is provided in [Checklist 1](#).

### Search Strategy

A systematic search was conducted across 8 databases: PubMed, Embase, Cochrane Library, Web of Science, SinoMed, China National Knowledge Infrastructure, Wanfang Database, and VIP Database. The search covered the inception of each database up to November 2, 2024. Boolean operators (AND and OR) were used to combine Medical Subject Headings and free-text keywords, and detailed search strategies for all databases are provided in [Multimedia Appendix 1](#).

### Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were systematically structured based on the population, intervention, comparator, outcome, and study design (PICOS) framework, with additional relevant domains (eg, language) incorporated, as summarized in [Textbox 1](#).

**Textbox 1.** Inclusion and exclusion criteria categorized by eligibility domain, based on the population, intervention, comparator, outcome, and study design (PICOS) framework with additional language criteria.

#### Inclusion criteria

- Population: adult patients diagnosed with ulcerative colitis (UC)
- Intervention: development and validation of prognostic prediction models
- Comparator: not applicable
- Outcome: clinical outcomes, such as relapse, colectomy, hospitalization, steroid dependence, and so on
- Study design: original studies using statistical or machine learning models for prognosis
- Language: English or Chinese

**Exclusion criteria**

- Population: studies involving non-UC populations (eg, Crohn disease and inflammatory bowel disease without UC-specific data)
- Intervention: studies focusing solely on diagnostic models or without a clear time lag between predictors and outcomes
- Comparator: not applicable
- Outcome: studies not reporting prognostic outcomes or focusing only on diagnostic classification
- Study design: reviews, systematic reviews, case reports, editorials, or studies without predictive model development or validation
- Language: languages other than English or Chinese

**Study Selection and Screening Process**

Initially, 1 researcher (ZB) imported all retrieved records into EndNote X9 (Clarivate) software for deduplication. The deduplicated records were independently reviewed by 2 researchers (YS and ZS) through titles, abstracts, and keywords to identify studies that met the inclusion criteria. The same researchers reviewed the full texts of potentially eligible studies to finalize inclusion based on the criteria. At each stage, another researcher (LH) cross-checked the results to ensure consistency. Discrepancies were resolved through discussions involving all 3 researchers until a consensus was reached.

**Data Extraction**

Data from the included studies were independently extracted by 2 researchers (YS and ZS) and documented in a Microsoft Excel spreadsheet. Data extraction and organization were performed using Microsoft Excel (version 2024). Extracted data included basic study information such as the title, first author, publication year, study region, and so on.

1. Study background: specific subgroups of patients with UC (eg, mild, moderate, or severe cases)
2. Study population: sample sizes of training, validation, and test datasets; single-center or multicenter studies; and so on
3. Prediction model development: number and types of predictor variables, methods for variable selection, modeling approaches, data partitioning methods, and use of cross-validation or bootstrap validation
4. Model performance: metrics for internal and external validation, such as the area under the curve (AUC) and concordance index (C-index)
5. Applicability and limitations: clinical application scenarios and study limitations

Another researcher (CJ) cross-checked the extracted data with the original reviewers. Disagreements were addressed through discussion among the 3 researchers until mutual agreement was achieved.

**Quality Assessment of Included Studies**

Two researchers (YL and JA) independently evaluated the quality of the included studies using the Prediction Model

Risk of Bias Assessment Tool (PROBAST). This tool assesses bias risk and applicability in prediction model studies across 4 domains: participants, predictors, outcomes, and analysis. It also includes 20 signaling questions designed to pinpoint potential biases in study design, data analysis, and reporting. Additionally, PROBAST assesses the applicability of models within specific clinical contexts. Each study's risk of bias was categorized as "high risk," "low risk," or "unclear risk" [13]. Differences were addressed through discussion involving a third researcher (HS).

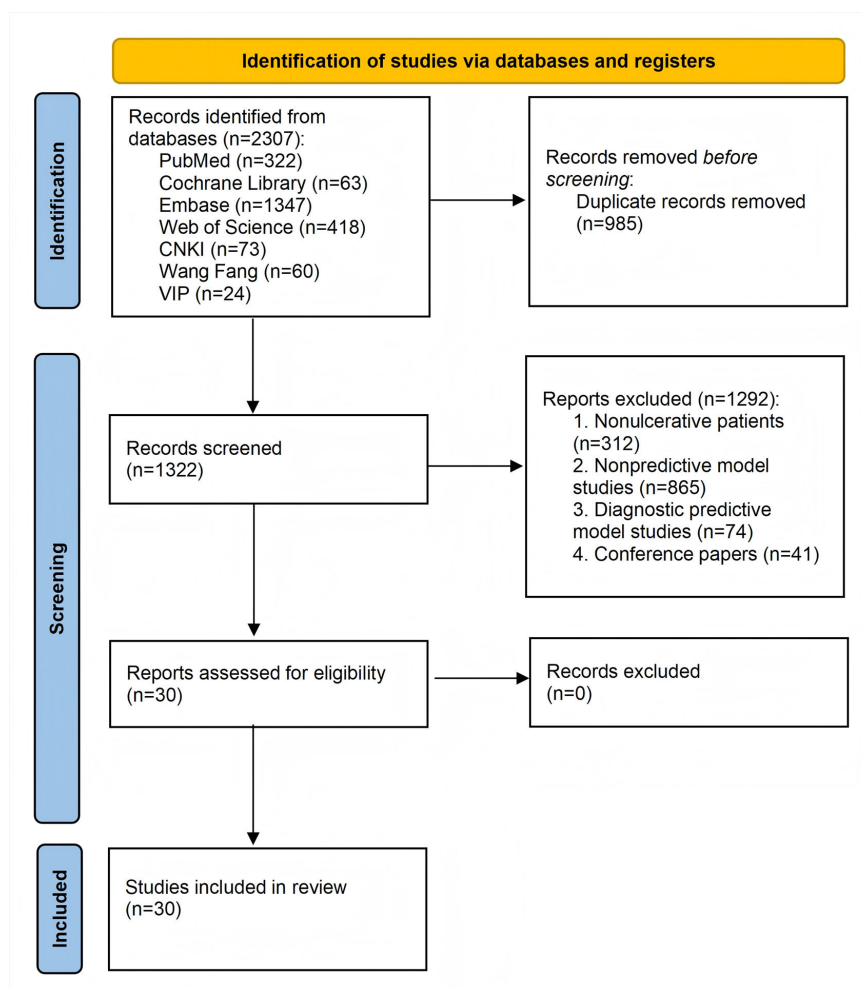
**Data Synthesis and Statistical Analysis**

The synthesis of data and statistical analysis were conducted using Stata 17 (StataCorp LLC) software. Meta-analyses of performance metrics were conducted using the meta package and the metagen function, with subgroup analyses comparing internal and external validation datasets. A random effects model was used to estimate pooled effect sizes with 95% CI, and heterogeneity was evaluated using conventional metrics. Funnel plots and the Egger test were used to assess potential publication bias. Sensitivity analyses were performed by excluding individual studies one at a time to evaluate the stability of the results and the impact of each study on the pooled effect sizes.

**Results****Study Selection**

A total of 2307 studies were identified from the databases. After removing 985 duplicates (985/2307, 42.70%), 1322 studies (1322/2307, 57.30%) remained for screening. During the preliminary screening, we excluded 312 studies (312/1322, 23.6%) that did not include patients with UC, 865 studies (865/1322, 65.4%) unrelated to prediction models, 74 studies (74/1322, 5.6%) focusing on diagnostic prediction models, and 41 (41/1322, 3.1%) studies focusing on conference abstracts. Consequently, 30 studies (30/1322, 2.3%) were chosen for a full-text review. After independent verification and discussion, all 30 studies were ultimately included in the analysis (Figure 1).

**Figure 1.** Study selection flow diagram based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines for a systematic review and meta-analysis of prognostic prediction models in ulcerative colitis (UC). This figure illustrates the number of records identified, screened, excluded (with reasons), and included in the final analysis. The systematic review included studies that developed or validated prognostic models for patients with UC, published in either English or Chinese, from the inception of each database up to November 2, 2024, across various countries. The PRISMA 2020 framework was used to ensure transparency and reproducibility in the study selection process.



## Study Characteristics

The studies included in the analysis were published between 2015 and 2024, involving a total of 7452 patients with UC. Among these, 12 studies (12/30, 40%) focused on patients with moderate-to-severe UC, and 6 (6/30, 20%) exclusively targeted patients with severe UC. Geographically, the largest number of studies was conducted in China (11/30, 37%), followed by Japan (4/30, 13%). With respect to study design, 6 studies (6/30, 20%) were prospective, 22 (22/30, 73%) were retrospective, and 2 (2/30, 7%) used a mixed prospective-retrospective approach. Ten were single-center studies (10/30, 33%), whereas 20 involved multiple centers (20/30, 67%).

Key objectives of the included UC prognostic models included predicting treatment efficacy and response, assessing surgery-related risks, forecasting disease progression or relapse, and exploring molecular biomarkers or disease mechanisms. Treatment efficacy and response prediction was the most common objective (18/30, 60%), focusing on interventions such as tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) inhibitors (eg, infliximab and adalimumab), vedolizumab, ustekinumab, tofacitinib, fecal microbiota transplantation, leukocyte apheresis, and the Chinese herbal medicine

Wuwei Kushen Enteric-coated capsules. Surgery-related risk assessments (5/30, 17%) primarily aimed to predict pouchitis, postoperative complications, and long-term surgical outcomes. Disease progression and relapse prediction (6/30, 20%) focused on risks of acute severe UC, relapse probabilities, and remission likelihoods ([Multimedia Appendix 2](#)).

Most studies (19/30, 63%) did not specify how missing data were handled. Among the remaining studies, the most common approach was direct deletion (6/30, 20%), whereas others used random forest (1/30, 3%), k-nearest neighbors (1/30, 3%), mean imputation (2/30, 7%), and multiple imputations (1/30, 3%). Data splitting into training and validation sets was the most common method, with a 7:3 split used in 6 studies. Predictor selection methods included regression techniques (13/30, 43%), such as logistic regression (6/30, 20%), LASSO regression (4/30, 13%), Cox regression (1/30, 3%), and elastic net regularization regression (1/30, 3%). Machine learning methods were used in 6 studies (6/30, 20%), comprising random forest (5/30, 17%) and CatBoost (1/30, 3%).

Common predictors across studies were age, C-reactive protein (CRP), sex, albumin, erythrocyte sedimentation

rate, platelet count, hemoglobin, disease extent or severity, endoscopic findings, and Mayo scores. Logistic regression was the most frequently used modeling method (12/30, 40%), followed by neural networks (8/30, 27%) and random forest (3/30, 10%). Cross-validation was used in 17 studies (17/30, 57%), with 5-fold cross-validation being the most common. Bootstrap validation was performed in 9 studies (9/30, 30%). 12 studies (12/30, 40%) provided calibration curves, 3 (3/30, 10%) included decision curve analysis (DCA), and 12 (12/30, 40%) developed decision-support tools. Most studies adequately discussed their limitations, which included small

sample sizes, single-center designs, retrospective biases, and lack of external validation ([Multimedia Appendix 3](#)).

## Quality Assessment

### Overview

[Table 1](#) presents a summary of the risk of bias and applicability for the included studies. Of the 30 studies, 29 were assessed as having a high risk of bias, whereas only 1 was rated as low risk.

**Table 1.** Risk of bias and applicability assessment: comprehensive evaluation of participants, predictors, and analytical methods.

Author, year	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Croft et al, 2024 [14]	+ <sup>a</sup>	+	+	+	– <sup>b</sup>	+	+	+	–
Zhang et al, 2024 [15]	–	+	+	+	+	+	+	–	+
Iacucci et al, 2023 [16]	+	+	+	–	+	+	+	–	+
Chen et al, 2021 [17]	+	+	+	–	+	+	+	–	+
Morilla et al, 2019 [18]	–	+	+	–	–	+	+	–	–
Morilla et al, 2021 [19]	–	+	+	–	–	+	+	–	–
Takayama et al, 2015 [20]	–	+	+	? <sup>c</sup>	+	+	+	–	+
Bu et al, 2023 [21]	–	+	+	–	+	+	+	–	+
Li et al, 2022 [22]	–	+	+	?	+	+	+	–	+
Dai et al, 2024 [23]	–	+	+	?	+	+	+	–	+
Chen et al, 2023 [24]	–	+	+	–	+	+	+	–	+
Yu et al, 2022 [25]	–	+	+	?	–	+	+	–	–
Kang et al, 2022 [26]	+	+	+	–	–	+	+	–	–
Waljee et al, 2018 [27]	–	+	+	–	–	+	+	–	–
Miyoshi et al, 2021 [28]	–	+	+	?	–	+	+	–	–
Morikubo et al, 2024 [29]	–	+	+	–	–	+	+	–	–
Ghiassian et al, 2022 [30]	–	+	+	–	–	+	+	–	–
Sofo et al, 2020 [31]	–	+	?	–	–	–	+	–	–
Feng et al, 2021 [32]	–	+	+	–	–	+	+	–	–
Konikoff et al, 2024 [33]	–	+	+	–	–	+	+	–	–
Cesarini et al, 2017 [34]	–	+	+	+	+	+	+	–	+
Derakhshan Nazari et al, 2023 [35]	+	+	+	–	–	+	+	–	–
Kim et al, 2023 [36]	+	+	+	–	+	+	+	–	+
Lees et al, 2021 [37]	+	+	+	–	–	+	+	–	–
Ghoshal et al, 2020 [38]	–	+	?	?	–	+	+	–	–
Mizuno et al, 2022 [39]	–	+	+	–	–	+	+	–	–
Pang et al, 2023 [40]	–	+	+	–	+	+	+	–	+
Chen et al, 2022 [41]	+	+	+	–	–	+	+	–	–
Wang et al, 2023 [42]	–	+	+	–	–	+	+	–	–
Wang et al, 2023 [43]	–	+	+	–	+	+	+	–	+

<sup>a</sup>“+””: low risk of bias or low concern regarding applicability.

<sup>b</sup>“–””: high risk of bias or high concern regarding applicability.

<sup>c</sup>“?””: unclear risk of bias or unclear concern regarding applicability.



- Participants domain: 22 studies (22/30, 73%) were rated as high risk due to their retrospective cohort design.
- Predictors and outcomes domains: all studies adequately reported these aspects and were rated as having a low risk of bias.
- Analysis domain: 11 studies (11/30, 37%) had an events-per-variable ratio below 10 or a validation sample size smaller than 100. One study (1/30, 3%) did not report sample size estimation. Fourteen studies (14/30, 5%) did not assess model calibration or discrimination, and 4 (4/30, 13%) lacked sufficient information on either. Consequently, 21 studies (21/30, 70%) were rated as having a high risk of bias in this domain, and 6 studies (6/30, 20%) were rated as having unclear risk. In terms of applicability, 18 studies (18/30, 60%) were rated as having high concerns, whereas 12 (12/30, 40%) were rated as having low concerns. Participants domain: 18 studies (18/30, 60%) were judged to have high applicability concerns because of their focus on specific subgroups of patients with UC.
- Predictors domain: 1 study (1/30, 3%) had high applicability concerns due to unreported follow-up times.
- Outcomes domain: all studies had outcomes consistent with the systematic review question and were rated as having low applicability concerns.

## Model Validation and Meta-Analysis

Among the included studies, 16 (16/30, 53%) conducted internal validation, 14 (14/30, 47%) performed external validation, and only 6 (6/30, 20%) provided both internal and external validation. 5 (5/30, 17%) studies provided 95% CIs for internal validation AUC, and 4 (4/30, 13%) reported 95% CIs for external validation AUC. Only 2 studies (2/30, 7%) fully reported both internal and external validation AUC values along with their 95% CIs.

We conducted a meta-analysis of AUC values and their 95% CIs, with subgroup analyses for internal and external validation. The overall pooled AUC was 0.84 (95% CI 0.77-0.92). For internal validation, the pooled AUC was 0.83 (95% CI 0.70-0.96), whereas for external validation, it was 0.87 (95% CI 0.78-0.95). High heterogeneity was observed in internal validation results ([Multimedia Appendix 4](#)). Funnel plots and the Egger regression analysis indicated potential publication bias in internal validation results. Sensitivity analysis by sequentially removing individual studies revealed that excluding the study by Kim et al [36] significantly reduced heterogeneity. Detailed examination of the Kim et al [36] study revealed that its small sample size likely introduced random errors, leading to model overfitting and inflated performance estimates. After excluding this study, the pooled AUC for internal validation was recalculated as 0.78 (95% CI 0.74-0.81; [Multimedia Appendix 4a](#)).

Owing to the limited reporting of sensitivity and specificity values along with their 95% CIs, we calculated the average sensitivity and specificity values for both internal and external validation. For internal validation, the average sensitivity and specificity were 0.814 and 0.761,

respectively (8/30, 27% and 7/30, 23%, respectively). For external validation, the average sensitivity and specificity were 0.830 and 0.757, respectively (11/30, 37% and 9/30, 30%, respectively).

## Discussion

### Principal Findings

We systematically reviewed 30 studies on UC prognostic models. Most studies focused on predicting treatment outcomes, disease progression or relapse, and surgery-related risks, especially the effectiveness of biologic therapies such as TNF- $\alpha$  inhibitors. Logistic regression was the most commonly used method for predictor selection and model construction because of its simplicity, interpretability, and ability to provide insights into the relationships between predictors and outcomes [44]. Common predictors across most UC prognostic models included age, CRP, sex, albumin, erythrocyte sedimentation rate, platelet count, hemoglobin, disease extent/severity, endoscopic findings, and Mayo scores. Prior research has indicated a strong link between age and UC prognosis [45], whereas CRP has been recognized as an effective, noninvasive biomarker for evaluating treatment response in patients with UC [46]. Most studies did not clearly report how missing data were handled, and few developed decision support tools (eg, calculators or nomograms). Small sample size was a common limitation across the majority of studies.

The meta-analysis reported pooled AUCs of 0.84 (95% CI 0.77-0.92) for overall validation, 0.83 (95% CI 0.70-0.96) for internal validation, and 0.87 (95% CI 0.78-0.95) for external validation. A notable and somewhat unexpected finding was that models evaluated through external validation demonstrated a higher pooled AUC than those assessed internally. This contradicts conventional expectations, as external validation typically yields lower predictive performance due to increased heterogeneity and real-world variability. Several factors may account for this discrepancy. First, further analysis of the 4 studies (4/30, 13%) in the external validation subgroup revealed several shared characteristics. Some used advanced modeling techniques, such as convolutional neural networks and artificial neural networks, which may have contributed to improved predictive performance. Additionally, the outcomes predicted by these models, such as drug sustainability or short-term treatment response, were inherently more predictable, which may have facilitated higher accuracy. Second, publication bias may have contributed to the observed results. Studies reporting poor performance in external validation may have remained unpublished or underreported, thereby inflating the pooled estimates. Third, the number of studies reporting external validation AUCs was relatively small (4/30, 13%), which may have introduced statistical uncertainty and led to overestimation of model performance. Moreover, according to the PROBAST risk of bias assessment, all 4 studies (4/30, 13%) were rated as having a high overall risk of bias. Therefore, although the pooled AUC from external validation appears high, this result should be interpreted

with caution due to the limited number of studies, the high risk of bias, and the potential for selective reporting. Regarding sensitivity and specificity, the mean values for internal validation were 0.814 and 0.761, respectively, and for external validation, 0.830 and 0.757. It is important to note that these values were derived using simple unweighted averages due to the inconsistent reporting of 95% CIs across studies. Consequently, this approach does not account for variations in sample size or the precision of estimates, which may introduce bias. However, significant heterogeneity was observed across the included studies, particularly for internal validation. Sensitivity analysis indicated that studies with small sample sizes may have overestimated model performance due to overfitting. This finding underscores the importance of prioritizing high-quality training data and conducting appropriate sample size estimations during model development [47]. Some studies reported calibration curves and DCA, indicating the potential utility of these models in specific clinical scenarios. However, only a minority of studies comprehensively reported performance metrics (eg, 95% CIs for AUC, sensitivity, and specificity), limiting the comparability and robustness of the results.

The PROBAST tool-based quality assessment showed that the majority of studies carried a high risk of bias, especially in aspects concerning participant selection and model analysis. Many studies were limited to specific UC subgroups (eg, moderate-to-severe cases or patients receiving specific treatments), reducing the external applicability of the results. Additionally, some studies did not report sample size calculations or follow-up durations, further restricting the generalizability of the models. This limitation may lead to insufficient evaluation of model performance and reduce the credibility of the results. The lack of standardized approaches for predictor selection and the absence of explicit methods for handling missing data were notable limitations. The frequent use of direct deletion to address missing data may have introduced sample bias. Although 14 studies (14/30, 47%) conducted external validation, most validation datasets were small and lacked multicenter data. Only 2 studies (2/30, 7%) conducted both internal and external validation, raising concerns about the reliability and generalizability of current UC prognostic models in external settings. One study, which focused on a specific population, was identified as having a high risk of limited applicability [14]. However, it demonstrated a low risk of bias, underscoring its potential for future validation in targeted populations.

## Comparison to Prior Work

Artificial intelligence has attracted growing attention in the diagnosis and management of UC. Previous research has primarily concentrated on diagnostic models. For example, Jahagirdar et al [48] conducted a meta-analysis evaluating convolutional neural network-based algorithms for predicting endoscopic severity, whereas Puga-Tejada et al [49] examined artificial intelligence performance in detecting histological remission. However, despite their promise for real-time assessment, these studies primarily address current activity and overlook long-term outcomes.

In contrast, our study is the first to systematically review and meta-analyze prognostic prediction models for UC, focusing on outcomes, such as treatment response, relapse, disease progression, and the need for surgery. We used the PROBAST tool to evaluate model quality rigorously and synthesized key performance measures, including the AUC, sensitivity, and specificity. Furthermore, we assessed model calibration, decision-support tools, and the completeness of reporting. Unlike previous reviews, our study highlights the predictive utility and clinical relevance of prognostic models, providing valuable insights for advancing precision medicine in UC.

## Future Directions

Future studies should prioritize the complete and transparent reporting of performance metrics, such as AUC, sensitivity, specificity, and their 95% CIs, to enable more reliable and meaningful evaluations of model utility. Additionally, researchers should incorporate appropriate sample size calculations and clearly report follow-up durations to improve methodological rigor and ensure reproducibility [50]. Addressing these limitations will enhance the overall quality, transparency, and clinical applicability of predictive models for UC.

In light of the growing significance of prediction models in UC prognosis, we recommend that future research adhere to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guidelines and adopt the 9-step framework for developing and validating clinical prediction models. These frameworks emphasize best practices, including clearly defining the target population and intended end users, selecting high-quality data sources, properly handling missing data, exploring alternative modeling approaches, and rigorously assessing model performance through both internal and external validation [51,52]. For instance, when defining the target population (step 1 of the 9-step framework), researchers should clearly distinguish between patients with newly diagnosed UC and those with long-standing disease, as their clinical trajectories, risk profiles, and treatment responses may differ significantly. During model development, step 4 (handling missing data) is particularly relevant in UC research, where laboratory and endoscopic variables are frequently incomplete. Therefore, appropriate statistical approaches, such as multiple imputation, should be used in place of listwise deletion, which can introduce bias and reduce statistical power. For step 8 (external validation), it is crucial to evaluate model performance using datasets from distinct geographic regions or health care systems. For example, a model developed in a tertiary care center in East Asia could be externally validated using a population-based registry from Europe or North America. This approach helps assess the model's generalizability across diverse UC populations. Furthermore, in alignment with Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis recommendations, future studies should ensure transparent reporting of all aspects of model development, including predictor selection, handling of missing data, and strategies to address

overfitting. Reporting should also include calibration plots, discrimination metrics (eg, AUC or C-index), and their 95% CIs to support comprehensive and interpretable performance evaluation.

## Limitations

This study has several limitations. First, substantial heterogeneity existed among the included studies in terms of study design, population characteristics, modeling approaches, and outcome definitions, which may have affected comparability and introduced variability into the pooled estimates. Second, external validation was limited, with most models relying on small or single-center datasets, thereby restricting their generalizability. Third, many studies did not consistently report key performance metrics, such as 95% CIs for AUC, sensitivity, and specificity, limiting the ability to critically evaluate and compare model performance. Fourth, missing data were often poorly addressed, with many studies using complete-case analysis or listwise deletion, which increases the risk of bias and reduces statistical power. Finally, we only included studies published in English or Chinese and searched 8 major databases, which may have introduced language bias and led to the omission of relevant studies from other languages, sources, or the grey literature.

## Conclusions

This study identified 30 prognostic prediction models for UC, encompassing 7452 patients, with most studies conducted in China and Japan. The majority of studies were retrospective and multicenter, primarily aimed at predicting therapeutic responses, particularly to TNF- $\alpha$  inhibitors, such as infliximab and adalimumab, as well as estimating the risks of surgery, disease progression, or relapse. Logistic regression was the most frequently used method for predictor selection and model development, with commonly used predictors, including age, CRP, albumin, hemoglobin, disease extent, and Mayo scores. The meta-analysis revealed a pooled AUC of 0.84 (95% CI 0.77-0.92). However, notable limitations were identified, including a high risk of bias in most studies, insufficient external validation, and restricted generalization due to small sample sizes and subgroup-specific designs. These findings underscore the need for future research to prioritize robust model development, rigorous external validation, and enhanced generalization to support broader clinical application. Ultimately, well-developed and validated prognostic models hold the potential to guide personalized treatment strategies, enhance clinical decision-making, and improve outcomes for patients with UC in real-world settings.

## Acknowledgments

The authors are grateful to Yi-Ke Song from the Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China, for her valuable assistance in the language editing of this manuscript. Her expertise in English significantly improved the clarity and readability of the final version.

## Funding

This study was supported by several grants, including those from the National Natural Science Foundation of China (grant 82374298), the Beijing University of Chinese Medicine Discipline Backbone Successor Support Program (grant 90010960920033), and the High-Level Traditional Chinese Medicine Key Subject Construction Project of the National Administration of Traditional Chinese Medicine–Evidence-Based Traditional Chinese Medicine (grant zyyzdxk-2023249).

## Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: JL, ZL

Data collection: ZB, YS, ZS, LH, CJ, YL, JA, HS

Data analysis: ZB, YS, ZS, LH, CJ, YL, JA, HS

Methodology development: ZB, YS, ZS, LH, CJ, YL, JA, HS

Supervision: JL, ZL

Writing – original draft: ZB, YS, ZS, LH, CJ, YL, JA, HS

Writing – review & editing: ZB, YS, ZS, LH, CJ, YL, JA, HS, ZL

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Detailed search strategy for all included databases (PubMed, Embase, Cochrane Library, Web of Science, SinoMed, CNKI, Wanfang, VIP).

[\[DOCX File \(Microsoft Word File\), 24 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Comprehensive summary of study characteristics: study design, objectives, and subgroup descriptions of patients with ulcerative colitis.

[\[DOCX File \(Microsoft Word File\), 27 KB-Multimedia Appendix 2\]](#)



### Multimedia Appendix 3

Detailed summary of predictive model characteristics: variable selection methods, model construction techniques, and validation results.

[[DOCX File \(Microsoft Word File\)](#), 46 KB-Multimedia Appendix 3]

### Multimedia Appendix 4

Meta-analysis results on internal validation of area under the curve (AUC) values.

[[DOCX File \(Microsoft Word File\)](#), 265 KB-Multimedia Appendix 4]

### Checklist 1

PRISMA 2020 checklist.

[[DOCX File \(Microsoft Word File\)](#), 20 KB-Checklist 1]

### References

1. Voelker R. What is ulcerative colitis? JAMA. Feb 27, 2024;331(8):716. [doi: [10.1001/jama.2023.23814](#)] [Medline: [38306113](#)]
2. Yu T, Li W, Liu Y, Jin C, Wang Z, Cao H. Application of internet hospitals in the disease management of patients with ulcerative colitis: retrospective study. J Med Internet Res. Mar 18, 2025;27:e60019. [doi: [10.2196/60019](#)] [Medline: [40101745](#)]
3. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet. Dec 23, 2017;390(10114):2769-2778. [doi: [10.1016/S0140-6736\(17\)32448-0](#)] [Medline: [29050646](#)]
4. Hraes L, Windsor JW, Gorospe J, et al. Global evolution of inflammatory bowel disease across epidemiologic stages. Nature New Biol. Jun 2025;642(8067):458-466. [doi: [10.1038/s41586-025-08940-0](#)] [Medline: [40307548](#)]
5. Kamm MA. Rapid changes in epidemiology of inflammatory bowel disease. Lancet. Dec 23, 2017;390(10114):2741-2742. [doi: [10.1016/S0140-6736\(17\)32669-7](#)] [Medline: [29050647](#)]
6. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. Lancet. Apr 29, 2017;389(10080):1756-1770. [doi: [10.1016/S0140-6736\(16\)32126-2](#)] [Medline: [27914657](#)]
7. Knowles SR, Graff LA, Wilding H, Hewitt C, Keefer L, Mikocka-Walus A. Quality of life in inflammatory bowel disease: a systematic review and meta-analyses-part I. Inflamm Bowel Dis. Mar 19, 2018;24(4):742-751. [doi: [10.1093/ibd/izx100](#)] [Medline: [29562277](#)]
8. Sands BE. Biomarkers of inflammation in inflammatory bowel disease. Gastroenterology. Oct 2015;149(5):1275-1285. [doi: [10.1053/j.gastro.2015.07.003](#)] [Medline: [26166315](#)]
9. Colombel JF, Narula N, Peyrin-Biroulet L. Management strategies to improve outcomes of patients with inflammatory bowel diseases. Gastroenterology. Feb 2017;152(2):351-361. [doi: [10.1053/j.gastro.2016.09.046](#)] [Medline: [27720840](#)]
10. Zhang C, Yang H, Liu X, et al. A knowledge-enhanced platform (MetaSepsisKnowHub) for retrieval augmented generation-based sepsis heterogeneity and personalized management: development study. J Med Internet Res. Jun 6, 2025;27:e67201. [doi: [10.2196/67201](#)] [Medline: [40478618](#)]
11. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. N Engl J Med. Jun 4, 2015;372(23):2229-2234. [doi: [10.1056/NEJMsbl503104](#)] [Medline: [26014593](#)]
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](#)] [Medline: [33782057](#)]
13. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. Jan 1, 2019;170(1):51-58. [doi: [10.7326/M18-1376](#)] [Medline: [30596875](#)]
14. Croft A, Okano S, Hartel G, et al. A personalised algorithm predicting the risk of intravenous corticosteroid failure in acute ulcerative colitis. Aliment Pharmacol Ther. Oct 2024;60(7):921-933. [doi: [10.1111/apt.18190](#)] [Medline: [39110549](#)]
15. Zhang S, Lu G, Wang W, et al. A predictive machine-learning model for clinical decision-making in washed microbiota transplantation on ulcerative colitis. Comput Struct Biotechnol J. Dec 2024;24:583-592. [doi: [10.1016/j.csbj.2024.08.021](#)] [Medline: [39281978](#)]
16. Iacucci M, Cannatelli R, Parigi TL, et al. A virtual chromoendoscopy artificial intelligence system to detect endoscopic and histologic activity/remission and predict clinical outcomes in ulcerative colitis. Endoscopy. Apr 2023;55(4):332-341. [doi: [10.1055/a-1960-3645](#)] [Medline: [36228649](#)]
17. Chen X, Jiang L, Han W, et al. Artificial neural network analysis-based immune-related signatures of primary non-response to infliximab in patients with ulcerative colitis. Front Immunol. 2021;12:742080. [doi: [10.3389/fimmu.2021.742080](#)]

18. Morilla I, Uzzan M, Laharie D, et al. Colonic microRNA profiles, identified by a deep learning algorithm, that predict responses to therapy of patients with acute severe ulcerative colitis. *Clin Gastroenterol Hepatol*. Apr 2019;17(5):905-913. [doi: [10.1016/j.cgh.2018.08.068](https://doi.org/10.1016/j.cgh.2018.08.068)] [Medline: [30223112](https://pubmed.ncbi.nlm.nih.gov/30223112/)]
19. Morilla I, Uzzan M, Cazals-Hatem D, et al. Computational learning of microRNA-based prediction of pouchitis outcome after restorative proctocolectomy in patients with ulcerative colitis. *Inflamm Bowel Dis*. Oct 18, 2021;27(10):1653-1660. [doi: [10.1093/ibd/izab030](https://doi.org/10.1093/ibd/izab030)] [Medline: [33609036](https://pubmed.ncbi.nlm.nih.gov/33609036/)]
20. Takayama T, Okamoto S, Hisamatsu T, et al. Computer-aided prediction of long-term prognosis of patients with ulcerative colitis after cytoapheresis therapy. *PLoS One*. 2015;10(6):e0131197. [doi: [10.1371/journal.pone.0131197](https://doi.org/10.1371/journal.pone.0131197)] [Medline: [26111148](https://pubmed.ncbi.nlm.nih.gov/26111148/)]
21. Bu ZJ, Huang ZR, Chen YR, et al. Development and multiple visualization methods for the therapeutic effects prediction model of five-flavor Sophora Flavescens enteric-coated capsules in the treatment of active ulcerative colitis: a study on model development and result visualization. *Eur J Integr Med*. Oct 2023;63:102297. [doi: [10.1016/j.eujim.2023.102297](https://doi.org/10.1016/j.eujim.2023.102297)]
22. Li N, Zhan S, Liu C, et al. Development and validation of a nomogram to predict indolent course in patients with ulcerative colitis: a single-center retrospective study. *Gastroenterol Rep (Oxf)*. 2022;10:goac029. [doi: [10.1093/gastro/goac029](https://doi.org/10.1093/gastro/goac029)] [Medline: [35785264](https://pubmed.ncbi.nlm.nih.gov/35785264/)]
23. Dai C, Dong ZY, Wang YN, Huang YH, Jiang M. Development and validation of a nomogram to predict non-response to 5-aminosalicylic acid in patients with ulcerative colitis. *Rev Esp Enferm Dig*. Mar 2024;116(3):124-131. [doi: [10.17235/reed.2023.9703/2023](https://doi.org/10.17235/reed.2023.9703/2023)] [Medline: [37706463](https://pubmed.ncbi.nlm.nih.gov/37706463/)]
24. Chen J, Zhang Y, Guo Q, et al. Development and validation of a risk model to predict the progression of ulcerative colitis patients to acute severe disease within one year. *Expert Rev Gastroenterol Hepatol*. Dec 2023;17(12):1341-1348. [doi: [10.1080/17474124.2023.2279737](https://doi.org/10.1080/17474124.2023.2279737)] [Medline: [37935651](https://pubmed.ncbi.nlm.nih.gov/37935651/)]
25. Yu S, Li H, Li Y, et al. Development and validation of novel models for the prediction of intravenous corticosteroid resistance in acute severe ulcerative colitis using logistic regression and machine learning. *Gastroenterol Rep (Oxf)*. 2022;10:goac053. [doi: [10.1093/gastro/goac053](https://doi.org/10.1093/gastro/goac053)] [Medline: [36196253](https://pubmed.ncbi.nlm.nih.gov/36196253/)]
26. Kang GU, Park S, Jung Y, et al. Exploration of potential gut microbiota-derived biomarkers to predict the success of fecal microbiota transplantation in ulcerative colitis: a prospective cohort in Korea. *Gut Liver*. Sep 15, 2022;16(5):775-785. [doi: [10.5009/gnl210369](https://doi.org/10.5009/gnl210369)] [Medline: [35975640](https://pubmed.ncbi.nlm.nih.gov/35975640/)]
27. Waljee AK, Liu B, Sauder K, et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther*. Mar 2018;47(6):763-772. [doi: [10.1111/apt.14510](https://doi.org/10.1111/apt.14510)] [Medline: [29359519](https://pubmed.ncbi.nlm.nih.gov/29359519/)]
28. Miyoshi J, Maeda T, Matsuoka K, et al. Machine learning using clinical data at baseline predicts the efficacy of vedolizumab at week 22 in patients with ulcerative colitis. *Sci Rep*. Aug 12, 2021;11(1):16440. [doi: [10.1038/s41598-021-96019-x](https://doi.org/10.1038/s41598-021-96019-x)] [Medline: [34385588](https://pubmed.ncbi.nlm.nih.gov/34385588/)]
29. Morikubo H, Tojima R, Maeda T, et al. Machine learning using clinical data at baseline predicts the medium-term efficacy of ustekinumab in patients with ulcerative colitis. *Sci Rep*. Feb 22, 2024;14(1):4386. [doi: [10.1038/s41598-024-55126-1](https://doi.org/10.1038/s41598-024-55126-1)] [Medline: [38388662](https://pubmed.ncbi.nlm.nih.gov/38388662/)]
30. Ghiassian SD, Voitalov I, Withers JB, Santolini M, Saleh A, Akmaev VR. Network-based response module comprised of gene expression biomarkers predicts response to infliximab at treatment initiation in ulcerative colitis. *Transl Res*. Aug 2022;246:78-86. [doi: [10.1016/j.trsl.2022.03.006](https://doi.org/10.1016/j.trsl.2022.03.006)] [Medline: [35306220](https://pubmed.ncbi.nlm.nih.gov/35306220/)]
31. Sofo L, Caprino P, Schena CA, Sacchetti F, Potenza AE, Ciociola A. New perspectives in the prediction of postoperative complications for high-risk ulcerative colitis patients: machine learning preliminary approach. *Eur Rev Med Pharmacol Sci*. Dec 2020;24(24):12781-12787. [doi: [10.26355/eurrev\\_202012\\_24178](https://doi.org/10.26355/eurrev_202012_24178)] [Medline: [33378027](https://pubmed.ncbi.nlm.nih.gov/33378027/)]
32. Feng J, Chen Y, Feng Q, Ran Z, Shen J. Novel gene signatures predicting primary non-response to infliximab in ulcerative colitis: development and validation combining random forest with artificial neural network. *Front Med (Lausanne)*. 2021;8:678424. [doi: [10.3389/fmed.2021.678424](https://doi.org/10.3389/fmed.2021.678424)] [Medline: [34650991](https://pubmed.ncbi.nlm.nih.gov/34650991/)]
33. Konikoff T, Loebl N, Yanai H, et al. Precision medicine: externally validated explainable AI support tool for predicting sustainability of infliximab and vedolizumab in ulcerative colitis. *Dig Liver Dis*. Dec 2024;56(12):2069-2076. [doi: [10.1016/j.dld.2024.06.008](https://doi.org/10.1016/j.dld.2024.06.008)] [Medline: [38960819](https://pubmed.ncbi.nlm.nih.gov/38960819/)]
34. Cesarini M, Collins GS, Rönnblom A, et al. Predicting the individual risk of acute severe colitis at diagnosis. *J Crohns Colitis*. Mar 1, 2017;11(3):335-341. [doi: [10.1093/ecco-jcc/jjw159](https://doi.org/10.1093/ecco-jcc/jjw159)] [Medline: [27647858](https://pubmed.ncbi.nlm.nih.gov/27647858/)]
35. Derakhshan Nazari MH, Shahrokh S, Ghanbari-Maman L, Maleknia S, Ghorbaninejad M, Meyfour A. Prediction of anti-TNF therapy failure in ulcerative colitis patients by ensemble machine learning: a prospective study. *Heliyon*. Nov 2023;9(11):e21154. [doi: [10.1016/j.heliyon.2023.e21154](https://doi.org/10.1016/j.heliyon.2023.e21154)] [Medline: [37928018](https://pubmed.ncbi.nlm.nih.gov/37928018/)]
36. Kim SY, Shin SY, Saeed M, et al. Prediction of clinical remission with adalimumab therapy in patients with ulcerative colitis by Fourier transform-infrared spectroscopy coupled with machine learning algorithms. *Metabolites*. Dec 19, 2023;14(1):2. [doi: [10.3390/metabo14010002](https://doi.org/10.3390/metabo14010002)] [Medline: [38276292](https://pubmed.ncbi.nlm.nih.gov/38276292/)]

37. Lees CW, Deuring JJ, Chiorean M, et al. Prediction of early clinical response in patients receiving tofacitinib in the OCTAVE Induction 1 and 2 studies. *Therap Adv Gastroenterol*. 2021;14:17562848211054710. [doi: [10.1177/17562848211054710](https://doi.org/10.1177/17562848211054710)] [Medline: [35154388](https://pubmed.ncbi.nlm.nih.gov/35154388/)]
38. Ghoshal UC, Rai S, Kulkarni A, Gupta A. Prediction of outcome of treatment of acute severe ulcerative colitis using principal component analysis and artificial intelligence. *JGH Open*. Oct 2020;4(5):889-897. [doi: [10.1002/jgh3.12342](https://doi.org/10.1002/jgh3.12342)] [Medline: [33102760](https://pubmed.ncbi.nlm.nih.gov/33102760/)]
39. Mizuno S, Okabayashi K, Ikebata A, et al. Prediction of pouchitis after ileal pouch-anal anastomosis in patients with ulcerative colitis using artificial intelligence and deep learning. *Tech Coloproctol*. Jun 2022;26(6):471-478. [doi: [10.1007/s10151-022-02602-3](https://doi.org/10.1007/s10151-022-02602-3)] [Medline: [35233723](https://pubmed.ncbi.nlm.nih.gov/35233723/)]
40. Pang W, Zhang B, Jin L, Yao Y, Han Q, Zheng X. Serological biomarker-based machine learning models for predicting the relapse of ulcerative colitis. *J Inflamm Res*. 2023;16:3531-3545. [doi: [10.2147/JIR.S423086](https://doi.org/10.2147/JIR.S423086)] [Medline: [37636275](https://pubmed.ncbi.nlm.nih.gov/37636275/)]
41. Chen J, Girard M, Wang S, Kisfalvi K, Lirio R. Using supervised machine learning approach to predict treatment outcomes of vedolizumab in ulcerative colitis patients. *J Biopharm Stat*. Mar 2022;32(2):330-345. [doi: [10.1080/10543406.2021.2009500](https://doi.org/10.1080/10543406.2021.2009500)] [Medline: [34882518](https://pubmed.ncbi.nlm.nih.gov/34882518/)]
42. Wang ZY, Tan D, Li S, Abudurexiti W, Gong JF. Prediction of pouchitis after ileal pouch-anal anastomosis based on clinical nomogram of perioperative factors. *Chin J Pract Surg*. 2023;43(10):1158-1161. [doi: [10.19538/j.cjps.issn1005-2208.2023.10.18](https://doi.org/10.19538/j.cjps.issn1005-2208.2023.10.18)]
43. Wang XH, Chen YR, Zheng YY, Wang XY, Sun HY, Wang JY, et al. Construction of an efficacy prediction model for active ulcerative colitis. *J Beijing Univ Tradit Chin Med*. 2023;46(4):528-535. [doi: [10.3969/j.issn.1006-2157.2023.04.012](https://doi.org/10.3969/j.issn.1006-2157.2023.04.012)]
44. Bedogni G. Clinical prediction models—a practical approach to development, validation and updating. *J R Stat Soc Ser A Stat Soc*. Oct 1, 2009;172(4):944. [doi: [10.1111/j.1467-985X.2009.00614.16.x](https://doi.org/10.1111/j.1467-985X.2009.00614.16.x)]
45. Malham M, Vestergaard MV, Bataillon T, et al. The composition of the fecal and mucosa-adherent microbiota varies based on age and disease activity in ulcerative colitis. *Inflamm Bowel Dis*. Feb 10, 2025;31(2):501-513. [doi: [10.1093/ibd/izae179](https://doi.org/10.1093/ibd/izae179)] [Medline: [39150994](https://pubmed.ncbi.nlm.nih.gov/39150994/)]
46. Yarur AJ, Chiorean MV, Panés J, et al. Achievement of clinical, endoscopic, and histological outcomes in patients with ulcerative colitis treated with etrasimod, and association with faecal calprotectin and C-reactive protein: results from the phase 2 OASIS trial. *J Crohn Colitis*. Jun 3, 2024;18(6):885-894. [doi: [10.1093/ecco-jcc/jjae007](https://doi.org/10.1093/ecco-jcc/jjae007)]
47. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. 2019;14(11):e0224365. [doi: [10.1371/journal.pone.0224365](https://doi.org/10.1371/journal.pone.0224365)] [Medline: [31697686](https://pubmed.ncbi.nlm.nih.gov/31697686/)]
48. Jahagirdar V, Bapaye J, Chandan S, et al. Diagnostic accuracy of convolutional neural network-based machine learning algorithms in endoscopic severity prediction of ulcerative colitis: a systematic review and meta-analysis. *Gastrointest Endosc*. Aug 2023;98(2):145-154. [doi: [10.1016/j.gie.2023.04.2074](https://doi.org/10.1016/j.gie.2023.04.2074)] [Medline: [37094691](https://pubmed.ncbi.nlm.nih.gov/37094691/)]
49. Puga-Tejada M, Majumder S, Maeda Y, et al. Artificial intelligence-enabled histology exhibits comparable accuracy to pathologists in assessing histological remission in ulcerative colitis: a systematic review, meta-analysis, and meta-regression. *J Crohn Colitis*. Jan 11, 2025;19(1). [doi: [10.1093/ecco-jcc/jjae198](https://doi.org/10.1093/ecco-jcc/jjae198)]
50. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. Mar 18, 2020;368:m441. [doi: [10.1136/bmj.m441](https://doi.org/10.1136/bmj.m441)] [Medline: [32188600](https://pubmed.ncbi.nlm.nih.gov/32188600/)]
51. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. Jan 7, 2015;350:g7594. [doi: [10.1136/bmj.g7594](https://doi.org/10.1136/bmj.g7594)] [Medline: [25569120](https://pubmed.ncbi.nlm.nih.gov/25569120/)]
52. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. *BMJ*. Sep 3, 2024;386:e078276. [doi: [10.1136/bmj-2023-078276](https://doi.org/10.1136/bmj-2023-078276)]

## Abbreviations

**AUC:** area under the curve

**C-index:** concordance index

**CRP:** C-reactive protein

**DCA:** decision curve analysis

**PICOS:** population, intervention, comparator, outcome, and study design

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PROBAST:** Prediction Model Risk of Bias Assessment Tool

**TNF- $\alpha$ :** tumor necrosis factor-alpha

**UC:** ulcerative colitis

*Edited by Andrew Coristine; peer-reviewed by Ankita Wal, Levente Kovacs, Maryam Almashmoum; submitted 30.Jan.2025; final revised version received 26.Sep.2025; accepted 27.Sep.2025; published 22.Dec.2025*

Please cite as:

Bu Z, Sun Y, Shi Z, Hu L, Ju C, Liu Y, An J, Sun H, Liu J, Liu Z

*Prognostic Prediction Models for Ulcerative Colitis: Systematic Review and Meta-Analysis*

*J Med Internet Res* 2025;27:e71944

URL: <https://www.jmir.org/2025/1/e71944>

doi: [10.2196/71944](https://doi.org/10.2196/71944)

© Zhijun Bu, Yuan Sun, Zeyang Shi, Liming Hu, Chuanlan Ju, Yuan Liu, Jing An, Huiyi Sun, Jianping Liu, Zhaolan Liu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.Dec.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.