Cochrane Library

Cochrane Database of Systematic Reviews

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis (Review)

Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KGM, Monsef I, Foroutan F, Kreuzer KA, Skoetz N

Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K-A, Skoetz N.
Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis.
Cochrane Database of Systematic Reviews 2020, Issue 7. Art. No.: CD012022.
DOI: 10.1002/14651858.CD012022.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Prognosis Review]

Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis

Nina Kreuzberger[1], Johanna AAG Damen[2], Marialena Trivella[3], Lise J Estcourt[4], Angela Aldin[1], Lisa Umlauff[1], Maria DLA Vazquez-Montes[5], Robert Wolff[6], Karel GM Moons[7], Ina Monsef[1], Farid Foroutan[8], Karl-Anton Kreuzer[9], Nicole Skoetz[10]

[1]Cochrane Haematology, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [2]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [3]Centre for Statistics in Medicine, University of Oxford, Oxford, UK. [4]Haematology/Transfusion Medicine, NHS Blood and Transplant, Oxford, UK. [5]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. [6]Kleijnen Systematic Reviews Ltd, York, UK. [7]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands. [8]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. [9]Center of Integrated Oncology Cologne-Bonn, Department I of Internal Medicine, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany. [10]Cochrane Cancer, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

Contact address: Nicole Skoetz, nicole.skoetz@uk-koeln.de.

Editorial group: Cochrane Haematology Group.
Publication status and date: New, published in Issue 7, 2020.

Citation: Kreuzberger N, Damen JAAG, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes MDLA, Wolff R, Moons KG, Monsef I, Foroutan F, Kreuzer K-A, Skoetz N. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. *Cochrane Database of Systematic Reviews* 2020, Issue 7. Art. No.: CD012022. DOI: 10.1002/14651858.CD012022.pub2.

ABSTRACT

Background

Chronic lymphocytic leukaemia (CLL) is the most common cancer of the lymphatic system in Western countries. Several clinical and biological factors for CLL have been identified. However, it remains unclear which of the available prognostic models combining those factors can be used in clinical practice to predict long-term outcome in people newly-diagnosed with CLL.

Objectives

To identify, describe and appraise all prognostic models developed to predict overall survival (OS), progression-free survival (PFS) or treatment-free survival (TFS) in newly-diagnosed (previously untreated) adults with CLL, and meta-analyse their predictive performances.

Search methods

We searched MEDLINE (from January 1950 to June 2019 via Ovid), Embase (from 1974 to June 2019) and registries of ongoing trials (to 5 March 2020) for development and validation studies of prognostic models for untreated adults with CLL. In addition, we screened the reference lists and citation indices of included studies.

Selection criteria

We included all prognostic models developed for CLL which predict OS, PFS, or TFS, provided they combined prognostic factors known before treatment initiation, and any studies that tested the performance of these models in individuals other than the ones included in model development (i.e. 'external model validation studies'). We included studies of adults with confirmed B-cell CLL who had not received treatment prior to the start of the study. We did not restrict the search based on study design.

Data collection and analysis

We developed a data extraction form to collect information based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS). Independent pairs of review authors screened references, extracted data and assessed risk of bias according to the Prediction model Risk Of Bias ASsessment Tool (PROBAST). For models that were externally validated at least three times, we aimed to perform a quantitative meta-analysis of their predictive performance, notably their calibration (proportion of people predicted to experience the outcome who do so) and discrimination (ability to differentiate between people with and without the event) using a random-effects model. When a model categorised individuals into risk categories, we pooled outcome frequencies per risk group (low, intermediate, high and very high). We did not apply GRADE as guidance is not yet available for reviews of prognostic models.

Main results

From 52 eligible studies, we identified 12 externally validated models: six were developed for OS, one for PFS and five for TFS. In general, reporting of the studies was poor, especially predictive performance measures for calibration and discrimination; but also basic information, such as eligibility criteria and the recruitment period of participants was often missing. We rated almost all studies at high or unclear risk of bias according to PROBAST. Overall, the applicability of the models and their validation studies was low or unclear; the most common reasons were inappropriate handling of missing data and serious reporting deficiencies concerning eligibility criteria, recruitment period, observation time and prediction performance measures.
We report the results for three models predicting OS, which had available data from more than three external validation studies:

CLL International Prognostic Index (CLL-IPI)

This score includes five prognostic factors: age, clinical stage, IgHV mutational status, B2-microglobulin and TP53 status. *Calibration*: for the low-, intermediate- and high-risk groups, the pooled five-year survival per risk group from validation studies corresponded to the frequencies observed in the model development study. In the very high-risk group, predicted survival from CLL-IPI was lower than observed from external validation studies. *Discrimination:* the pooled c-statistic of seven external validation studies (3307 participants, 917 events) was 0.72 (95% confidence interval (CI) 0.67 to 0.77). The 95% prediction interval (PI) of this model for the c-statistic, which describes the expected interval for the model's discriminative ability in a new external validation study, ranged from 0.59 to 0.83.

Barcelona-Brno score

Aimed at simplifying the CLL-IPI, this score includes three prognostic factors: IgHV mutational status, del(17p) and del(11q). *Calibration*: for the low- and intermediate-risk group, the pooled survival per risk group corresponded to the frequencies observed in the model development study, although the score seems to overestimate survival for the high-risk group. *Discrimination*: the pooled c-statistic of four external validation studies (1755 participants, 416 events) was 0.64 (95% CI 0.60 to 0.67); 95% PI 0.59 to 0.68.

MDACC 2007 index score

The authors presented two versions of this model including six prognostic factors to predict OS: age, B2-microglobulin, absolute lymphocyte count, gender, clinical stage and number of nodal groups. Only one validation study was available for the more comprehensive version of the model, a formula with a nomogram, while seven studies (5127 participants, 994 events) validated the simplified version of the model, the index score. *Calibration*: for the low- and intermediate-risk groups, the pooled survival per risk group corresponded to the frequencies observed in the model development study, although the score seems to overestimate survival for the high-risk group. *Discrimination*: the pooled c-statistic of the seven external validation studies for the index score was 0.65 (95% CI 0.60 to 0.70); 95% PI 0.51 to 0.77.

Authors' conclusions

Despite the large number of published studies of prognostic models for OS, PFS or TFS for newly-diagnosed, untreated adults with CLL, only a minority of these (N = 12) have been externally validated for their respective primary outcome. Three models have undergone sufficient external validation to enable meta-analysis of the model's ability to predict survival outcomes. Lack of reporting prevented us from summarising calibration as recommended. Of the three models, the CLL-IPI shows the best discrimination, despite overestimation. However, performance of the models may change for individuals with CLL who receive improved treatment options, as the models included in this review were tested mostly on retrospective cohorts receiving a traditional treatment regimen. In conclusion, this review shows a clear need to improve the conducting and reporting of both prognostic model development and external validation studies. For prognostic models to be used as tools in clinical practice, the development of the models (and their subsequent validation studies) should adapt to include the latest therapy options to accurately predict performance. Adaptations should be timely.

PLAIN LANGUAGE SUMMARY

How well do tools predict what happens with adults with newly-diagnosed chronic lymphocytic leukaemia (CLL) over time?

What was the aim of this review?

There are many types of blood cancers called leukaemia. Chronic lymphocytic leukaemia (CLL) is the most common type. Twenty-five per cent of people who have leukaemia have CLL. It is natural for people with newly-diagnosed CLL and their families to want to know what will happen with their health in the future. They may be wondering if or when they will need treatment, if or when their disease will get worse or how long people live with CLL.

Researchers identified several characteristics that are associated with these outcomes. From these characteristics, they have tried to design tools that help predict what may happen to groups of people with newly-diagnosed CLL.

The aim of this Cochrane Review is to evaluate and summarise those tools and studies that test the tools with other patient data.

What are the key messages from this review?

Reviewers found that there is no reliable way to predict what might happen over time to people who have (untreated) CLL. One reason is because the prediction tools have not been tested enough times with enough different people to know how well they really work.

Another reason is because researchers continue to develop more effective CLL treatment options that have better results, and the prediction tools have not kept up with advances in treatment.

What are the main results of the review?

We identified 52 tools that were designed to predict what may happen to people newly-diagnosed with CLL. To find the best tools, we had to select the studies carefully. To apply these tools in clinical practice:

- a tool has to be tested by different researchers to predict what may happen with individuals with CLL in different geographic locations using different groups of people (i.e. age, gender, stage) with CLL. In other words, we would not include a tool if it was only tested on the people who provided their data to create it;

- the results of the tool should be consistent to prove that it works;

- the tests of the tool have to provide enough information to show how well the tool works. For example, the tests have to include large groups of people and enough information about the type of CLL they have.

We found three tools that met these requirements: the CLL International Prognostic Index (CLL-IPI), the Barcelona-Brno score, and the MDACC 2007 index score.

The CLL-IPI did the best job at identifying people who would survive longer with CLL and people who would survive less long. However, we rated the quality of the CLL-IPI studies as low because they did not provide all the information necessary to know how accurate the tool was. The Barcelona-Brno score and the MDACC 2007 index score, tested on a smaller overall number of patients, showed lower discrimination between persons with a good as compared to a worse prognosis, and showed a similarly low quality of the studies.

Conclusion

More and better research is needed to develop and test the tools to help predict how CLL will behave for different groups of people over time. The tools must also adapt to accurately predict the performance of new treatments.

Summary of findings 1. CLL International Prognostic Index

Prognostic models for chronic lymphocytic leukaemia in adult patients

Population: untreated individuals with CLL
Index model: CLL international prognostic index

Timing: moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)

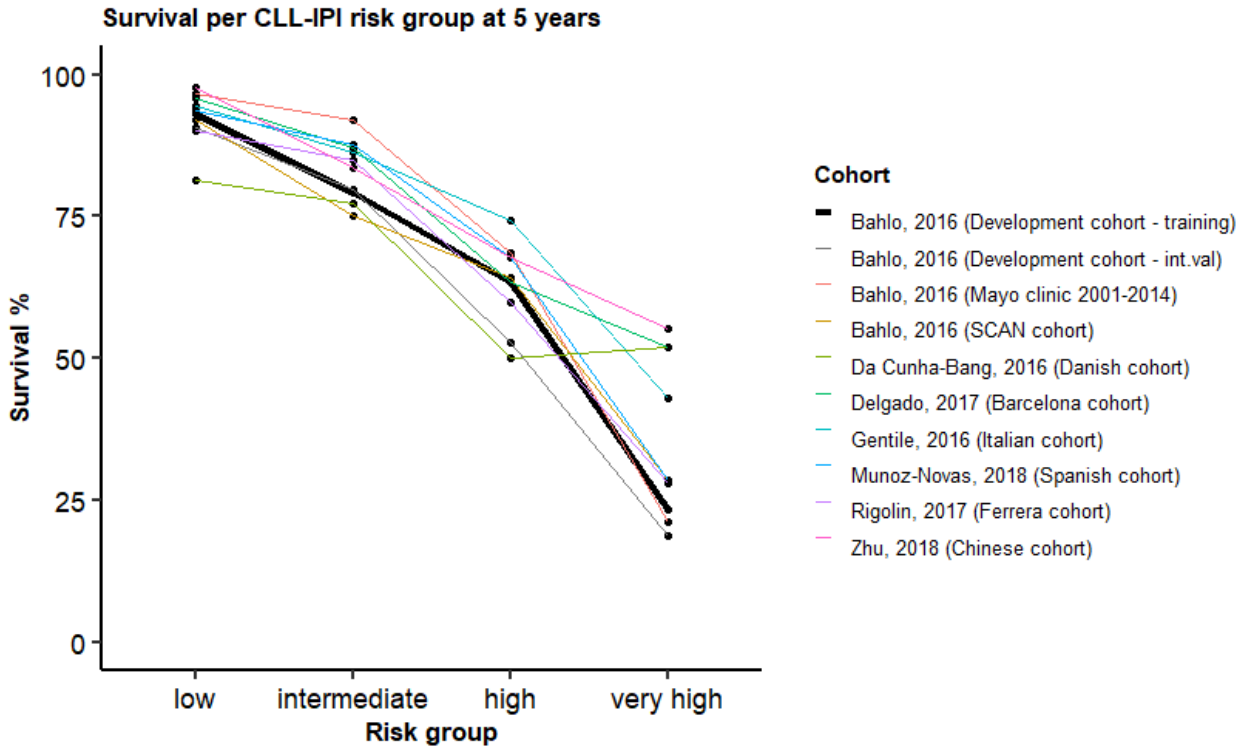
Setting: inpatient and outpatient care

Outcomes	Measure	№ of participants (studies)	Summary measure	Pooled result (95% CI)				Comments
				risk group	n (deaths)	pooled result (95% CI)	95%-PI	
Overall survival (OS)[a]	Discrimination	7 studies 3307 patients 917 deaths	c-statistic of external validation studies			0.72 (95% CI 0.67 to 0.77)	0.59 to 0.83	GRADE[c]
	Calibration[b]	8 studies 4891 patients 875 deaths	survival per risk group	low	2497 (249)	92.5% (89.2% to 94.8%)	82.4% to 97.0%	Survival at 5 years GRADE[c]
				intermediate	1428 (233)	85.0% (79.7% to 89.1%)	68.8% to 93.6%	
				high	765 (280)	64.9% (56.4% to 72.6%)	41.9% to 82.6%	
				very high	201 (113)	40.4% (29.3% to 52.6%)	13.2% to 72.4%	

CI: confidence interval; **PI:** prediction interval

[a]The CLL-IPI was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.
[b]No calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see Figure 19.
[c]GRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

Figure 19. Representation of survival per risk group per development and external validation study of the CLL-IPI (Bahlo 2016)



Summary of findings 2. Barcelona-Brno model

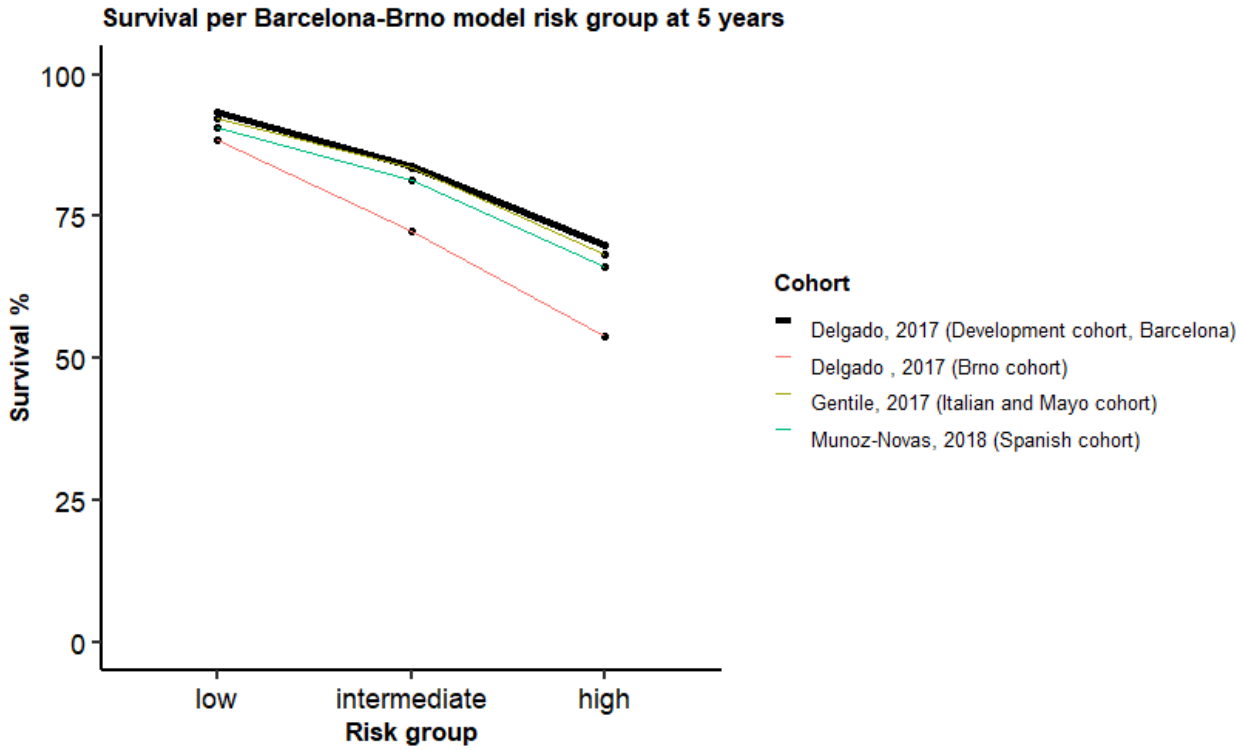
Prognostic models for chronic lymphocytic leukaemia in adult patients							
Population: untreated individuals with CLL **Index model:** Barcelona-Brno model **Timing:** moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included) **Setting:** inpatient and outpatient care							
Outcomes	**Measure**	**№ of participants (studies)**	**Summary measure**	**Pooled result (95% CI)**			**Comments**
				risk group	**n (deaths)**	**pooled result (95% CI)**	**95% PI**

Overall survival (OS)[a]	Discrimination	4 studies 1755 patients 416 deaths	c-statistic of external validation studies			0.64 (95% CI 0.60 to 0.67)	0.59 to 0.68	GRADE[c]
	Calibration[b]	3 studies 1974 patients 317 deaths	survival per risk group	low	1042 (88)	90.5% (85.1% to 94.0%)	80.4% to 95.7%	Survival at 5 years GRADE[c]
				intermediate	673 (131)	79.7% (70.7% to 86.5%)	63.1% to 90.0%	
				high	259 (98)	62.5% (49.3% to 74.1%)	41.3% to 79.7%	

CI: confidence interval; **PI:** prediction interval

[a]The Barcelona-Brno model was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.
[b]No calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see Figure 20.
[c]GRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

Figure 20. Representation of survival per risk group per development and external validation study of the Barcelona-Brno model (Delgado 2017)



Summary of findings 3. MDACC 2007 index score

Prognostic models for chronic lymphocytic leukaemia in adult patients							
Population: untreated individuals with CLL **Index model:** MDACC 2007 index score **Timing:** moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included) **Setting:** inpatient and outpatient care							
Outcomes	**Measure**	**№ of participants (studies)**	**Summary measure**	**Pooled result (95% CI)**			**Comments**
				Risk group	**n (deaths)**	**pooled result (95% CI)**	**95% PI**

Overall survival (OS)[a]	Discrimination	7 studies 5127 patients 994 deaths	c-statistic of external validation studies			0.65 (95% CI 0.60 to 0.70)	0.51 to 0.77	GRADE[c]
	Calibration[b]	5 studies 3786 patients 511 deaths	survival per risk group	low	1202 (35)	97.0% (94.3% to 98.4%)	90.9% to 99.0%	Survival at 5 years GRADE[c]
				intermediate	2425 (393)	82.3% (74.6% to 88.0%)	61.5% to 93.1%	
				high	159 (83)	45.6% (31.3% to 60.5%)	21.2% to 72.3%	

CI: confidence interval; **PI:** prediction interval

[a]The MDACC 2007 index score was developed to predict overall survival. Although we identified external validation studies for other outcomes, we limited our analysis to the primary outcome of the development study.
[b]No calibration measures were reported, and so we used data on survival frequencies per group to compare expected versus observed survival – see Figure 21.
[c]GRADE was not conducted, as currently, no GRADE guidance for prognostic models exists.

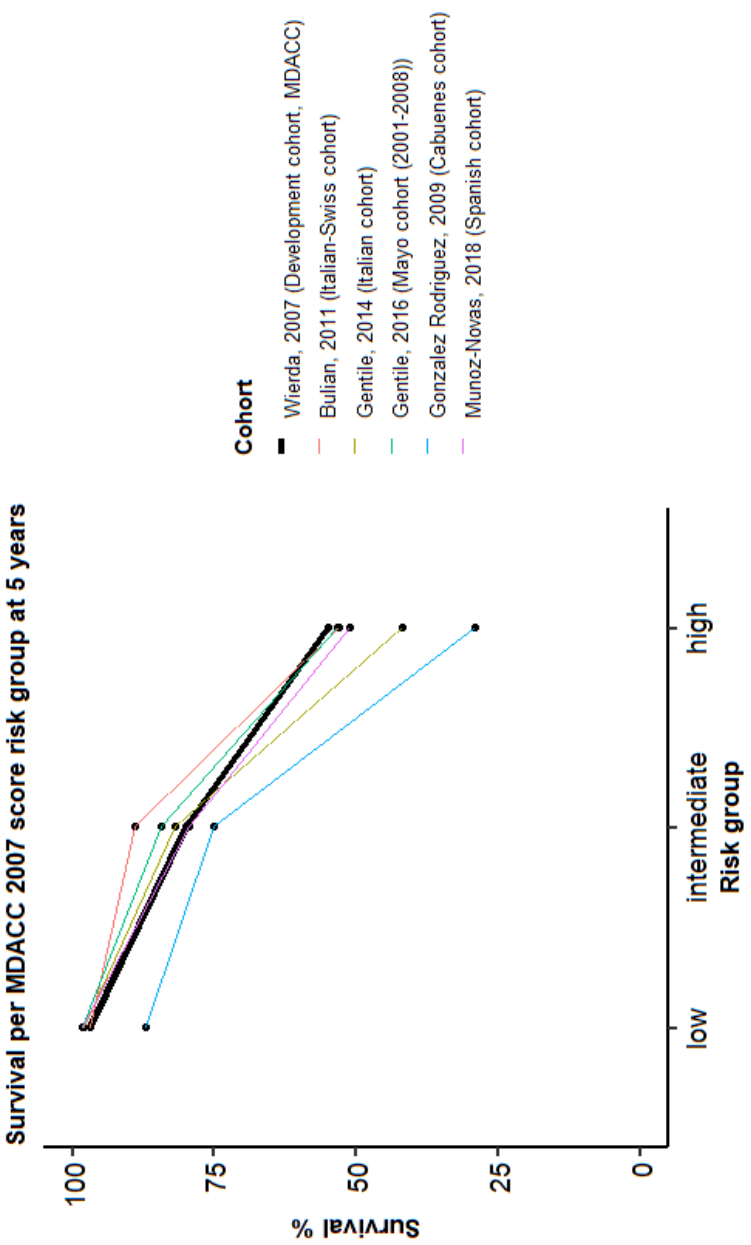


Figure 21. Representation of survival per risk group per development and external validation study of the MDACC 2007 model (Wierda 2007)

BACKGROUND

Description of the condition

Chronic lymphocytic leukaemia (CLL) is the most common form of malignant neoplasm (cancer) of the lymphatic system in Western countries. It is responsible for 25% of all leukaemias and occurs mainly in the elderly population (Chiorazzi 2005). The reported age-adjusted incidence rate of CLL in the USA between 2012 and 2016 was 4.9 per 100,000 persons with an estimated 20,720 new cases and an age-adjusted death rate of 1.2 per 100,000 persons per year (Howlader 2019). In the European Union, an estimated 46,000 individuals were living with CLL five years post-diagnosis in 2006 (Watson 2008).

In CLL, mature B cells accumulate in the blood, bone marrow, lymph nodes and spleen (Herishanu 2013). The diagnosis of CLL is generally established based on blood counts, differential counts, blood smears and immunophenotyping (Hallek 2017). The requirement for a diagnosis of CLL has been modified from a chronic absolute lymphocytosis with more than 5.0×10^9 cells per L to an absolute count of more than 5.0×10^9 monoclonal B cells with CLL immunophenotype per L peripheral blood for the duration of at least three months (Hallek 2008; Hallek 2017). In the case of a monoclonal B-cell count lower than 5.0×10^9 per L and the absence of disease-related symptoms, cytopenias or tissue involvement other than bone marrow, the condition is defined as monoclonal B-lymphocytosis (Hallek 2008). A diagnosis of small lymphocytic lymphoma (SLL) is made when lymphadenopathy (enlarged lymph nodes) or splenomegaly (enlarged spleen) are caused by infiltrating CLL cells, and B lymphocytes in the peripheral blood do not exceed 5.0×10^9 per L (Hallek 2008).

CLL is characterised by a highly variable clinical course and prognosis. Some individuals with CLL experience no or only few symptoms over many years, do not require treatment, and have a life expectancy comparable to that of a healthy individual. Other individuals already experience symptoms at diagnosis or shortly thereafter, and die within a few years despite treatment. The heterogeneity in clinical presentation makes it difficult for the physician to predict accurately whether a patient may benefit from an early, aggressive treatment strategy, and to provide the patient with relevant prognostic information.

The most commonly used early attempts to group individuals with CLL according to their risk are the two staging systems by Binet 1981 (Binet 1981) and Rai 1975 (Rai 1975), which distinguish between early (Rai 0; Binet A), intermediate (Rai I, II; Binet B) and advanced stages (Rai III, IV; Binet C). The disease stage is determined by the number of lymphocytes in the peripheral blood, presence of enlarged lymph nodes, presence of anaemia or thrombocytopenia (low platelet count), and presence of an enlarged liver or spleen. The prognostic value of the two staging systems is limited as survival times vary significantly within these stages. According to the guidelines by the European Society for Medical Oncology, B-symptoms at diagnosis (fever, night sweats or weight loss) categorise people with early stage disease according to Binet or Rai, and are considered to have aggressive disease, requiring treatment (Eichhorst 2015).

Prognostic models

A prognostic model is a mathematical function that considers at least two prognostic factors, also called predictors, simultaneously with the aim to provide an estimate of an individual patient's probability to experience a certain health event within a defined time frame in the future (Alba 2017; Moons 2009; Riley 2019; Steyerberg 2013). Prognostic factors may be characteristics of the individual or the disease (e.g. age, gender, disease stage, biological or genetic information), which are likely to predict a patient-relevant outcome, such as overall survival or disease progression (Riley 2013). By statistical modelling methods, these factors are often combined to form a weighted model able to accurately predict the likelihood of this outcome. Such a model aims at assisting the clinician to estimate the patient's prognosis and enhance shared decision-making. Establishing a prognosis for the individual patient may also lead to risk-stratified treatment recommendations (Alba 2017; Debray 2017; Moons 2009; Riley 2019; Steyerberg 2013).

To develop a prognostic model for a specific outcome, such as death or disease progression, various data sources such as cohort, nested case-control or case-cohort studies are considered appropriate, especially when data are prospectively collected. Data originating from randomised clinical trials, as a special form of prospectively collected data, can be used, but may limit generalisability due to more restrictive eligibility criteria, selective participation of specialised centres, trial effects and unrealistically precise predictor assessments (Collins 2015; Moons 2014; Moons 2019; Pajouheshnia 2019). Before a prognostic model is used in clinical practice, its predictive performance should be quantified. Apparent performance of the model is the model performance estimated from the same data as used for model development and usually provides an overly optimistic estimate due to overfitting of the model to this specific dataset. Internal validation tests the model performance in the development dataset by using techniques such as bootstrapping or cross-validation. In order to test the generalisability of a prognostic model, predictive performance should be ideally assessed in several independent sets of data of individuals that were not used in the development and internal validation of the model, preferably by independent investigators to reduce bias. This process is called external validation (Moons 2015).

In the past few years, a number of clinical and biological CLL prognostic factors have been identified, including genomic aberrations, gene abnormalities (p53, ATM), mutation status of the variable segments of the immunoglobulin heavy chain genes (IgHV), or surrogate markers for these factors, such as CD38 and ZAP-70 expression (Döhner 2000; El Rouby 1993; Kay 2007). Assessing recent molecular markers at time of diagnosis is expected to provide more reliable information regarding optimal time for treatment initiation, type of therapy and individual prognosis (Kay 2007; Shanafelt 2004; Wierda 2011; Zenz 2011). Thus, progressive and smouldering forms of the disease can now be separated more accurately than by using Rai or Binet staging systems alone. Moreover, the early recognition of aggressive stage A and indolent stage B and C disease would allow rational application of risk-adapted treatment strategies. Factors influencing the choice of treatment include age, fitness to tolerate chemotherapy or immunotherapy or both, TP53 status, previous or current immune

cytopenias, and evidence of lymphomatous transformation (Goede 2012).

The recent increase in availability of biological markers for CLL presents a challenge, as well as an opportunity to develop more precise prognostic models or algorithms that integrate a combination of markers and may guide counselling and treatment decisions for the individual patient. In order to identify the best-performing tool to estimate prognosis for untreated individuals with CLL, a comprehensive evaluation of all currently available prognostic models and meta-analysis of their predictive performance in external validation studies is urgently needed.

Health outcomes

The highly variable course of CLL and the possibility of having a normal life expectancy without progression or need for treatment entails that overall survival (OS) is one of the most important outcomes to be predicted by a prognostic model. In the USA, the median age at diagnosis is 70 years and the five-year survival rate with CLL is 85.1% (Howlader 2019). Therefore, it is important to observe patients as long as possible to obtain a prognostic model that is meaningful not only for high-risk individuals, but also for people with a less aggressive disease and longer survival.

As individuals with CLL are usually older, which implies an increased prevalence of comorbidities and decreased physical fitness, treatment may lead to serious adverse events and interactions with other medications. Hence, alternative meaningful outcomes to be predicted by a prognostic model include progression-free survival (PFS) or treatment-free survival (TFS, also sometimes referred to as time-to-first-treatment). Treatment options for CLL have improved over time, thus, affecting survival rates but not the rates of treatment indication.

Why it is important to do this review

Although several prognostic factors have been identified during the last decade (Pflug 2014; Stilgenbauer 2014; Zaja 2013), they are controversial and there is no single prognostic factor available to determine treatment options in CLL patients. These factors have also been combined into numerous prognostic models. To date, no systematic review has been conducted to evaluate and assess the predictive performance of prognostic models in CLL, which would inform us which models have the greatest validity and therefore, would be preferred to guide clinical decision-making. To shed light on this important research question, we conducted a systematic review and, where possible, a meta-analysis of existing prognostic models for CLL and their corresponding validation studies.

OBJECTIVES

To identify, describe and appraise all prognostic models developed to predict overall survival (OS), progression-free survival (PFS) or treatment-free survival (TFS) in newly-diagnosed (previously untreated) adults with CLL, and meta-analyse their predictive performances.

METHODS

Criteria for considering studies for this review

Table 1. PICOTS system for prognostic models

P	Population	Untreated individuals with chronic lymphocytic leukaemia (CLL) at time of prediction
I	Index model(s)	All developed prognostic models for CLL and their corresponding external validation studies
C	Comparator	No predefined comparator
O	Outcome(s)	Overall survival (OS), progression-free survival (PFS), treatment-free survival (TFS)
T	Timing	Moment of prediction at diagnosis of CLL; moment of outcome occurrence not prespecified (any moment after diagnosis was included)
S	Setting	Not specified

Types of studies

According to the checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) (Moons 2014), we included:

- prognostic model development studies without external validation in independent data;
- prognostic model development studies with external validation in independent data reported in the same study; and
- external model validation studies of previously reported models.

A prognostic model was defined as some form of mathematical function including at least two independent factors to predict OS, PFS or TFS. We included only models that combined prognostic factors known and assessed before treatment initiation. We excluded prognostic factor finding studies (i.e. studies aiming to establish one or several variables as independent prognostic factor(s) associated with an outcome, but not aiming to develop

a model to be used for individualised predictions (i.e. to provide absolute prognostic outcome probabilities) in new patients, and model impact studies (i.e. studies aiming to investigate the impact of the use of a model in practice) (Bouwmeester 2012). We excluded studies that were published only as conference proceedings. We did not exclude studies with treatment during follow-up time.

Participants

We included studies on individuals with previously untreated, confirmed B-cell CLL. We included both male and female adults (age ≥ 18 years).

Types of outcomes to be predicted

Primary outcome

- Models that predict OS as an outcome

We chose this outcome because it has the greatest clinical relevance and is most important for patients. Furthermore, death due to any cause is an objective endpoint not susceptible to bias of the outcome assessor. We did not require studies to have a minimum follow-up time for inclusion in this review.

Secondary outcomes

- Models that predict either PFS or TFS

We chose these outcomes as patients with similar survival may nevertheless have differing lengths of time without symptoms or need for treatment, depending both on initial treatment and disease characteristics. In case of immediate start of treatment, identification of patients with a lower probability to obtain a good response will help in making decisions regarding treatment, for example, deciding which patients might receive new or more aggressive therapy regimens. In case of a watch-and-wait strategy, differences in estimated prognosis can influence patient management regarding surveillance and treatment.

Search methods for identification of studies

Electronic searches

Reporting and therefore retrieval of prognostic model studies is very poor, as guidelines on reporting of prediction models have only recently been published (Collins 2015). For our first search, we did not use a specific search filter (Appendix 1). As this search strategy was not very specific, it yielded many results, which had to be screened in detail by two review authors. For the updated search, we integrated the search filter by Geersing 2012 to allow for a more specific strategy (Appendix 2).

We searched the following databases without applying any language restrictions in order to reduce language bias.

- MEDLINE via Ovid (searched 24 June 2019; Appendix 1; Appendix 2).
- Embase (searched 24 June 2019; Appendix 3).

We searched the following databases for ongoing trials.

- ClinicalTrials.gov (clinicaltrials.gov; searched 5 March 2020; Appendix 4).
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) (searched 5 March 2020; Appendix 5).

Initially, we planned to include only models published after 1990. However, prognostic model studies are often conducted based on retrospective data, which in some cases included the analysis of blood samples frozen at diagnosis and analysed many years later. Thus, we decided not to limit our search strategy to date of publication. We did not screen conference proceedings, according to the section 'Types of prognostic models', where we stated that we would exclude conference abstracts based on limited information that would not allow us to assess risk of bias. In contrast to our planning (Skoetz 2016), we did not search the database of prognostic studies by the Prognostic Methods Group as it has not been developed. We did not search conference proceedings because of their limited information that would have complicated the inclusion and exclusion of studies as well as the risk of bias rating.

Searching other resources

As prespecified in the protocol (Skoetz 2016), we searched the following sources.

- Handsearching of references
  - References of all identified trials, relevant review articles and current treatment guidelines for further literature.
- Personal contacts
  - We contacted authors of relevant studies, study groups, experts and investigators from transplantation centres worldwide who are known to be active in the field for unpublished material or further information on ongoing studies.

In addition, and not reported in the protocol of this review, we tracked the citations of all included studies in October 2018 to identify possible additional validation studies of a developed model (Web of Science citation indexing).

Data collection and analysis

Selection of studies

During independent screening of titles and abstracts by two review authors (AA, LE, MT, NK, NS), we organised regular group discussions for congruency due to the novelty of the topic. The main discussion points can be found below. After title and abstract screening, full texts for all eligible studies were obtained and independently selected by two review authors (AA, LE, MT, NK, NS). Disagreements were solved by involving one or more additional review authors (Lefebvre 2019).

As recommended in the PRISMA statement (Moher 2009), we documented the total number of retrieved references and numbers of included and excluded studies in a flow chart.

Solving disagreements in study inclusion

The poor reporting and the fine line between the development of a prognostic model and prognostic factor (identification) studies made the screening of titles and abstracts and the final decision regarding the inclusion of studies challenging. During several group discussions, the following issues emerged repeatedly.

- As we changed our pre-planned limitation of the search to include studies from inception of the database, we obtained many studies labelled as "staging system." A staging system intends to describe the severity of a disease and, thus, indicate

some prognostic and therapeutic information. Two staging systems for CLL, Rai and Binet, are commonly used in clinical practice. However, the predicted survival of individuals with the same disease stage can be highly variable. We decided to exclude staging systems, as they are not used by clinicians any longer (except for Rai and Binet). Instead, we focused on new algorithms with a high chance to include more recently identified prognostic factors.

- We decided to exclude studies that build genetic signatures aiming to distinguish between good and poor prognosis of an individual, due to several reasons. First, the development of a genetic signature is often preceded by the identification of genetic markers on the same cohort from several hundred candidate markers, which introduces a high potential for overfitting of the algorithm to the development cohort (e.g. Houldsworth 2014). Second, genetic signatures can consist of a tremendous number of genes compared to the number of individuals included in the cohort (e.g. Ferreira 2014). Third, it is currently difficult to apply the algorithm of a genetic signature to an external cohort due to the complexity of the model.
- The fine line between a multivariable model, which includes several prognostic factors to prove their independence from each other, and a prognostic model was not always evident from the abstract. When an abstract mentioned a multivariable model and the formation of risk groups, we considered the full text of the paper. We then excluded papers that did not proceed to build a prognostic model or score explicitly.
- References were classified as an external validation study if the term "validation" was explicitly stated or the application of the prognostic model was clearly the main focus of the paper. We also included some publications where validation was not explicitly mentioned and performance measures were not reported, but where the authors put their focus on the application of one of the previously developed prognostic models (e.g. CLL-IPI V - Reda 2017 (Milano cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort)).

The decisions detailed above may seem subjective and were made in the case of this specific review. We would advise future authors who aim to evaluate prognostic models for a certain disease to clearly define the eligibility of different types of prognostic models at the protocol stage. This will help with objectivity throughout the screening process.

Data extraction and data management

Teams of two review authors (AA, LE, MT, NK, NS) independently extracted the data to enable assessment of applicability of the model for the review and the 'Risk of bias' assessment (see below). We contacted authors of individual studies for additional information, where required. We developed a standardised data extraction form containing the following items based on the CHARMS checklist (Moons 2014).

- General information
  - Author, title, publication date, country, language, duplicate publications
- Source of data
  - E.g. cohort, case-control, randomised trial participants, or registry data
- Participants
  - Participant eligibility and recruitment method (e.g. consecutive participants, location, number of centres, setting, inclusion and exclusion criteria)
  - Participant description
  - Details of treatments received
  - Study dates
- Outcomes to be predicted
  - Definition and method for measurement of outcome
  - Was the same outcome definition (and method for measurement) used in all patients?
  - Was the outcome assessed without knowledge of the candidate predictors (i.e. blinded)?
  - Time of outcome occurrence or summary of duration of follow-up
- Candidate predictors
  - Number and type of predictors (e.g. demographics, patient history, physical examination, additional testing, disease characteristics, tumour markers)
  - Definition and method for measurement of candidate predictors
  - Timing of predictor measurement (e.g. at patient presentation, at diagnosis, at treatment initiation)
  - Were predictors assessed blinded for outcome, and for each other (if relevant)?
  - Handling of predictors in the modelling (e.g. continuous, linear, non-linear transformations or categorised)
- Sample size
  - Number of participants and number of outcomes/events
  - Number of outcomes/events in relation to the number of candidate predictors (events per variable)
- Missing data
  - Number of participants with any missing value (include predictors and outcomes)
  - Number of participants with missing data for each predictor
  - Handling of missing data (e.g. complete-case analysis, imputation, or other methods)
- Model development
  - Modelling method (e.g. logistic, survival, neural networks, or machine learning techniques)
  - Modelling assumptions satisfied
  - Method for selection of predictors for inclusion in multivariable modelling (e.g. all candidate predictors, pre-selection based on unadjusted association with the outcome)
  - Method for selection of predictors during multivariable modelling (e.g. full model approach, backward or forward selection) and criteria used (e.g. P value, Akaike Information Criterion)
  - Shrinkage of predictor weights or regression coefficients (e.g. no shrinkage, uniform shrinkage, penalised estimation)
- Model performance
  - Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals
  - Classification measures (e.g. sensitivity, specificity, predictive values, net reclassification improvement) and whether a priori cut-points were used

- Model evaluation
  - Method used for testing model performance: development dataset only (random split of data, re-sampling methods, e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)
  - In case of poor validation, whether model was adjusted or updated (e.g. intercept re-calibrated, predictor effects adjusted, or new predictors added)
- Results
  - Final and other multivariable models (e.g. basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)
  - Any alternative presentation of the final prediction models, e.g. sum score, nomogram, score chart, predictions for specific risk subgroups with performance
  - Comparison of the distribution of predictors (including missing data) for development and validation datasets
- Interpretation and discussion
  - Interpretation of presented models (confirmatory, i.e. model useful for practice versus exploratory, i.e. more research needed)
  - Comparison with other studies, discussion of generalisability, strengths and limitations

References of model development studies, for which no external validation studies have yet been published, are summarised in Studies awaiting classification. These studies fit the formal inclusion criteria of this review. However, they have not yet been tested in any additional cohort, which renders summary of performance in external cohorts impossible. They cannot yet be evaluated regarding their clinical usefulness, and are therefore not yet described in detail.

Risk of bias and applicability assessment

Since the publication of the review protocol, a 'Risk of bias' tool specifically designed for prognostic model studies (Prediction model Risk Of Bias ASsessment Tool, PROBAST) was published (Moons 2019; Wolff 2019). Instead of the CHARMS checklist for critical appraisal of prognostic modelling studies as reported in our protocol, we used this new tool recommended by the Cochrane Prognosis Methods Group to assess the risk of bias of the individual prediction models investigated in the included primary studies. Teams of two review authors (AA, LE, MT, NK, NS) independently assessed the risk of bias and applicability for each study. We assessed only validation studies with the outcome that a model was developed for, in line with meta-analyses, hence PROBAST assessments reflect only the outcome of interest and not any additional outcomes reported in the same study.

After classifying each study into one of the three categories (model development with or without external validation in the same publication and external validation study of a previously developed model only), we assessed risk of bias according to the following four PROBAST domains (once per developed or validated model and per outcome).

- Participants
- Predictors
- Outcome
- Analysis

We answered signalling questions within each domain with one out of five options ('yes', 'probably yes', 'probably no', 'no', 'no information'); 'yes' always implied the absence of a potential bias-generating aspect. We rated domain-level 'Risk of bias' assessments using one of the following three options.

- Low risk of bias: if the criterion is adequately fulfilled in the study, i.e. the study is at a low risk of bias for the given domain.
- High risk of bias: if the criterion is not fulfilled in the study, i.e. the study is at high risk of bias for the given domain.
- Unclear risk of bias: if the study report does not provide sufficient information to allow for a clear judgement or if the risk of bias is unknown for one of the domains listed above.

PROBAST additionally requires the judgement of the applicability of the model to the research question. The assessment occurs per domain (for the first three domains only) with the following response options: 'low concern regarding applicability', 'high concern regarding applicability' and 'unclear concern regarding applicability' (equivalent to the categories for risk of bias).

Based on these domain-level judgements, we established an overall judgement per study based on PROBAST guidance for risk of bias:

- Low risk of bias: if a prediction model evaluation is judged as low on all domains relating to bias or applicability.
- High risk of bias: if an evaluation is judged as high for at least one domain.
- Unclear risk of bias: if the prediction model evaluation is unclear in one or more domains and was rated as low in the remaining domains.

When information for a complete judgement was missing, we contacted the corresponding authors via email to request additional information to be able to make sound judgements on the 'Risk of bias' assessment. According to PROBAST, models that were developed, but not (yet) externally validated can be classified as 'high risk of bias', with the exception of extremely large samples (Wolff 2019). Due to the novelty of PROBAST and the subjectivity of ratings, we devoted a small section of the results to present our experience and agreements. Two authors of PROBAST (RW, KGMM) are also authors in our review and RW was involved in group discussions addressing disagreements between ratings.

Solving disagreements regarding PROBAST ratings

Due to the novelty of the 'Risk of bias' tool (PROBAST), we arranged frequent team meetings to discuss disagreements and discrepancies regarding ratings between the six authors who extracted data and assessed risk of bias and applicability. For reasons of transparency, we report the following challenges and our agreements.

- The relevant observation time varies between diseases. For an indolent neoplasm like CLL, the clinicians in our team considered a median follow-up of five years as appropriate for OS in a cohort with a normal case mix (i.e. no limitation to patients with late stage or high-risk disease such as patients with del(17p) only).

- The signalling question for the participant domain asks whether all inclusions and exclusions of participants were appropriate (item 1.2). Some publications listed missing values as part of the exclusion criteria. Missing values are also treated in the analysis domain (items 4.3 and 4.4). To avoid duplicate rating of the same issue, we decided to rate the risk of bias as high for the participant domain when complete data were required for eligibility, and rated the risk of bias as high for the analysis domain when the number and characteristics of the background sample were described but persons were excluded from analysis due to missing values.
- Some publications did not report eligibility criteria and/or recruitment period. We rated these studies as unclear not only for risk of bias, but also for concerns of applicability, because we could not be sure if the included individuals matched our review question and a current application of the model.
- When a publication lacked information necessary to answer a signalling question, but other available information could be used as an indication, we made assumptions regarding the answer (probably yes, probably no) to avoid the option 'no information'.
- For the CLL-IPI, we frequently encountered the problem that one predictor was replaced by a proxy predictor in the analysis of an external validation study. We assumed that this may happen in clinical practice, and included the external validation study to validate the CLL-IPI, although strictly speaking, they did not apply the original model. We rated the concern for applicability as unclear in the domain predictor.
- Some models developed to predict TFS include predictors also used for treatment indication, which means that the predictors are not excluded from the outcome definition (PROBAST item 3.3; e.g. O-CLL1 model: Rai staging is included in the prognostic score, and is, in combination with other characteristics, a factor for treatment indication). This may lead to an overestimation of the association between predictor and outcome, which is called incorporation bias (Moons 2019). We answered the corresponding PROBAST item 3.3 with 'no' and considered the risk of bias to be high for the outcome domain. We decided that all corresponding external validations, which validate a model that include a predictor also used for treatment indication, are also considered to contain a high risk of bias.
- The PROBAST guidance for item 4.1, which states that a study should include at least 20 events per candidate predictor for development studies and 100 events for validation studies is recommended for regression analysis with a binary outcome. The models included in our systematic review are mainly based on Cox proportional hazard models (Moons 2019). We therefore considered a cohort with more than 100 events to be an appropriate sample size for development and external validation samples (Collins 2016).
- The testing of model assumptions (PROBAST item 4.6) was very rarely reported. We did not let this item influence our judgement regarding the analysis domain because, by default, all except one study would have been considered to be of unclear risk of bias. We assumed that authors tested the pre-conditions before using their analysis type. However, we would advise authors of primary studies to improve reporting of the entire model development process.

Measures of prediction model performance

In contrast to 'classical' meta-analysis focusing on treatment efficacy as the parameter of interest, there is no single recommended methodology to meta-analyse the predictive performance of prognostic models (Debray 2012; Debray 2014). Hence, we extracted the reported predictive performance measures for each studied model from the development and validation studies, notably the model's discrimination and calibration in the development cohort (i.e. apparent performance, the performance measures after internal validation) and the performance found in a model's external validation.

Calibration refers to the accuracy of the predicted risk probabilities, which means the agreement between estimated and observed number of events in a cohort. A model might predict more events than the number of observed events, i.e. there is overestimation of the number of events. In contrast, a model can predict fewer events than observed meaning that the number of events is underestimated. In both situations, the model is miscalibrated. The use of a miscalibrated model in clinical practice can misguide clinical decisions based on such models (Van Calster 2019). Calibration can be presented as a calibration plot (expected probabilities plotted against observed outcome frequencies), as ratios between observed and expected number of events or outcome frequencies (O:E ratios) or calibration table (Debray 2017).

Discrimination refers to the ability of a prediction model to differentiate between those who do or do not experience the outcome event. A model has perfect discrimination if the predicted risks for all individuals who develop the outcome are higher than those for all individuals who do not experience the outcome. Discrimination is commonly estimated by the so-called concordance index (c-statistic, also sometimes called c-index). The c-statistic reflects the probability that for any randomly selected pair of individuals, one with and one without the outcome, the model assigns a higher probability to the individual with the outcome. The c-index can range from 0 to 1, with 1 indicating perfect discriminative ability and 0.5 meaning the model's predictions equal chance. The c-index is identical to the area under the receiver-operating characteristic (ROC) curve for models with binary endpoints, and can be generalised for time-to-event (survival) models accounting for censoring (Debray 2017).

Both measures are needed to assess the performance of a prognostic model.

Unit of analysis issues

Ideally, a prognostic model is developed and tested on specially designed, prospectively followed cohort studies consisting of a representative case mix. However, more often we identified models or validations derived from retrospective data or data from randomised controlled trials (RCTs). Data from these populations are used for several purposes and models. Where a model was developed and externally validated in the same publication, we compared the institution and year of inclusion of individuals to assure independence of the cohorts. We also compared the institution and year of inclusion of all external validation cohorts belonging to one developed model to avoid overlaps in participants. When the same cohort was used to validate two or more different prognostic models, we considered them as separate studies. When RCT data were used, the randomisation of

individuals was not considered by the authors of prognostic model studies.

Dealing with missing data

When data were missing, we requested additional information from the original investigators. When confidence intervals for measures of discrimination were still missing after we contacted the author, we calculated these according to a method proposed by Newcombe (Debray 2017; Debray 2018a; Newcombe 2006).

Investigation/description of heterogeneity

For the c-statistic, we used the between-study standard deviation (tau - τ) to quantify possible heterogeneity. As we did not summarise O:E ratios to pool calibration, we merely visually inspected differences in survival per subgroup between cohorts. We planned to investigate and discuss clinical and statistical heterogeneity and design aspects of included studies mentioned in the section 'Data extraction and data management' based on subgroup analysis. We were unable to perform prespecified subgroup analysis based on diagnostic criteria due to the low number of studies with clear distinction between criteria.

Discussing reporting deficiencies

It is widely recommended that all prognostic models assess and report calibration and discrimination (Collins 2015; Moons 2015). However, it is known from numerous systematic reviews on methodological conduct and reporting of prognostic models in various disciplines that calibration is rarely reported and when it is reported (Heus 2018), it is done quite poorly. Hence, we also evaluated reporting deficiencies.

Methods and reporting in prognostic research often do not follow current methodological recommendations, limiting retrieval, reliability and applicability of these publications (Bouwmeester 2012; Peat 2014). There are some indications that prognosis research is cluttered with false-positive studies which would not have been published if the results were negative. Moreover, studies evaluating development studies of prognostic models are not prospectively registered, and usually no protocol is published (Peat 2014). Therefore, it is difficult to assess publication bias. We used sensitive search strategies to increase retrieval (Geersing 2012).

Data synthesis

Data synthesis and meta-analysis approaches

In the protocol, we explained that we would pool the performance measures for calibration and discrimination. During the review process however, new methodological developments became available and we chose to adopt them, as follows. Although we assessed both model development and external validations for risk of bias, we performed meta-analysis only for validation studies, as there was only one development study per model.

- When a particular model has been validated at least three times for the primary outcome, we applied a random-effects model for pooling the logit transformation of the discrimination measure (the c-statistic) using the meta-analysis packages 'metamisc' and 'metafor' in the R statistical language (Debray 2014; Debray 2018b; Viechtbauer 2010).
- We pooled only validation studies with the outcome that a model was developed for (e.g. we did not meta-analyse validation studies that tested a model to predict TFS if the model was developed for OS).
- We used random-effects meta-analysis, because validation studies typically differ regarding several parameters, including patient characteristics and design.
- Where the c-statistic was not reported, we did not estimate it. When the c-statistic was provided without measures of uncertainties, we calculated the standard error and confidence intervals based on the P value or the combination of sample size and number of events, if available, according to Newcombe and colleagues (Debray 2017; Debray 2018a; Newcombe 2006).

We planned to summarise the measures of calibration. Unfortunately, calibration measures and the expected survival, both in development and external validation cohorts, were rarely reported. Instead, many studies reported the observed outcome frequency per subgroup at a specific time point. Though it is possible to estimate approximate O:E ratios from the reported observed survival per risk category and probabilities reported in the model development study (for scores), in collaboration with the Cochrane Prognosis Methods Group, we decided against this estimation because it is merely an approximate indication for the calibration of a model. For time-to-event outcomes, we would need to account for censoring (Debray 2018a), which was not possible based on the limited reporting of the studies. Instead, we illustrated the survival per risk category for each model with sufficient data graphically. We pooled survival per risk category for all external validations of one model and reported a prediction interval. The 95% prediction interval (95% PI) is an estimate of the range in which a future average survival frequency in a new validation study of the prognostic model will fall with a 95% probability.

Subgroup analysis

We planned to investigate whether the change of definition of CLL over time has affected the performance of a model (i.e. Cheson 1996; Hallek 2008). We did not perform this analysis because the number of studies that reported a clear distinction between the definitions was too low.

Sensitivity analysis

During the review process, and before data extraction and analysis, we decided to conduct sensitivity analysis based on the 'Risk of bias' rating, if a sufficient number of validation studies were available per model.

We decided post hoc to conduct the following model-specific meta-analyses.

- Test the effect of studies that reported the area under the curve (AUC) instead of the c-statistic (which we meta-analysed together). These two measures for discrimination are the same for binary outcomes, but may differ for time-to-event data depending on the time point chosen to calculate the AUC.
- Explore the effect of the estimation of the 95% CI for the c-statistic. Estimating confidence intervals can introduce imprecision, therefore we aimed to explore the extend of the changes between reported and estimated 95% CIs.
- For the CLL-IPI, several external validation studies used a proxy prognostic factor for TP53 deletion or mutation, the cytogenetic aberration based on FISH, del(17p). Although we expected the effect to be small as the concordance between the two

predictors is larger than 90% (Dicker 2009), using a proxy can reduce predictive performance.

Rating the certainty of evidence and summary of findings

Originally, we planned to use the most up-to-date GRADE guidance for this systematic review. However, no GRADE guidance for grading the certainty of results from meta-analysis of prognostic models yet exists. Hence, for this review, we decided to refrain from applying GRADE.

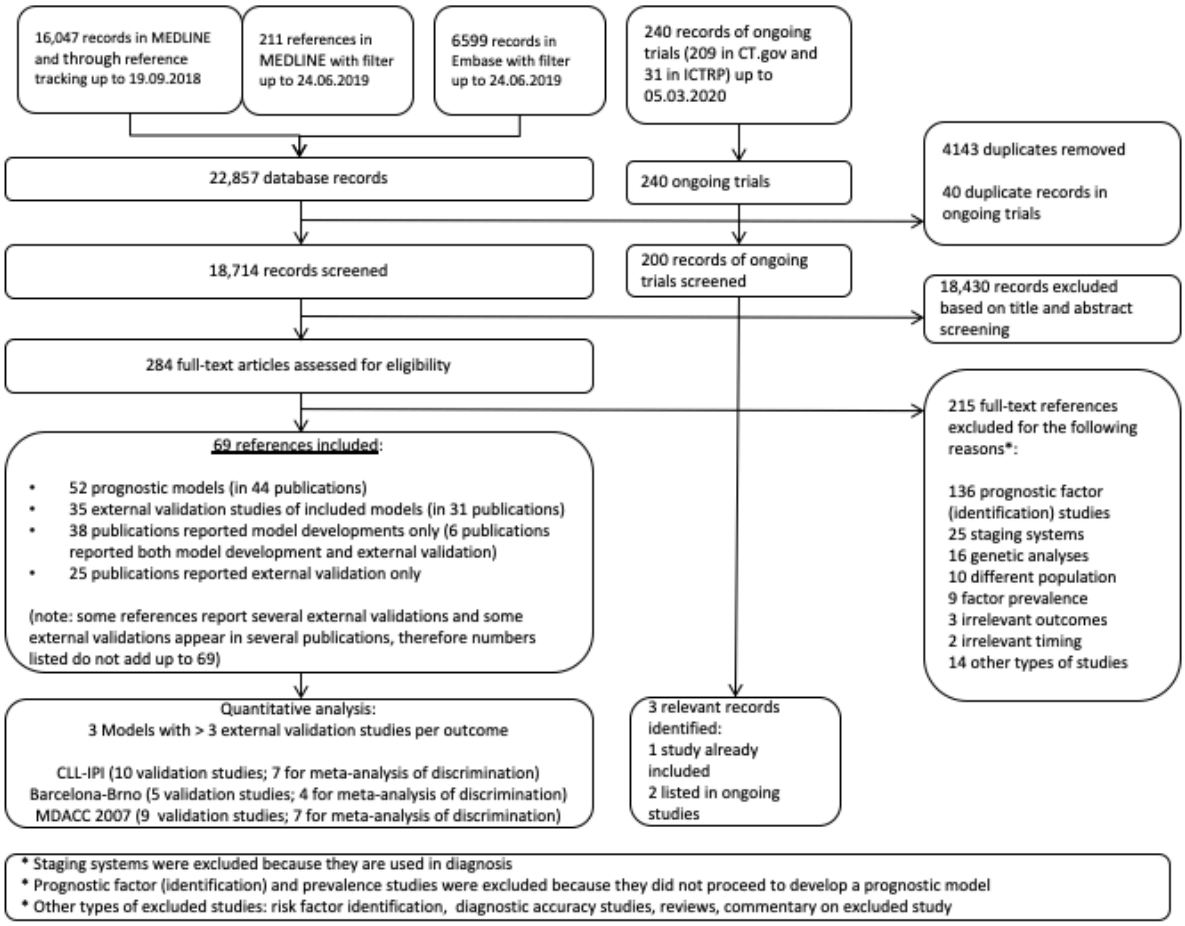
RESULTS

Description of studies

Results of the search

Our literature search (24 June 2019; MEDLINE and Embase) resulted in 22,857 potentially relevant references related to prognostic models for newly-diagnosed untreated individuals with chronic lymphocytic leukaemia (CLL). Of these, we removed 4144 duplicates, leaving 18,713 references for title and abstract screening. We identified 284 full-text references, which might fulfil our predefined inclusion criteria, including two references identified by reference and citation tracking. Of these, we excluded 214 references. We documented the overall number of screened, included and excluded references in a PRISMA flow diagram (Figure 1).

Figure 1. Study flow diagram



This led to a total of 70 included references. Of these, 38 reported model developments only, six publications reported both model development and external validation(s), and 25 publications reported external validation only. In total, we identified 52 prognostic models (in 44 publications), and 35 external validation studies of included models that predicted the primary outcome of one of the included developed models (in 31 publications). Several of these publications offered secondary analyses with another outcome. These are only listed in this review if the primary outcome was not reported.

Out of the 52 developed models, 12 models were externally validated at least once either in the primary publication or as described in an additional publication (for an overview, see Appendix 6). Forty-one models did not undergo any external validation.

The search of databases for ongoing trials resulted in 240 records (ClinicalTrials.gov: 209 and ICTRP: 31; 5 March 2020). Of these, 40 records were duplicates. We identified three records that referred to the potential use of developing a prognostic model. One cohort had already been included from the database search (O-CLL-GISL, development of a model and several validations, e.g. O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). One study is currently recruiting (NCT03436524), and one mentioned the aim to develop a staging system, however, it does not refer to any relevant publication yet (NCT00275054).

It is important to note the difference between 'reference' and 'study': one reference of a model development study may in addition to the model development also report one or several external validation studies of their model in independent cohorts. Furthermore, some references of external validations may report and compare the performances of several prognostic models in their cohort. Several publications used the same dataset to validate the same model(s).

References of model development studies without any external validation studies are summarised in Studies awaiting classification and Appendix 7.

Included studies

We described the characteristics of included studies per model and per predicted outcome. The numbers of external validations belonging to each developed model and outcome may not correspond to the number of included references. We defined a study as a validation of one model in one independent cohort, which implies that the validation of two different models in the same cohort are counted as two separate studies for the description in this review. In some publications, one model was externally validated in several cohorts, while other publications validated several models in one cohort. We presented the number of participants included in analysis instead of the number of participants in the study sample or complete registry as, often, missing values were reported to be an exclusion reason. We only included external validation studies that predicted the same outcome that the prognostic model was developed for.

Models with meta-analysis (> 3 external validations)

- CLL-IPI (OS, CLL-IPI D - Bahlo 2016 (development cohort))
- Barcelona-Brno model (OS, Barcelona-Brno D - Delgado 2017 (Barcelona cohort))
- MDACC 2007 (OS, MDACC 2007 D - Wierda 2007 (MDACC))

Models without meta-analysis (1-3 external validations)

- GCLLSG model (OS, GCLLSG D - Pflug 2014 (GCLLSG))
- Rossi model (OS, Rossi D - Rossi 2013 (Italian cohort))
- Stephens (OS, Stephens OS D - Stephens 2015 (Ohio cohort))
- Baliakas model (TFS, Baliakas D - Baliakas 2019 (multicentre))
- GIMEMA model (PFS, GIMEMA D - Molica 2005 (GIMEMA cohort))
- MDACC 2011 (TFS, MDACC 2007 D - Wierda 2007 (MDACC))
- Morabito model (TFS, Morabito D - Morabito 2009 (Italian cohort))
- O-CLL1 model (TFS, O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))
- Stephens model (TFS, Stephens TFS D - Stephens 2015 (Ohio cohort))

Models with more than three external validation studies per outcome

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI *(CLL-IPI D - Bahlo 2016 (development cohort))*

The CLL-IPI model predicts OS and was derived from combined data from eight phase three RCTs conducted in multiple countries (France, Germany, Poland, UK, USA), with a total of 1799 individuals included for the model development. The recruitment period for the trials lasted from 1997 to 2009 (CLL-IPI D - Bahlo 2016 (development cohort)). The median follow-up time was 79.9 months (interquartile range (IQR) 79.9 to 101.4 months). All disease stages were included; most individuals were classified as Rai I/II or Binet B stage (50% and 41% respectively). The final model includes five predictors: TP53 status, IgHV mutational status, serum B2-microglobulin, clinical stage (Rai or Binet), and age. The original weighting of predictors was simplified to present as a risk score for easier use in clinical practice. We identified 11 external validation cohorts (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort); CLL-IPI V - Gentile 2016 (Italian cohort); CLL-IPI V - Molica 2016 (O-CLL1-GISL); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort); CLL-IPI V - Rani 2018 (Indian cohort); CLL-IPI V - Reda 2017 (Milano cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort); CLL-IPI V - Zhu 2018 (Chinese cohort)), with 10 validations for OS and eight for TFS. The validation studies for OS included a total of 5485 individuals with available data presenting between 1983 and 2016. Two relevant validation studies lacked information regarding the recruitment period (CLL-IPI V - Delgado 2017 (Barcelona cohort); CLL-IPI V - Rani 2018 (Indian cohort)). Due to lack of reporting, only 3307 individuals with 917 deaths from seven external validation studies were included in the meta-analysis of the discrimination (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort); CLL-IPI V - Gentile 2016 (Italian cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort); CLL-IPI V - Rani 2018 (Indian cohort); CLL-IPI V - Zhu 2018 (Chinese cohort)). We provide a summary of the main characteristics of studies regarding the CLL-IPI in Figure 2.

Figure 2. CLL international prognostic index - summary of characteristics of included studies

Characteristics of included studies for the CLL-IPI externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
CLL-IPI model (overall survival)	Development	Bahlo 2016 (Development cohort)	1997 - 2009	France, Germany, Poland, UK, USA, etc.	462 events 1214 persons (train) 243 events 585 persons (test)	median 61 years (range: 27 - 86 years)	70% male	Rai 0: 14% Rai I/II: 50% Rai III/IV: 36%	690 F 1017 FC 113 CC 475 FCR 81 FCCAM 479 CHR	combined RCT data	median 79.9 months, IQR 79.9 - 101.4
	Meta-analysed c-statistic	Bahlo 2016 (Mayo clinic 2001-2014)	2001 - 2014	USA	144 events 838 persons	≤ 65 years: 533 persons > 65 years: 305 persons	67.8% male	Rai 0: 57.4% Rai I-II: 38.5% Rai III-IV: 4.1%	168 purine analogs w/o MoAbs 37 Alkylator-based Chemoimmunotherapy 11 Alkylator w/o MoAbs 97 antibody only 7 RTK inhibitors 6 other	prospective cohort	median 63.2 months, IQR 30.2 - 91.8
		Bahlo 2016 (SCAN cohort)	1999 - 2002	Denmark, Sweden	215 events 416 persons	≤ 65 years: 242 persons > 65 years: 174 persons	62.7% male	Binet A: 78.1% Binet B: 16.6% Binet C: 5.3%	NR	unclear	median 151 months, IQR 124.8 - 163.0
		Delgado 2017 (Barcelona cohort)	NR	Spain	212 events 524 persons	median 62 years (range: 22 - 93 years)	60% male	Rai 0: 62%; Rai I-IV: 38%	83 FCR or similar 82 purine alanogs w/o MoAbs 86 received alkylating agents 23 received others	retrospective cohort	median 99.6 months (range 1 - 456 months)
		Gentile 2016 (Italian cohort)	1985 - 2015	Italy	174 events 858 persons	median 65.5 years (range: not reported)	56.2% male	Rai 0: 58.4% Rai I: 19.8% Rai II: 15.2% Rai III: 1.5% Rai IV: 5.1%	130 received chemotherapy 174 received chemo-immunotherapy	retrospective cohort	median 69.6 months, range not reported
		Muñoz-Novaz 2018 (Spanish cohort)	1989 - 2013	Spain	47 events 258 persons	median 65.7 years (IQR: 55.2 - 73.5)	62.8% male	Rai 0: 58.9%; Rai I-IV: 41.1%	NR	retrospective cohort	median 68 months (range 3 - 277 months)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
		Zhu 2018 (Chinese cohort)	2002 - 2017	China	46 events 215 persons	median 60 years (range 16 - 85 years)	66.5% male	Rai 0: 16.7% Rai I-II: 54.0% Rai III-IV: 29.3%	147 treated with benzodiazepine FCR bendamoxetine R R+FFP+HDMP	retrospective cohort	median 48 months, (range: 1 - 192 months)
	Narrative	Da Cunha-Bang 2016 (Danish cohort)	2008 - 2015	Denmark	249 events 1514 persons	median 69 years (range: not reported)	60% male	Binet A: 80% Binet B/C: 20%	NR	prospective cohort	median 38.4 months, range not reported
		Molica 2016 (O-CLL1-GISL)	2007 - NR	Italy	91 events (TFS) 337 persons	median 61 years (range: 33 - 70 years)	57.2% male	Rai 0: 77.8% Rai ≥ I: 22.2%	NR	prospective cohort	median 42 months, 2038 person-years
		Reda 2017 (Milan cohort)	1983 - 2016	Italy	NR events 329 persons	median 65 years (range: 32 - 91 years)	57% male	Rai 0-I: 74%; Rai II-IV: NR	190 CHL 108 FCR 12 O-Benda 44 BR 65 alemtuzumab 29 Ibrutinib or Idelalisib	retrospective cohort	median 144 months (range 0 - 360 months)
		Rigolin 2017 (Ferrera cohort)	2006 - 2016	Italy	NR events 335 persons	median 68.7 years (range: 33 - 96 years)	57.9% male	Binet A: 77.9% Binet B: 14.3% Binet C: 7.8%	114 chemotherapy	retrospective cohort	NR

Abbreviations: NR: not reported; CHL: chlorambucil; F: fluradabine; CC: cyclophosphamide + cladribine; FC: fluradabine + cyclophosphamide; FCR: fluradabine + cyclophosphamide + rituximab; FCCAM: FC + alemtuzumab; MoAbs: monoclonal antibody; R: rituximab; FFP: fresh frozen plasma; HDMP: high-dose methylprednisolone

Barcelona-Brno model *(*Barcelona-Brno D - Delgado 2017 (Barcelona cohort)*)*

The Barcelona-Brno model predicts OS and was derived from data from a retrospective, single-centre cohort study conducted in Spain (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)). TFS was reported as a secondary outcome. Model development included 524 individuals. The recruitment period was not reported. The median follow-up time was 99.6 months, ranging from 1 to 456 months. All disease stages were included; most individuals were Rai 0 or Binet A stage (62% and 83% respectively). The model aimed to simplify the CLL-IPI described below, therefore only combinations of the factors included in the CLL-IPI were tested. The final model included three predictors: IgHV mutational status, and the genomic aberrations del(17p), and del(11q). We identified six external validation cohorts (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Molica 2017 (O-CLL1-GISL); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.); Barcelona-Brno V - Rani 2018 (Indian cohort); Barcelona-Brno V - Reda 2017 (Milan cohort)), with five validations for OS and six for TFS. The validation studies for OS included a total of 2501 individuals presenting between 1983 and 2016. Two validation studies lacked information regarding the recruitment period (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Rani 2018 (Indian cohort)). Due to lack of reporting, only 1755 individuals with 416 deaths from three studies were included in meta-analysis of discrimination (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.); Barcelona-Brno V - Rani 2018 (Indian cohort)). We provide a summary of the main characteristics of studies regarding the Barcelona-Brno model in Figure 3.

Figure 3. Barcelona-Brno score - summary of characteristics of included studies

Characteristics of included studies for the Barcelona-Brno externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Barcelona cohort)	NR	Spain	212 events 524 persons	median 62 years (range: 22 - 93 years)	60% male	Rai 0: 62%; Rai I-IV: 38%	83 FCR or similar 82 purine alanogs w/o MoAbs 86 received alkylating agents 23 received others	retrospective cohort	median 99.6 months (range 1 - 456 months)
	Meta-analysed c-statistic	Delgado 2017 (Brno cohort, CZ)	NR	Czech Republic	158 events 417 persons	median 62 years (range: 32 - 85 years)	66% male	Rai 0: 41%; Rai I-II: 43%; Rai III-IV: 16%	125 FCR or similar 41 purine analogs w/o MoAbs 53 alkylating agents 57 received others	NR	median 59 months (range 3 - 330 months)
		Gentile 2017 (Italian multicentre and Mayo cohort)	1985 - 2015	Italy & USA	283 events 1299 persons	median 63 years (range: 27 - 92 years)	61.3% male	Rai 0: 57.9%, Rai I: 28.6% Rai II: 8.6%, Rai III: 1.5%; Rai IV: 3.5%	194 chemotherapy 316 chemoimmunotherapy	retrospective cohort	median 82 months (range 3 - 330 months)
		Muñoz-Novaz 2018 (Spanish cohort)	1989 - 2013	Spain	47 events 258 persons	median 65.7 years (IQR: 55.2 - 73.5)	62.8% male	Rai 0: 58.9%; Rai I-IV: 41.1%	137 received therapy, unclear which therapy	retrospective cohort	median 68 months (range 3 - 277 months)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
	Narrative	Molica 2017 (O-CLL1-GISL)	2007 - NR	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	57.2% male	Rai 0: 77.8%; Rai I-II: 22.2%	NR	prospective cohort	median 42 months (range 1 - 82 months)
		Reda 2017 (Milan cohort)	1983 - 2016	Italy	NR events 329 persons	median 65 years (range: 32 - 91 years)	57% male	Rai 0-I: 74%; Rai II-IV: NR	190 CHL 108 FCR 12 O-Benda 44 BR 65 alemtuzumab 29 Ibrutinib or Idelalisib	retrospective cohort	median 144 months (range 0 - 360 months)

Abbreviations: NR: not reported; CHL: chlorambucil; O-Benda: ofatumumabbendamustine; BR: bendamustine-rituximab; MoAbs: monoclonal antibodies

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

The MDACC 2007 model predicts OS and was derived from data from a retrospective, single-centre cohort study conducted in the USA (MDACC 2007 D - Wierda 2007 (MDACC)). A total of 1674 individuals were included for the model development. The recruitment period for the study lasted from 1981 to 2004. The median follow-up time was 58.8 months (95% confidence interval (CI) 55.2 to 61.2 months). All disease stages were included; most individuals were classified as Rai I or Binet A stage (739/1674 and 1019/1674 respectively). The final model included six predictors: age, absolute lymphocyte count, serum B2-microglobulin, nodal groups, Rai stage, and gender. The publication reports both the formula to be used in combination with a nomogram, and a simplified index score. We identified 10 external cohorts (MDACC 2007 V - Trajkova 2013 (Macedonia); MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - González Rodríguez (Cabueñes); MDACC 2007 V - Molica 2010 (GIMEMA cohort); MDACC 2007 V - Molica 2015 (O-CLL1-GISL); MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort); MDACC 2007 V - Pflug 2014 (3 RCTs); MDACC 2007 V - Rani 2018 (Indian cohort)), with eight validations for OS and six for TFS. Seven studies used the index score and three additionally calculated the formula to predict OS. For one study, it was unclear if the index score or the formula were used. The eight cohorts for which OS was predicted included 6928 individuals presenting between 1983 and 2013. One validation study lacked information regarding the recruitment period (MDACC 2007 V - Rani 2018 (Indian cohort)). Only the simplified index score was sufficiently externally validated to be summarised in meta-analysis. In total, 5127 individuals with 994 deaths from seven studies remained for inclusion in the meta-analysis of discrimination (MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - González Rodríguez (Cabueñes); MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort); MDACC 2007 V - Pflug 2014 (3 RCTs); MDACC 2007 V - Rani 2018 (Indian cohort)). We provide a summary of the main characteristics of studies regarding the MDACC 2007 model in Figure 4.

Figure 4. MDACC 2007 index score - summary of characteristics of included studies

Characteristics of included studies for the MDACC 2007 externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number persons	age	gender	stage	treatment	design	follow-up
MDACC 2007 model (overall survival)	Development	Wierda 2007 (Development cohort)	1981 - 2004	USA	437 events 1617 persons	median 58 years (range: 0 - 90 years)	1029 male	Rai 0: 469 Rai I: 739 Rai II: 235 Rai III: 93 Rai IV: 127	NR	retrospective cohort	median 58.8 months, 95% CI: 55.2 - 61.2
	Meta-analysed c-statistic for the index score	Bulian (Italian-Swiss cohort)	1996 - 2011	Italy, Switzerland	151 events 1037 persons	median 65 years (range: 21 - 94 years)	58.2% male	Binet A: 74.6% Binet B: 18.4% Binet C: 6.9%	306 chemotherapy 129 chemoimmunotherapy 42 FCR 49 FR 38 R-other 40 with missing data	retrospective cohort	median 63.6 months, range not reported
		Gentile 2014 (Italian multicentre cohort)	1983 - 2013	Italy	277 events 1502 persons	median 67 years (range: not reported)	55.7% male	Rai 0: 56.5% Rai I: 21.4% Rai II: 15% Rai III: 2.6% Rai IV: 4.5%	337 chemotherapy 142 chemo-immunotherapy	retrospective cohort	median 68.4 months, range not reported
		Gentile 2016 (Mayo cohort 2001-2008)	2001 - 2008	USA	114 events 506 persons	median 62.5 years (range: 36 - 89 years)	68.6% male	Rai 0: 62.5% Rai I: 28.9% Rai II: 5.1% Rai III: 1.2% Rai IV: 2.4%	NR	prospective cohort	median 86.4 months, range not reported
		González Rodríguez 2009 (Cabueñes hospital cohort)	1997 - 2007	Spain	76 events 257 persons	median 71.7 years (range: 42 - 94 years)	NR	Binet A: 76.8% Binet B: 14.8% Binet C: 8.4%	NR	retrospective cohort	NR
		Muñoz-Novaz 2018 (Spain multicentre cohort)	1989 - 2013	Spain	92 events 483 persons	median 67 years (range: 25 - 90 years)	64.2 % male	Rai 0-II: 93.8% Rai III-IV: 6.2%	NR	retrospective cohort	median 46 months, range not reported
		Pflug 2014 (German CLL study group cohort)	1997 - 2006	Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain	180 events 1144 persons	median 60 years (range: 30 - 81 years)	68.4% male	Rai 0: 20.8% Rai I: 18.4% Rai II: 37.9% Rai III: 8.6% Rai IV: 14.3%	early F FC FCR	combined RCT data	median 63.4 months, range not reported
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
	Narrative	Molica 2010 (GIMEMA cohort 1991-2000)	1991 - 2000	Italy	NR events 310 persons	median 64 years (range: not reported)	56.1% male	Rai 0: 67% Rai I-III: 32.9% Rai III/IV: 0%	NR	retrospective cohort	median 41.5 months, range not reported
		Molica 2015 (O-CLL1-GISL)	2007 - NR	Italy	91 events (TFS) 337 persons	median 61 years (range: 33 - 70 years)	57.1% male	Rai 0: 77.5% Rai I-II: 22.5%	NR	prospective cohort	median 42 months, range not reported
		Trajkova 2013 (Macedonian cohort)	2011 - 2013	Macedonia		median 64.8 years (range: 47 - 78 years)	63% male	Rai 0: 54% Rai I: 5% Rai II: 32% Rai III: 5% Rai IV: 4%	NR	unclear	NR

Abbreviations: NR: not reported; F: fludarabine; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabine + rituximab; R-other: rituximab with other regimens or rituximab monotherapy; FC: fludarabine + cyclophosphamide; CHL: chlorambucil

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS or TFS.

Models with one to three external validation studies per outcome

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

The GCLLSG model predicts OS and was derived from combined data from three RCTs conducted in multiple countries (Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand and Spain), with a total of 2007 individuals included for the model development (GCLLSG D - Pflug 2014 (GCLLSG)). The recruitment period for the trials lasted from 1997 to 2006. The median follow-up time was 63.4 months; ranges were not reported. All disease stages were included; most individuals were Rai II or Binet A stage (37.9% and 42.6%, respectively). The final model included eight predictors: gender, age, Eastern Co-operative Oncology Group Performance Status (ECOG PS), del17p, del11q, IgHV mutational status, serum thymidine kinase, and serum B2-microglobulin. We identified three external validation cohorts, with one validation for OS (676 individuals, GCLLSG V - Pflug 2014 (Mayo cohort)), and two for TFS (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)). All validation studies lacked information regarding the recruitment period. We provide a summary of the main characteristics of studies regarding the GCLLSG model in Figure 5.

Figure 5. GCLLSG model - summary of characteristics of included studies

Characteristics of included studies for the GCLLSG externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
GCLLSG model (overall survival)	Development	Pflug 2014 (Development cohort)	1997 - 2006	Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain	180 events 1144 persons	median 60 years (range: 30 - 81 years)	68.4% male	Rai 0: 20.8% Rai I: 18.4% Rai II: 37.9% Rai III: 8.6% Rai IV: 14.3%	early F FC FCR	combined RCT data	median 63.4 months, range not reported
	Narrative	Molica 2015 (O-CLL1-GISL)	2007 - NR	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	57.1% male	Rai 0: 77.8%; Rai I-IV: NR	NR	prospective cohort	median 42 months (range 1 - 82 months; 2038 person-years)
		Pflug 2014 (Mayo cohort)	NR	USA	190 events 676 persons	median 61.5 years (range: 32 - 89 years)	67% male	Rai 0: 57.1% Rai I: 34% Rai II: 5.9% Rai III: 1% Rai VI: 1.9%	NR	prospective cohort	median 57 months (range not reported)
		Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)

Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide-rituximab; CHL: chlorambucil

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

The Rossi model predicts OS and was derived from data from a retrospective, multicentre cohort study conducted in Italy (Rossi D - Rossi 2013 (Italian cohort)). A total of 637 individuals were included for the model development. The recruitment period for the study lasted from 1996 to 2011. The median follow-up time was 67.2 months; ranges were not reported. All disease stages were included; most individuals were classified as Rai 0 stage (74.9%). The final model included five predictors: TP53, BIRC3 DIS, SF3B1 M, NOTCH1 M, and del11q22-q23. We identified two external validations (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)). One validation study included a total of 1160 individuals presenting between 2005 and 2010 (Rossi V - Jeromin 2014 (Munich cohort)). The second validation study was conducted between 1996 and 2011; the authors failed to report the number of participants included in the validation cohort (Rossi V - Rossi 2013 (unclear)). We provide a summary of the main characteristics of studies regarding the Rossi model in Figure 6.

Figure 6. Rossi model - summary of characteristics of included studies

Characteristics of included studies for the Rossi externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Rossi model (overall survival)	Development	Rossi 2013 (Development cohort)	1996 - 2011	Italy	178 events 583 persons	51.2% >= 65 years	58.4% male	Rai 0: 74.9% Rai I-II: 11.5%	122 FCR, FR, or pentostatin + CR 64 F or FC 80 alkylator-based (e.g. CHL)	retrospective cohort	median 67.2 months (range not reported)
	Narrative	Jeromin 2014 (Munich cohort)	2005 - 2010	Unclear	NR events 930 persons	median 67 years (range: 29.6 - 90.5 years)	64.6% male	NR	NR	prospective cohort	median 55.2 months, (range not reported)
		Rossi 2013 (External validation cohort)	1996 - 2011	Italy	62 events 370 persons	NR	55.9% male	Rai 0-I: 75.4% Rai II: 17.3% Rai III-IV: 7.3%	NR	retrospective cohort	median 70.8 months, (range not reported)

Abbreviations: NR: not reported; F: fludarabine; FC: fludarabine + cyclophosphamide; FCR: fludarabine + cyclophosphamide + rituximab; FR: fludarabinerituximab, CR: cyclophosphamide + rituximab; CHL: chlorambucil

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

Stephens and colleagues developed models for two outcomes, OS and TFS (Stephens OS D - Stephens 2015 (Ohio cohort); Stephens TFS D - Stephens 2015 (Ohio cohort)). They were derived from data from a retrospective, single-centre cohort study conducted in the USA. A total of 114 individuals were included for the model development. The recruitment period for the study lasted from 2002 to 2012. The follow-up time was not reported. All disease stages were included; most individuals were classified as Rai I/II stage (46%). The final models for both outcomes were simplified for use in clinical practice. The simplified risk score for OS includes three predictors: ECOG PS, age, and lactate dehydrogenase. The simplified risk score for TFS included five predictors: ECOG PS, Rai stage, age, white blood cell count, and del11q22.3. We identified one external validation study (Stephens OS V - Stephens 2015 (MDACC)), which included 129 individuals. The recruitment period of the validation study was not reported. We provide a summary of the main characteristics of studies regarding the Stephens model in Figure 7.

Figure 7. Stephens model for OS - summary of characteristics of included studies

Characteristics of included studies for the Stephens externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Stephens model (overall survival)	Development	Stephens 2015 (Development cohort)	2002 - 2012	USA	NR events 114 persons	median 62 years (range: 40 - 92 years)	NR	Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21%	35 purine analogs 18 treatment on a clinical trial 5 other treatment	retrospective cohort	NR
	Narrative	Stephens 2015 (MDACC cohort)	NR	USA	NR events 129 persons	median 63 years (range: 40 - 85 years)	NR	Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19%	NR	retrospective cohort	NR
Abbreviations: NR: not reported											

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

The Baliakas model predicts TFS and was derived from data from a retrospective, multicentre cohort study conducted in 10 European institutions (Baliakas D - Baliakas 2019 (multicentre)). A total of 1900 individuals were included for the development of two separate models, dividing the study population into mutated (M-CLL) and unmutated CLL (U-CLL) patients. The recruitment period for the study was not reported. The median follow-up time was 7.1 years, ranging from 0.1 to 33.1 years. Only individuals with disease stage Binet A were included. The final model for M-CLL included three predictors: TP53 abnormality, trisomy 12, and stereotyped subset #2 defined as IGHV3-21/IGLV3-21 BcR IG. The final model for U-CLL included three predictors: TP53 abnormality, del11q, and gender. We identified one external validation study (Baliakas V - Baliakas 2019 (MLL + Scan.)), which included a total of 649 persons from two separate studies, one conducted at the Munich Leukemia Laboratory (508 persons) and one Scandinavian population-based study (141 persons). We provide a summary of the main characteristics of studies regarding the Baliakas model in Figure 8.

Figure 8. Baliakas model - summary of characteristics of included studies

Characteristics of included studies for the Baliakas externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Baliakas model (time-to-first-treatment)	Development	Baliakas 2019 (Development cohort)	NR	multiple European countries	NR events 1900 persons	median 64.3 years (range: 22 - 92 years)	61% male	Binet A: 80% Binet B: 12% Binet C: 8%	NR	retrospective cohort	median 7.1 years (range 0.1 - 33.1 years)
	Narrative	Baliakas 2019 (MLL + Scan. cohort)	NR	Germany, Scandinavia	NR events 649 persons	median 63.6 years (range: 29 - 89 years)	62% male	Binet A: 100%	NR	retrospective cohort	NR
Abbreviations: NR: not reported											

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

The GIMEMA model predicts PFS and was derived from data from a retrospective, multicentre cohort study conducted in Italy (GIMEMA D - Molica 2005 (GIMEMA cohort)). A total of 1138 individuals were included for the model development. The recruitment period for the study lasted from 1991 to 2000. The median follow-up time was 54 months, ranging from 4 to 309 months. Only individuals with disease stage Binet A were included. The final model included four predictors: lymphocyte doubling time, absolute peripheral blood lymphocytosis, Rai stage, and gender. We identified one external validation study (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)), which included 265 persons presenting between 1997 and 2007. We provide a summary of the main characteristics of studies regarding the GIMEMA model in Figure 9.

Figure 9. GIMEMA model - summary of characteristics of included studies

Characteristics of included studies for the GIMEMA externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
GIMEMA model (progression-free survival)	Development	Molica 2005 (Development cohort)	1991 - 2000	Italy	235 events 593 persons	median 65 years (range: 27 - 100 years)	54% male	Rai 0: 77.2 % Rai I-III: 22.7%	133 CHL 20 F or FC 7 CHOP-like regimen 40 other (CVP, C +/-prednisone)	retrospective cohort	median 54 months, range 4 - 309 months
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	1997 - 2007	Spain	76 events 257 persons	median 71.7 years (range: 42 - 94 years)	NR	Binet A: 76.8% Binet B: 14.8% Binet C: 8.4%	NR	retrospective cohort	NR
Abbreviations: NR: not reported; CHL: chlorambucil; F: fludarabine; FC: fludarabine + cyclophosphamide; CHOP: doxorubicin + cyclophosphamide + vincristine + prednisone; CVP: cyclophosphamide + vincristine sulfate + prednisone											

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

The MDACC 2011 model predicts TFS and was derived from data from a retrospective, single-centre cohort study conducted in the USA (MDACC 2011 D - Wierda 2011 (MDACC)). A total of 930 individuals were included for the model development. The recruitment period for the study lasted from 2004 to 2009. The median follow-up time was 26 months, ranging from three to 73 months. All disease stages were included; most individuals were classified as Rai I stage (51%). The final model included six predictors: IgHV mutational status, diameter of largest palpated lymph node, FISH category (del11q or del17p versus none), number of involved lymph node sites, lactate dehydrogenase (LDH), and the IgHV mutational status and LDH interaction term. We identified one external validation study (MDACC 2011 V - Molica 2016 (O-CLL1-

GISL)), which included 328 persons presenting between 2006 and 2010. We provide a summary of the main characteristics of studies regarding the MDACC 2011 model in Figure 10.

Figure 10. MDACC 2011 model - summary of characteristics of included studies

Characteristics of included studies for the MDACC 2011 externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
MDACC 2011 model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	2004 - 2009	USA	193 events 687 persons	median 59 years (range: 30 - 89 years)	61% male	Rai 0: 36% Rai I: 51% Rai II: 8% Rai III-IV: 5%	NR	retrospective cohort	median 26 months (range 3 - 73 months)
	Narrative	Molica 2016 (O-CLL1-GISL)	2006 - 2010	Italy	91 events 337 persons	median 61 years (range: 33 - 70 years)	58.2% male	Rai 0: 76.5% Rai I-II: 23.5%	NR	prospective cohort	median 30 months (range 1 - 65 months)
Abbreviations: NR: not reported											

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

The Morabito model predicts TFS and was derived from data from a multicentre study conducted in Italy (Morabito D - Morabito 2009 (Italian cohort)). The study design was not reported. A total of 262 individuals were included for the model development. The recruitment period was not reported. The median follow-up time was 36 months, ranging from 12 to 180 months. Only individuals with disease stage Binet A were included. The final model included three predictors: IgHV mutational status, CD38, and ZAP-70. We identified one external validation study (Morabito V - Gentile 2014 (O-CLL1-GISL)), which included 480 persons presenting from 2007. There was no information regarding the end of the recruitment period. We provide a summary of the main characteristics of studies regarding the Morabito model in Figure 11.

Figure 11. Morabito model - summary of characteristics of included studies

Characteristics of included studies for the Morabito externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Morabito model (time-to-first-treatment)	Development	Morabito 2009 (Development cohort)	NR	Italy	51 events 252 persons	median 65 years (range not reported)	61.5% male	Binet A: 100%	NR	NR	median 36 months (range 12 - 180 months)
	Narrative	Gentile 2014 (O-CLL1-GISL)	2007 - NR	Italy	121 events 468 persons	47.4 % < 60 years	58.5% male	Rai 0: 77.3% Rai I-II: 22.7%	NR	prospective cohort	median 38.5 months (range 6 - 82 months)
Abbreviations: NR: not reported											

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

The O-CLL1 model predicts TFS and was derived from data from a prospective, multicentre cohort study conducted in Italy (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). A total of 480 individuals were included for the model development. Recruitment for the study started in 2007; the end date was not reported. The median follow-up time was 42 months, ranging from 6 to 82 months. Only individuals with disease stage Binet A were included. The final model included four predictors: Rai stage, absolute lymphocyte count (ALC), serum B2-microglobulin, and IgHV mutational status. We identified two external validation cohorts for TFS (O-CLL1 V - Gentile 2016 (Mayo cohort); O-CLL-1 V - Rani 2018 (Indian cohort)), which included a total of 626 persons. One validation study recruited participants between 2001 and 2008 (O-CLL1 V - Gentile 2016 (Mayo cohort)); the other study lacked information regarding the recruitment period (O-CLL-1 V - Rani 2018 (Indian cohort)). We provide a summary of the main characteristics of studies regarding the O-CLL1 model in Figure 12.

Figure 12. O-CLL-1 model - summary of characteristics of included studies

Characteristics of included studies for the O-CLL1 externally validated model											
Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
O-CLL1 model (time-to-first-treatment)	Development	Gentile 2016 (O-CLL1-GISL)	2007 - NR	Italy	84 events 337 persons	46.7% <= 60 years	57.3% male	Rai 0: 77.7% Rai I-II: 22.3%	NR	prospective cohort	median 42 months (range 6 - 82 months)
	Narrative	Rani 2018 (Indian cohort)	NR	India	86 events 198 persons	median 60 years (range: not reported)	77% male	Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%	62 CHL 56 rituximab-based therapy 20 other therapies	NR	median 40.5 months (range 1 - 215 months)
		Gentile 2016 (Mayo cohort 2001-2008)	2001 - 2008	USA	130 events 428 persons	40.2% <= 60 years	66.4% male	Rai 0: 67.5% Rai I-II: 32.5%	NR	NR	median 97 months (range not reported)
Abbreviations: NR: not reported; CHL: chlorambucil											

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

For a description of the model for TFS in Stephens and colleagues (Stephens TFS D - Stephens 2015 (Ohio cohort)), please see description in section OS (above), as the authors used the same development and validation cohorts for both outcomes. We provide a summary of the main characteristics of studies regarding the Stephens model in Figure 13.

Figure 13. Stephens model for TFS - summary of characteristics of included studies

Characteristics of included studies for the Stephens externally validated model

Model		Study	Characteristics of included studies								
			recruitment	country	number of events / number	age	gender	stage	treatment	design	follow-up
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	2002 - 2012	USA	58 events 114 persons	median 62 years (range: 40 - 92 years)	NR	Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21%	35 purine analogs 18 treatment on a clinical trial 5 other treatment	retrospective cohort	NR
	Narrative	Stephens 2015 (MDACC cohort)	NR	USA	NR events 129 persons	median 63 years (range: 40 - 85 years)	NR	Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19%	NR	retrospective cohort	NR

Abbreviations: NR: not reported

Models without external validation studies

The 34 references with 41 model or score developments can be found in Studies awaiting classification, as it is widely recommended that developed prediction models should not be used in daily practice before they were validated at least once (Moons 2009; Moons 2015; Moons 2019; Steyerberg 2013). We have provided an overview of the main characteristics of these developed (but never validated) models in Appendix 7. In total, 19 models were developed to predict OS, 19 to predict TFS and three to predict PFS. These models were published between 1982 and 2019 and contained an average number of 3.8 predictors (range between 2 and 6). Nineteen models were derived from retrospective cohorts, three were on prospective cohorts and 19 did not report the study design clearly. The most commonly included predictors were IgHV status (mutated versus unmutated), B2-microglobulin, age, clinical stage, genomic aberrations as defined by Döhner 2000, ZAP-70 expression, CD38 expression and gender.

Excluded studies

During abstract screening, we excluded 18,713 references that clearly did not match our inclusion criteria. Of the remaining 283 full-text references, we excluded 213 for the following reasons.

- References of prognostic factor studies or prognostic factor identification studies (130 references) (Berke 2019; Bo 2014; Brejcha 2010; Brugiatelli 2007; Bulian 2014; Byrd 2006; Cailliod 2005; Callea 1999; Catovsky 1989; Cesano 2013; Chang 2003; Chauzeix 2018; Chen 1997; Chena 2008; Chevallier 2002; Chiaretti 2014; Christiansen 1994; Ciccone 2012; Claus 2012; Claus 2014; Cmunt 2002; Cocco 2005; Corcoran 2005; Cordone 1998; Cortese 2014; Coscia 2012; Crespo 2003; Cro 2009; D'Arena 2001; D'Arena 2007; Damle 1999; DeAndres-Galiana 2016; Degan 2004; Delgado 2009; Delgado 2014; Del Guidice 2011; Del Poeta 2010; Del Principe 2004; Del Principe 2006; Di Raimondo 2001; Dong 2011; Dong 2014; Durak 2009; El-Kinawy 2012; Gattei 2008; Gdynia 2018; Gentile 2016; Giudice 2018; Gogia 2014; Grabowski 2005; Han 1984; Hock 2010; Hus 2006; Jaksic 1981; Josefsson 2007; Juliusson 1986; Juliusson 1990; Kahraman 2014; Kardum-Skelin 2008; Karmiris 1994; Khalifa 2002; Kim 2004; Kimby 1988; Knospe 1977; Koberda 1989; Korycka-Wolowiec 2011; Krober 2002; Kryachok 2011; Kurec 1992; Lai 2002; Lech-Maranda 2012; Lech-Maranda 2013; Lecouvet 1997; Li 2008; Li 2017a; Lin 2002; Lin 2014; Lozano-Santos 2014; Lucas 2015; Maffei 2007; Maffei 2010; Mansouri 2013; Marasca 2005; Marasca 2013; Martinelli 2008; Masic 1998; Mateva 2001; Matthews 2006; Matthews 2007; Matutes 2013; Miao 2018; Molica 1986; Molica 1988; Molica 1991; Molica 1994; Molica 1998; Molica 1999a; Molica 1999b; Molica 2008; Montserrat 1991; Morabito 2001; Morabito 2015a; Morabito 2018c; Morilla 2008; Nenova 2000; Nipp 2014; Nowakowski 2009; Nuckel 2006; Nuckel 2009; Ocana 2007; Oliveira 2011; Oscier 1990; Paolino 1984; Prokocimer 1985; Qin 2017; Resegotti 1989; Rissiek 2014; Ronchetti 2016; Sarmiento 2002; Shanafelt 2010; Spacek 2009; Stamatopoulos 2017; Strefford 2015; Szymczyk 2018; Vojdeman 2017; Vural 2014; Weiss 2011; Wierda 2003; Winkler 2010; Zenz 2009), and five references aimed at identifying prognostic factor thresholds (Dasgupta 2015; Davis 2016; Rossi 2010a; Tobin 2005a; Tobin 2005b).
- The main focus was on the development of a diagnostic CLL staging system (21 references) (Apelgren 2006; Baccarani 1982; Binet 1981; Binet 1977; Chelazzi 1979; Ciocoiu 1988; De Faria 2000; De Rossi 1989; Ferrara 1981; Jaksic 1992; Molica 1984; Moreno 2019; Rai 1990; Rai 1975; Rossi 1986; Rozman 1979; Santoro 1979; Scolozzi 1981; Velardi 1980; Wu 2010; Zengin 1997).
- Involved genetic analysis only (16 references), e.g. genetic subgrouping, genetic signature(s) or genetic clustering (Baliakas 2015; Bomben 2009; Bou Samra 2014; Chuang 2012; Ferreira 2014; Friedman 2009; Herold 2011; Houldsworth 2014; Morabito 2015b; Orgueira 2019; Queiros 2015; Raponi 2018; Rodriguez 2007; Vallat 2013; Van Damme 2012; Zucchetto 2006).
- Included a population which did not match our PICOTS question (10 references), e.g. previously treated individuals with CLL (Giles 2003; Kardum-Skelin 2009; Keating 2000; Krober 2006; Melo 1987; Nola 2004; O'Brien 1993; Rossi 2010b; Rossi 2011; Weinberg 2007).
- Nine references focused on the characterisation of CLL and the prevalence of prognostic factors (Criel 1997; Cro 2010; Cuneo 2004; D'Arena 2012; Geisler 1997; Gonzalez 2013; Gonzalez-Gascon 2015; Gonzalez-Rodriguez 2010; Hallek 1999).
- Nine references of other types of studies, e.g. risk factor study, diagnostic accuracy study (Cuneo 2018; Deslandes 2007; Dimier 2018; Fang 2019; Kay 2018; Kleinstern 2018; Plesingerova 2017; Salomon-Nguyen 1995; Savvopoulos 2016).
- Four references developed or used an outdated staging system, which is no longer used in today's clinical practice (Bettini 1986; Chastang 1985; French Cooperative Group on CLL 1988; Mandelli 1987).
- Five other types of references, e.g. review or comment to an excluded study (Bomben 2005; Grever 2006; Jaksic 2014; Matutes 2017; Nedeva 2018).
- Three references evaluated outcomes not relevant to this review (Shanafelt 2017; Stamatopoulos 2009; Tallarico 2018).
- Two references focused on scores or models which included predictors available at treatment initiation only (incorrect timing) (Gentile 2018; Nabhan 2017).

Reporting deficiencies

Appendix 8 describes the reporting deficiencies of the included studies per model for the primary outcome each model was developed for. The number of studies includes the development study. Data obtained by contacting authors are included as available.

For the Barcelona-Brno model, the number of events for two studies and calibration for one study was obtained by contacting the corresponding authors of the primary studies. For the CLL-IPI, information on total sample size and number of events was provided by one author. Information on calibration was not reported in publications at all, but was obtained by contacting the corresponding authors. For the GCLLSG model, the confidence intervals of the c-statistic and calibration for the development study publication with one external validation was obtained from the authors. For the MDACC 2007 model, the c-statistic and 95% CI for one external validation study was provided by the corresponding author. We did not obtain any additional information for the other models. We received primary data of one study, but were unable to reconstruct the analysis.

Of the 41 models without external validations, all studies reported the total number of individuals included for analysis. Eighteen studies lacked information on the recruitment period, 19 did not clearly report their study design, 24 did not report the calibration of their model and 31 did not report any measure of discrimination for their model (Appendix 7). Since we included studies with a wide range of publication years, the large amount of missing information may be a result of the methodological evolution of prognostic model research. Moreover, with our decision to also include scores where authors decided seemingly ad hoc to assign points and create a prognostic score, we have included studies that show more resemblance with prognostic factor studies (i.e. testing of the independence of factors from each other) than with a prognostic model development.

Summarised over all categories of included studies (development studies without external validations (N = 40), development studies with external validations (N = 12), external validations of primary outcome of an included model (N = 35)), information was especially lacking on calibration and discrimination. Calibration was reported in 17 and discrimination in 45 out of 87 studies.

Risk of bias and applicability assessment of included studies

Models with more than three external validation studies per outcome

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI *(CLL-IPI D - Bahlo 2016 (development cohort))*

We rated the risk of bias of the development study as low for the domains: participants, predictors and outcome (CLL-IPI D - Bahlo 2016 (development cohort)), and high for the domain 'analysis' due to several reasons (univariable selection and dichotomisation of factors, missing data handling). We rated the risk of bias for the validation studies as low for the domain 'participants' for two studies (CLL-IPI V - Gentile 2016 (Italian cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)), unclear in the case of six studies due to unclear eligibility criteria, and high for two studies due to inappropriate inclusion criteria (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Rani 2018 (Indian cohort)). We rated the risk of bias for the domains 'predictors' and 'outcome' as low for most studies, except one study with an unclear rating for the domain 'predictors' because predictor assessment has probably changed (CLL-IPI V - Gentile 2016 (Italian cohort)), and three studies with an unclear rating for the domain ''outcome' due to short observation time (CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort); CLL-IPI V - Rani 2018 (Indian cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort)). Concerning the domain 'analysis', we considered seven studies to have a high risk of bias due to inappropriate handling of missing data, low number of events and lack of reporting of performance measures. Three studies had a low rating (CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort)), and one had an unclear rating due to lack of performance measures (CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort)). One validation study did not examine the outcome of interest (CLL-IPI V - Molica 2016 (O-CLL1-GISL)) (Figure 14).

Figure 14. Risk of bias (PROBAST) assessment of the CLL-IPI model (Bahlo 2016)

Model		Study	Prediction model Risk Of Bias ASsessment Tool (PROBAST)				
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
CLL-IPI model (overall survival)	Development	Bahlo 2016 (Development cohort)	low	low	low	high	high
	Meta-analysed c-statistic	Bahlo 2016 (Mayo clinic 2001-2014)	high	low	low	low	high
		Bahlo 2016 (SCAN cohort)	unclear	low	low	low	unclear
		Delgado 2017 (Barcelona cohort)	unclear	low	low	low	high
		Gentile 2016 (Italian cohort)	low	unclear	low	high	high
		Muñoz-Novaz 2018 (Spanish cohort)	low	low	low	high	high
		Rani 2018 (Indian cohort)	high	low	unclear	high	high
		Zhu 2018 (Chinese cohort)	unclear	low	low	high	high
	Narrative	Da Cunha-Bang 2016 (Danish cohort)	unclear	low	unclear	unclear	unclear
		Molica 2016 (O-CLL1-GISL)	NA	NA	NA	NA	NA
		Reda 2017 (Milan cohort)	unclear	low	low	high	high
		Rigolin 2017 (Ferrera cohort)	unclear	low	unclear	high	high

We rated the concern for applicability of the development study as high for the domain 'participants' due to the high proportion of individuals with treatment indication (CLL-IPI D - Bahlo 2016 (development cohort)). We rated this low for the domains 'predictors' and 'outcome'. We rated concern for applicability as unclear for the domain 'participants' in all validation studies but two, which received a low rating (CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)), due to unclear eligibility criteria. We rated concern for applicability as unclear for the domain 'predictors' in six validation studies because one predictor (TP53 mutation) was replaced by a proxy predictor (del(17p). The remaining four were rated as low (CLL-IPI V - Bahlo 2016 (SCAN cohort); CLL-IPI V - Delgado 2017 (Barcelona cohort); CLL-IPI V - Rigolin 2017 (Ferrera cohort); CLL-IPI V - Zhu

2018 (Chinese cohort)). We considered the concern for applicability to be low for the domain 'outcome' in all validation studies. One validation study (CLL-IPI V - Molica 2016 (O-CLL1-GISL))) did not examine the outcome of interest (Figure 15).

Figure 15. Applicability assessment for all developed models with external validations

Model		Study	PROBAST – Applicability		
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome
Baliakas model (time-to-first-treatment)	Development	Baliakas 2019 (Development cohort)	high	low	low
	Narrative	Baliakas 2019 (MLL + Scan. cohort)	unclear	low	low
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Development cohort)	unclear	low	low
	Meta-analysed c-statistic	Delgado 2017 (Brno cohort, CZ)	unclear	low	low
		Gentile 2017 (Italian multicentre and Mayo cohort)	unclear	low	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	low	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
	Narrative	Molica 2017 (O-CLL1-GISL)	NA	NA	high
		Reda 2017 (Milan cohort)	unclear	unclear	low
CLL-IPI model (overall survival)	Development	Bahlo 2016 (Development cohort)	high	low	low
	Meta-analysed c-statistic	Bahlo 2016 (Mayo clinic 2001-2014)	unclear	unclear	low
		Bahlo 2016 (SCAN cohort)	unclear	low	low
		Delgado 2017 (Barcelona hospital cohort)	unclear	low	low
		Gentile 2016 (Italian multicentre cohort)	unclear	unclear	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	unclear	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
		Zhu 2018 (Chinese cohort)	unclear	low	low
	Narrative	Da Cunha-Bang 2016 (Danish cohort)	low	unclear	low
		Molica 2016 (O-CLL1-GISL)	NA	NA	high
		Reda 2017 (Milan cohort)	unclear	unclear	low
		Rigolin 2017 (Ferrera cohort)	unclear	low	low
MDACC 2007 model (overall survival)	Development	Wierda 2007 (Development cohort)	low	low	low
	Meta-analysed c-statistic for the index score	Bulian (Italian-Swiss cohort)	low	low	low
		Gentile 2014 (Italian multicentre cohort)	unclear	low	low
		Gentile 2016 (Mayo cohort 2001-2008)	unclear	low	low
		González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	low
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	low	low
		Rani 2018 (Indian cohort)	unclear	unclear	low
	Narrative	Molica 2010 (GIMEMA cohort 1991-2000)	low	high	high
		Molica 2015 (O-CLL1-GISL)	NA	NA	high
		Pflug 2014 (German CLL study group cohort)	unclear	low	low
		Trajkova 2013 (Macedonian cohort)	low	low	low
GCLLSG model (overall survival)	Development	Pflug 2014 (Development cohort)	unclear	low	low
	Narrative	Molica 2015 (O-CLL1-GISL)	NA	NA	high
		Pflug 2014 (Mayo cohort)	unclear	low	low
		Rani 2018 (Indian cohort)	NA	NA	high
GIMEMA model (progression-free survival)	Development	Molica 2005 (Development cohort)	low	low	low
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	low
MDACC 2011 model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	unclear	low	low
	Narrative	Molica 2013 (O-CLL1-GISL)	low	low	low
Morabito model (time-to-first-treatment)	Development	Morabito 2009 (Development cohort)	unclear	high	low
	Narrative	Gentile 2014 (O-CLL1-GISL)	low	low	low
O-CLL1 model (time-to-first-treatment)	Development	Gentile 2016 (O-CLL1-GISL)	low	low	low
	Narrative	Rani 2018 (Indian cohort)	unclear	unclear	low
		Gentile 2016 (Mayo cohort 2001-2008)	unclear	low	low
Rossi model (overall survival)	Development	Rossi 2013 (Development cohort)	low	low	low
	Narrative	Jeromin 2014 (Munich cohort)	low	low	low
		Rossi 2013 (External validation cohort)	low	low	low
Stephens model (overall survival)	Development	Stephens 2015 (Development cohort)	high	low	low
	Narrative	Stephens 2015 (MDACC cohort)	high	low	low
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	high	low	low
	Narrative	Stephens 2015 (MDACC cohort)	high	low	low

Barcelona-Brno model *(Barcelona-Brno D - Delgado 2017 (Barcelona cohort))*

We rated the risk of bias of the development study as unclear for the domain 'participants' due to missing information on eligibility criteria (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)), low for the domains 'predictors' and 'outcome', and high for the domain 'analysis' due to the model-building process (factor selection and weighting). We rated the risk of bias between the validation studies as unclear for the domain 'participants' in three studies (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Rani 2018 (Indian cohort); Barcelona-Brno V - Reda 2017 (Milan cohort)) due to missing eligibility criteria and recruitment period, and low

in two studies (Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.)). We rated the risk of bias for the domains 'predictors' and 'outcome' as low for all validation studies except one (Barcelona-Brno V - Rani 2018 (Indian cohort)), which we rated as unclear for the domain 'outcome' due to the short observation time. Concerning the domain 'analysis', we considered all studies to be at high risk of bias (especially due to handling missing data, low number of events and reporting of performance measures) except one (Barcelona-Brno V - Delgado 2017 (Brno cohort)), which we rated as unclear. One validation study did not examine the outcome of interest (Barcelona-Brno V - Molica 2017 (O-CLL1-GISL)) (Figure 16).

Figure 16. Risk of bias (PROBAST) assessment of the Barcelona-Brno model (Delgado 2017)

Model		Study	Prediction model Risk Of Bias ASsessment Tool (PROBAST)				
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
Barcelona-Brno model (overall survival)	Development	Delgado 2017 (Development cohort)	unclear	low	low	high	high
	Meta-analysed c-statistic	Delgado 2017 (Brno cohort, CZ)	unclear	low	low	unclear	unclear
		Gentile 2017 (Italian multicentre and Mayo cohort)	low	low	low	high	high
		Muñoz-Novaz 2018 (Spanish cohort)	low	low	low	high	high
		Rani 2018 (Indian cohort)	unclear	low	unclear	high	high
	Narrative	Molica 2017 (O-CLL1-GISL)	NA	NA	NA	NA	NA
		Reda 2017 (Milan cohort)	unclear	low	low	high	high

We rated the concern for applicability of the development study as unclear for the domain 'participants' and low for the domains 'predictors' and 'outcome' (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)). Similarly, we rated concern for applicability as unclear for the domain 'participants', and low for the domains 'predictors' and 'outcome' in three validation studies (Barcelona-Brno V - Delgado 2017 (Brno cohort); Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.)) due to undefined eligibility criteria. We assigned two validation studies (Barcelona-Brno V - Rani 2018 (Indian cohort); Barcelona-Brno V - Reda 2017 (Milan cohort)) the rating unclear for the domain 'predictors' due to lack of information on the timing of predictor assessment. One validation study did not examine the outcome of interest (Barcelona-Brno V - Molica 2017 (O-CLL1-GISL)) (Figure 15).

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

We rated the risk of bias for the development study as low across all domains (MDACC 2007 D - Wierda 2007 (MDACC)). We rated the risk of bias across the validation studies as low for the domain 'participants' in six validation studies and high in three validation studies (MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - Rani 2018 (Indian cohort); MDACC 2007 V - Trajkova 2013 (Macedonia)) due to unclear eligibility criteria. We considered all validation studies to have a low risk of bias for the domains 'predictors'. We rated three studies as unclear for the domain 'outcome' (MDACC 2007 V - Molica 2010 (GIMEMA cohort); MDACC 2007 V - Rani 2018 (Indian cohort); MDACC 2007 V - Trajkova 2013 (Macedonia)) because of the short observation time. Concerning the domain 'analysis', we considered all validation studies to have a high or unclear risk of bias due to the low number of events and inappropriate handling of missing data. One validation study did not examine the outcome of interest (MDACC 2007 V - Molica 2015 (O-CLL1-GISL)) (Figure 17).

Figure 17. Risk of bias (PROBAST) assessment of the MDACC 2007 model (Wierda 2007)

Model		Study	Prediction model Risk Of Bias ASsessment Tool (PROBAST)				
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
MDACC 2007 model (overall survival)	Development	Wierda 2007 (Development cohort)	low	low	low	low	low
	Meta-analysed c-statistic for the index score	Bulian (Italian-Swiss cohort)	low	low	low	high	high
		Gentile 2014 (Italian multicentre cohort)	low	low	low	high	high
		Gentile 2016 (Mayo cohort 2001-2008)	high	low	low	unclear	high
		González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	low	high	high
		Muñoz-Novaz 2018 (Spain multicentre cohort)	low	low	low	high	high
		Pflug 2014 (German CLL study group cohort)	low	low	low	high	high
		Rani 2018 (Indian cohort)	high	low	unclear	high	high
	Narrative	Molica 2010 (GIMEMA cohort 1991-2000)	low	low	unclear	high	high
		Molica 2015 (O-CLL1-GISL)	NA	NA	NA	NA	NA
		Trajkova 2013 (Macedonian cohort)	high	low	unclear	high	high

We rated the concern for applicability of the development study as low across all domains (MDACC 2007 D - Wierda 2007 (MDACC)). We rated concern for applicability as low for the domain 'participants' in five validation studies and unclear in four validation studies due to unclear eligibility criteria (MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - Pflug 2014 (3 RCTs); MDACC 2007 V - Rani 2018 (Indian cohort)). Concerning the domain 'predictors', we considered the concern for applicability to be low in seven validation studies, high in one study (MDACC 2007 V - Molica 2010 (GIMEMA cohort)), and unclear in one study (MDACC 2007 V - Rani 2018 (Indian cohort)). Concern for applicability was rated as low for the domain 'outcome' in all validation studies except one, for which we assigned a high rating (MDACC 2007 V - Molica 2010 (GIMEMA cohort)). One validation study did not examine the outcome of interest (MDACC 2007 V - Molica 2015 (O-CLL1-GISL)) (Figure 15).

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies per outcome

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

We rated the risk of bias for the development study as low for all domains except 'analysis' (GCLLSG D - Pflug 2014 (GCLLSG)). We rated 'analysis' at high risk of bias due to dichotomisation and univariable selection of factors, inappropriate handling of missing data and simplification of the model (loss of original weighting). We considered the only validation study with a matching outcome to have a high risk of bias for the domains 'participants' and 'analysis' due to inclusion of participants with available data and low number of events (GCLLSG V - Pflug 2014 (Mayo cohort)). Risk of bias was low for the domains 'predictors' and 'outcome'. The other two validation studies did not examine the outcome of interest (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)) (Figure 18).

Figure 18. Risk of bias (PROBAST) assessment of other models that were externally validated

Model		Study	Prediction model Risk Of Bias ASsessment Tool (PROBAST)				
			Domain 1: Participants	Domain 2: Predictors	Domain 3: Outcome	Domain 4: Analysis	Overall judgement
Baliakas model (time-to-first-treatment)	Development	Baliakas 2019 (Development cohort)	high	low	low	high	high
	Narrative	Baliakas 2019 (MLL + Scan. cohort)	unclear	unclear	unclear	high	high
GCLLSG model (overall survival)	Development	Pflug 2014 (Development cohort)	low	low	low	high	high
	Narrative	Molica 2015 (O-CLL1-GISL)	NA	NA	NA	NA	NA
		Pflug 2014 (Mayo cohort)	high	low	low	high	high
		Rani 2018 (Indian cohort)	NA	NA	NA	NA	NA
GIMEMA model (progression-free survival)	Development	Molica 2005 (Development cohort)	low	low	high	high	high
	Narrative	González Rodríguez 2009 (Cabueñes hospital cohort)	low	low	high	high	high
MDACC 2011 model (time-to-first-treatment)	Development	Wierda 2011 (Development cohort)	high	low	high	high	high
	Narrative	Molica 2013 (O-CLL1-GISL)	low	low	high	high	high
Morabito model (time-to-first-treatment)	Development	Morabito 2009 (Development cohort)	unclear	low	low	high	high
	Narrative	Gentile 2014 (O-CLL1-GISL)	low	low	low	unclear	unclear
O-CLL1 model (time-to-first-treatment)	Development	Gentile 2016 (O-CLL1-GISL)	low	low	high	high	high
	Narrative	Rani 2018 (Indian cohort)	unclear	low	high	high	high
		Gentile 2016 (Mayo cohort 2001-2008)	high	low	high	unclear	high
Rossi model (overall survival)	Development	Rossi 2013 (Development cohort)	low	low	low	unclear	unclear
	Narrative	Jeromin 2014 (Munich cohort)	unclear	low	low	high	high
		Rossi 2013 (External validation cohort)	unclear	low	low	high	high
Stephens model (overall survival)	Development	Stephens 2015 (Development cohort)	high	low	low	high	high
	Narrative	Stephens 2015 (MDACC cohort)	high	unclear	low	high	high
Stephens model (treatment-free survival)	Development	Stephens 2015 (Development cohort)	high	low	high	high	high
	Narrative	Stephens 2015 (MDACC cohort)	high	unclear	high	high	high

We rated the concern for applicability of the development study as unclear for the domain 'participants' because the sample was based on RCT data that may not be representative for all individuals with CLL (GCLLSG D - Pflug 2014 (GCLLSG)). Concern was low for the domains 'predictors' and 'outcome'. Similarly, we rated concern for applicability of the validation study as unclear for the domain 'participants' due to eligibility criteria and low for the domains 'predictors' and 'outcome' (GCLLSG V - Pflug 2014 (Mayo cohort)). The other two validation studies did not examine the outcome of interest (GCLLSG V - Molica 2015 (O-CLL1-GISL); GCLLSG V - Rani 2018 (Indian cohort)) (Figure 15).

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

We rated the risk of bias of the development study as low across all domains except 'analysis' (Rossi D - Rossi 2013 (Italian cohort)), which we considered to be of unclear risk because handling of missing data handling was not reported. We rated the risk of bias for the two validation studies as unclear due to lack of information on study design for the domain ‘participants’, low for the domains ‘predictors’ and ‘outcome’, and high for the domain ‘analysis’ due to missing data handling and low number of events (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)) (Figure 18).

We rated the concern for applicability of the development study as low across all domains (Rossi D - Rossi 2013 (Italian cohort)). We rated concern for applicability of the two validation studies as low for the domains 'participants', 'predictors' and 'outcome' (Rossi V - Jeromin 2014 (Munich cohort); Rossi V - Rossi 2013 (unclear)) (Figure 15).

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

We rated the risk of bias for the development study as high for the domains 'participants' (Stephens OS D - Stephens 2015 (Ohio cohort)), because the study included only individuals with del(17p), which implies that only persons with available FISH assessment could have been included. We considered the domains 'predictors' and 'outcomes' to be of low risk. We rated the domain 'analysis' as high due to small sample size, inappropriate missing data handling and inconsistent reporting of performance measures. In a similar manner, we rated the risk of bias for the validation study as high for the domains 'participants' and 'analysis' due to the same reasons, unclear for the domain 'predictors' and low for the domain 'outcome' (Stephens OS V - Stephens 2015 (MDACC)) (Figure 18).

We rated the concern for applicability of the development study (Stephens OS D - Stephens 2015 (Ohio cohort)) and validation study (Stephens OS V - Stephens 2015 (MDACC)) as high for the domain 'participants' due to the selective sample, and low for the domains 'predictors' and 'outcome' (Figure 15).

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

We rated the risk of bias of the development study as low for the domains ‘predictors’ and ‘outcome’, and high for the domains ‘participants’ since participants with missing data were excluded (Baliakas D - Baliakas 2019 (multicentre)). We rated the

domain 'analysis' as high due to inappropriate handling of missing data, and univariable selection of predictors. Due to reporting deficiencies concerning inclusion criteria, predictor assessment and outcome definition of the validation study (Baliakas V - Baliakas 2019 (MLL + Scan.)), we rated the risk of bias as unclear across all domains except 'analysis', which we considered to be of high risk due to lack of reporting performance measures (Figure 18).

We rated the concern for applicability of the development study as high for the domain 'participants' due to the sample not being representative of all CLL patients and low for the domains 'predictors' and 'outcome' (Baliakas D - Baliakas 2019 (multicentre)). We rated as unclear concern for applicability of the validation study for the domain 'participants' due to lack of information about eligibility criteria and recruitment period (Baliakas V - Baliakas 2019 (MLL + Scan.)), and low for the domains 'predictors' and 'outcome' (Figure 15).

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

We rated the risk of bias for the development study as low for the domains 'participants' and 'predictors' (GIMEMA D - Molica 2005 (GIMEMA cohort)), and high for the domains 'outcome' because a predictor was not excluded from the outcome definition and 'analysis' due to inappropriate handling of missing data, univariable selection of predictors, and simplification of the model. Likewise, we considered the validation study to have a low risk of bias for the domains 'participants' and 'predictors' (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)), while we rated the domains 'outcome' and 'analysis' as high risk of bias due to lack of information on performance measures (Figure 18).

We rated the concern for applicability of the development study (GIMEMA D - Molica 2005 (GIMEMA cohort)) and validation study (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)) as low across all domains (Figure 15).

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

We rated the risk of bias for the development study as high for the domain 'participants' due to inappropriate inclusion criteria (MDACC 2011 D - Wierda 2011 (MDACC)), for the 'outcome' because outcome definition did not exclude one of the predictors and for 'analysis' because of univariable selection of predictors, unclear modelling procedure and lack of performance measures. We only gave a low rating to the domain 'predictors'. Similarly, we considered the risk of bias of the validation study to be high for the domains 'outcome' for the same reason and 'analysis' due to unclear handling of missing data and low number of events (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)), while we rated the domains 'predictors' and 'participants' as low risk (Figure 18).

We rated the concern for applicability of the development study as unclear for the domain 'participants' due to inappropriate inclusion criteria and low for the domains 'predictors' and 'outcome' (MDACC 2011 D - Wierda 2011 (MDACC)). We rated concern for applicability of the validation study as low across all domains (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)) (Figure 15).

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

We rated the risk of bias of the development study as unclear for the domain 'participants' because it was not clear if individuals with missing data were considered (Morabito D - Morabito 2009 (Italian cohort)), low for the domains 'predictors' and 'outcome', and high for the domain 'analysis' due to low number of events, dichotomisation and univariable selection of factors and missing performance measures. In the case of the validation study (Morabito V - Gentile 2014 (O-CLL1-GISL)), we rated the risk of bias as low across all domains except 'analysis', which we considered to be of unclear risk due to lack of reporting performance measures (Figure 18).

We rated the concern for applicability of the development study as unclear for the domain 'participants' due to unclear eligibility criteria, high for the domain 'predictors' and low for the domain 'outcome' (Morabito D - Morabito 2009 (Italian cohort)). We rated as low, concern for applicability of the validation study across all domains (Morabito V - Gentile 2014 (O-CLL1-GISL)) (Figure 15).

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

We rated the risk of bias of the development study as low across the domains 'participants' and 'predictors' (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). We rated the domain 'outcome' as high risk of bias because one predictor was also included in the outcome definition. We rated the domain 'analysis' as high risk of bias due to unclear handling of missing data, categorisation and univariable selection of factors.

Concerning the domain 'participants', we rated one validation study as unclear due to missing eligibility criteria (O-CLL-1 V - Rani 2018 (Indian cohort)), and one as high risk due to inappropriate inclusion criteria (O-CLL1 V - Gentile 2016 (Mayo cohort)). We rated both validation studies as low risk of bias for the domain 'predictors' and high risk for the domain 'outcome' because one predictor was also included in the outcome definition. We considered the risk of bias for the domain 'analysis' to be high for one validation study due to the low number of events and no report for handling of missing data (O-CLL-1 V - Rani 2018 (Indian cohort)), while we rated the other study as unclear due to unclear handling of missing data (O-CLL1 V - Gentile 2016 (Mayo cohort)) (Figure 18).

We rated the concern for applicability of the development study as low across all domains (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). We rated concern for applicability of the two validation studies (O-CLL1 V - Gentile 2016 (Mayo cohort); O-CLL-1 V - Rani 2018 (Indian cohort)) as unclear for the domains 'participants' due to undefined eligibility criteria, the 'predictors' as low for the first and unclear for the second due to undefined timing of prognostic factor measurement, and low for both studies for the domain 'outcome' (Figure 15).

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

For risk of bias of the model for TFS in Stephens and colleagues (Stephens TFS D - Stephens 2015 (Ohio cohort)), please see description in the section for OS (above), as the authors used the same development and validation cohorts for both outcomes.

Deficiencies of prognostic model development studies (models with PROBAST rating)

Among the 12 prognostic model development studies with at least one external validation study (including the three meta-analysed models), we rated nine at high or unclear risk of bias for the domain

'analysis'. We identified the following most common issues in these studies regarding model development.

- *Missing data handling*: only one model development study imputed missing values; all other studies based their analysis on complete-case analyses.
- *Predictor selection based on univariable analysis*: eight studies derived their multivariable model on prior univariable selection using P values as a criterion.
- *Correction of estimates*: several development studies used split-sample or bootstrapping techniques. However, none of the studies reported a correction of their regression coefficients to reduce overfitting to the development cohort.
- *Categorisation of factors*: at some stage during the development process, all studies reported categorisation or dichotomisation of continuous prognostic factors.
- *Model weights*: 10 development studies presented a simplification of the original model formula in the form of a point score (weighting, grouping, or counting of disadvantageous prognostic factors), which results in loss of information.

Findings

Reporting of calibration was rare in the studies that we identified. If reported or provided afterwards via email, the format varied so that we were unable to summarise the 'traditional' calibration measures (calibration plots, calibration tables, O:E ratios). Therefore, we decided to use a different way of presenting calibration graphically (Figure 19; Figure 20; Figure 21), based on something that most studies reported (survival frequencies per group). The black lines represent the development study and each coloured line one external validation cohort. The survival per risk group of the development cohort can be interpreted as the expected survival frequencies, and the survival frequencies in the external validation cohorts the observed survival frequencies. The pooled result gives the pooled observed survival frequencies.

Models with more than three external validation studies per outcome and meta-analysis

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

CLL-IPI *(CLL-IPI D - Bahlo 2016 (development cohort))*

The CLL-IPI is a prognostic score derived from univariable selection of predictors which were then entered into a forward-stepwise proportional-hazards Cox regression model with a hierarchy based on completeness of the predictor information (CLL-IPI D - Bahlo 2016 (development cohort)). The dataset was split into a training and external validation dataset after univariable selection of factors (66% and 33% of the data, respectively). Continuous factors were dichotomised based on published thresholds and quartiles. The authors integrated prognostic factors that were independently associated with the outcome in the final multivariable model in a weighted manner to construct their prognostic index (Table 2).

Table 2. Scoring of the CLL-IPI (Bahlo 2016)

Prognostic factor	Point distribution
Age (≥ 65 years)	1
Clinical stage (Rai I-IV or Binet B-C)	1
IgHV mutational status (unmutated)	2
B2-microglobulin (> 3.5 mg/L)	2
TP53 status (deleted or mutated)	4

Notes: IgHV: immunoglobulin heavy chain variable region genes

Calibration

None of the included studies regarding the CLL-IPI reported a measure of calibration in their publication.

We included eight external validation studies of this model, with a total of 4891 individuals to calculate the pooled survival frequencies per risk group at five years (Figure 19). The pooled observed survival frequencies of all validation studies were 92.5% (89.2% to 94.8%) for the low-risk group (score 0-1), 85.0% (79.7% to 89.1%) for the intermediate-risk group (score 2-3), 64.9% (56.4% to 72.6%) for the high-risk group (score 4-6) and 40.4% (29.3% to 52.6%) for the very high-risk group (score 7-10). For the low, intermediate and high-risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the very high-risk group, survival according to the development study would have been lower than overall in the external validation studies. For the low and intermediate-risk groups, the 95% PIs were relatively small. The other risk groups showed wide 95% PIs, indicating uncertainty (Table 3), probably due to the low number of individuals in these subgroups (N = 765 and N = 201, respectively). We judged heterogeneity (visual inspection) between external validation cohorts to be low for the calibration of the CLL-IPI.

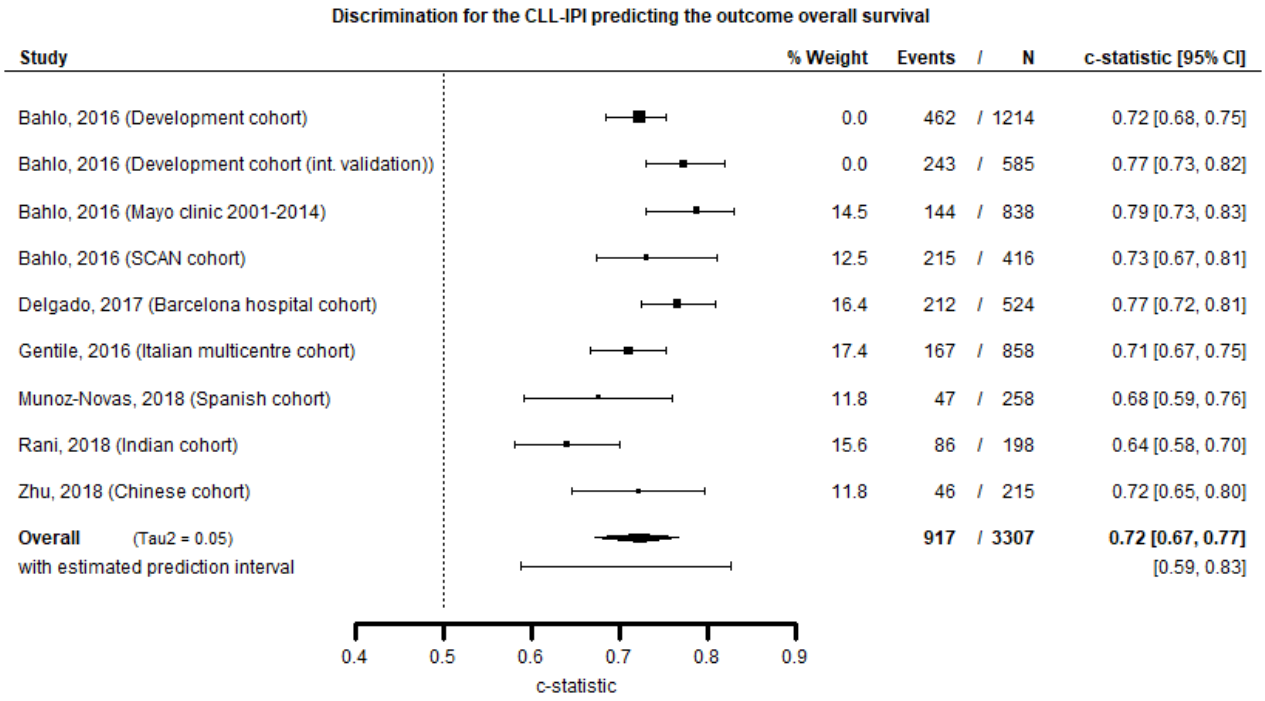
Table 3. Survival per CLL-IPI risk group

Risk group	Risk score	Development study (training dataset): percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	0-1	93.2% (90.5% to 96.0%)	92.5% (89.2% to 94.8%)	82.5% to 97.0%
Intermediate risk	2-3	79.3% (75.5% to 83.2%)	85.0% (79.7% to 89.1%)	68.8% to 93.6%
High risk	4-6	63.3% (57.9% to 68.8%)	64.9% (56.4% to 72.6%)	41.9% to 82.6%
Very high risk	7-10	23.3% (12.5% to 34.1%)	40.4% (29.3% to 52.6%)	13.2% to 72.4%

Discrimination

We included seven external validation studies with a total of 3307 individuals and 917 deaths observed during the overall observation time. Two studies that presented a measure for the c-statistic did not report a measure of uncertainty for this measure, which we estimated using the Newcombe method (CLL-IPI V - Gentile 2016 (Italian cohort); CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)). For one study (CLL-IPI V - Zhu 2018 (Chinese cohort)), we included the AUC as a measure of discrimination instead of the c-statistic. The pooled c-statistic of the seven studies was 0.72 (95% CI 0.67 to 0.77), with some degree of heterogeneity (between-study standard deviation, tau = 0.21). The 95% PI ranged from 0.59 to 0.83 (Figure 22).

Figure 22. Meta-analysis of the c-statistic for the CLL-IPI model (Bahlo 2016)



The c-statistic of the model development study was 0.72 (95% CI 0.68 to 0.75). Based on sensitivity analysis omitting the study that used the AUC instead of the c-statistic (Figure 23), and comparing estimated and reported 95% CI of the c-statistic (Figure 24; Figure 25), we conclude that this had no effect on the result. Four out of seven external validation studies replaced the prognostic factor TP53 with a proxy, del(17p); sensitivity analysis showed no substantial differences in discriminative performance (Figure 26).

Figure 23. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting AUC

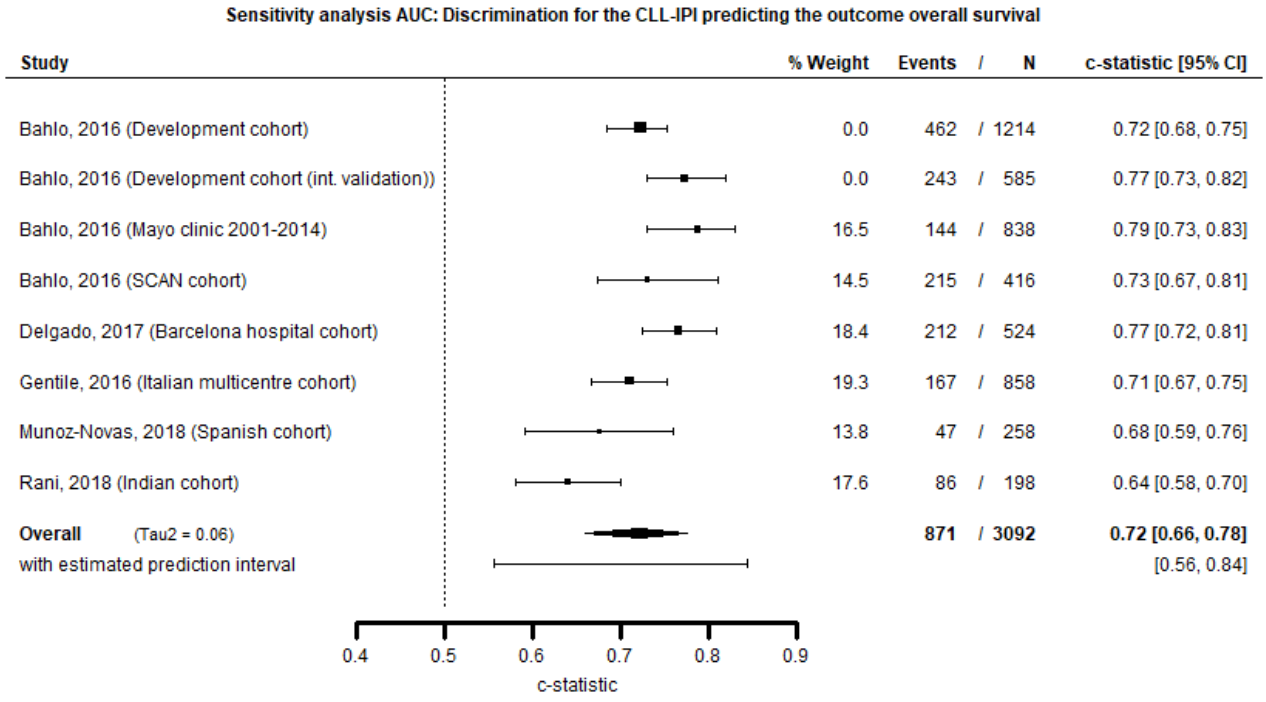


Figure 24. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting no 95% CI

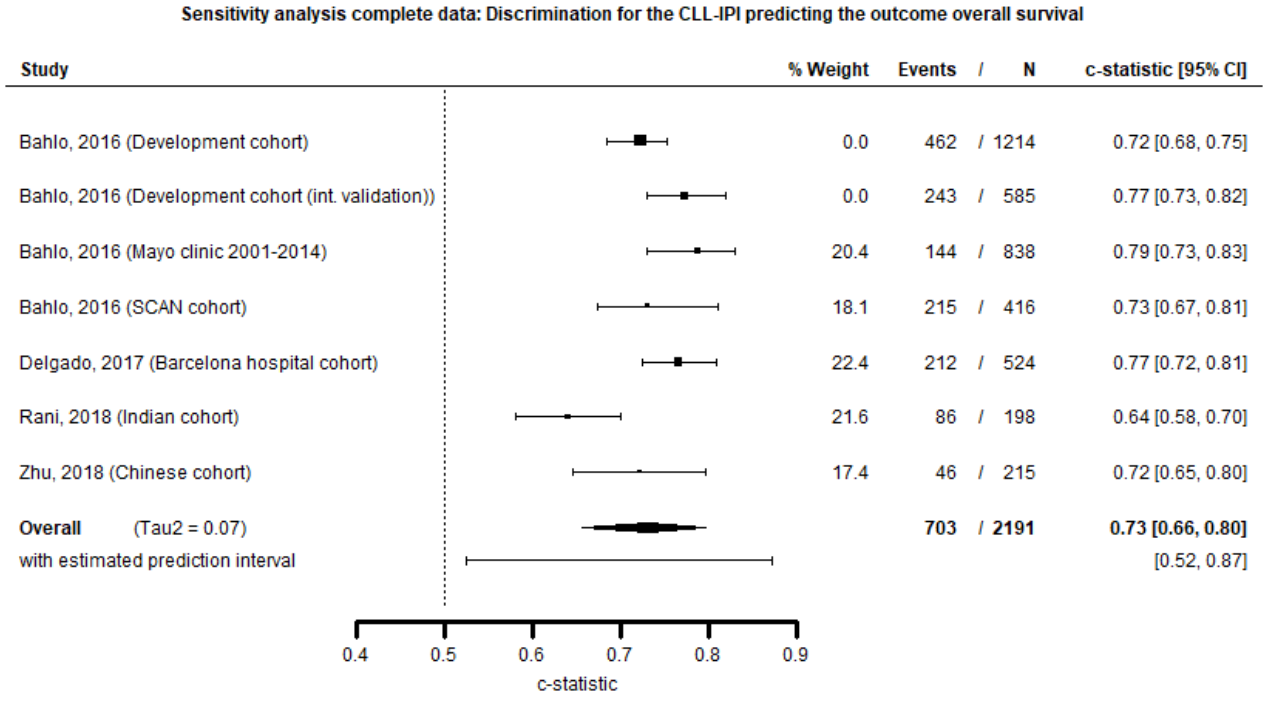


Figure 25. Sensitivity analysis of the c-statistic for the CLL-IPI excluding the study reporting no 95% CI. All 95% CIs were replaced by estimates using the Newcombe method for a comparison with Figure 27

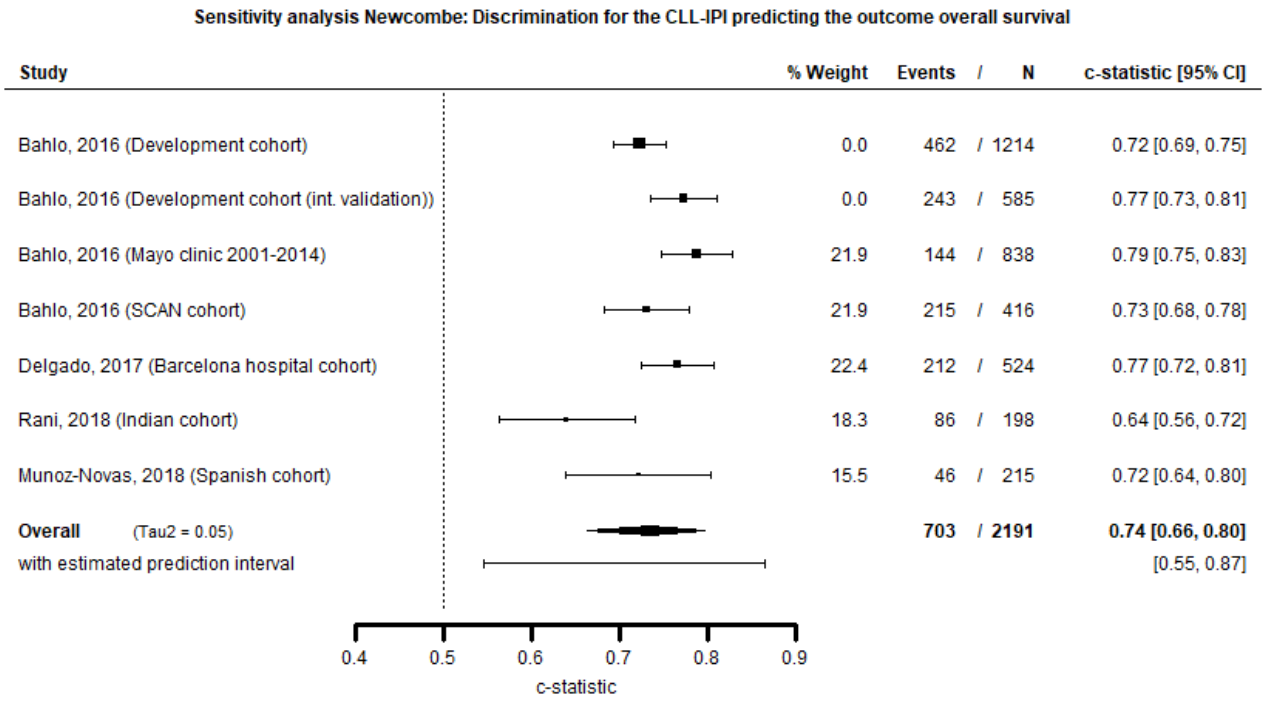
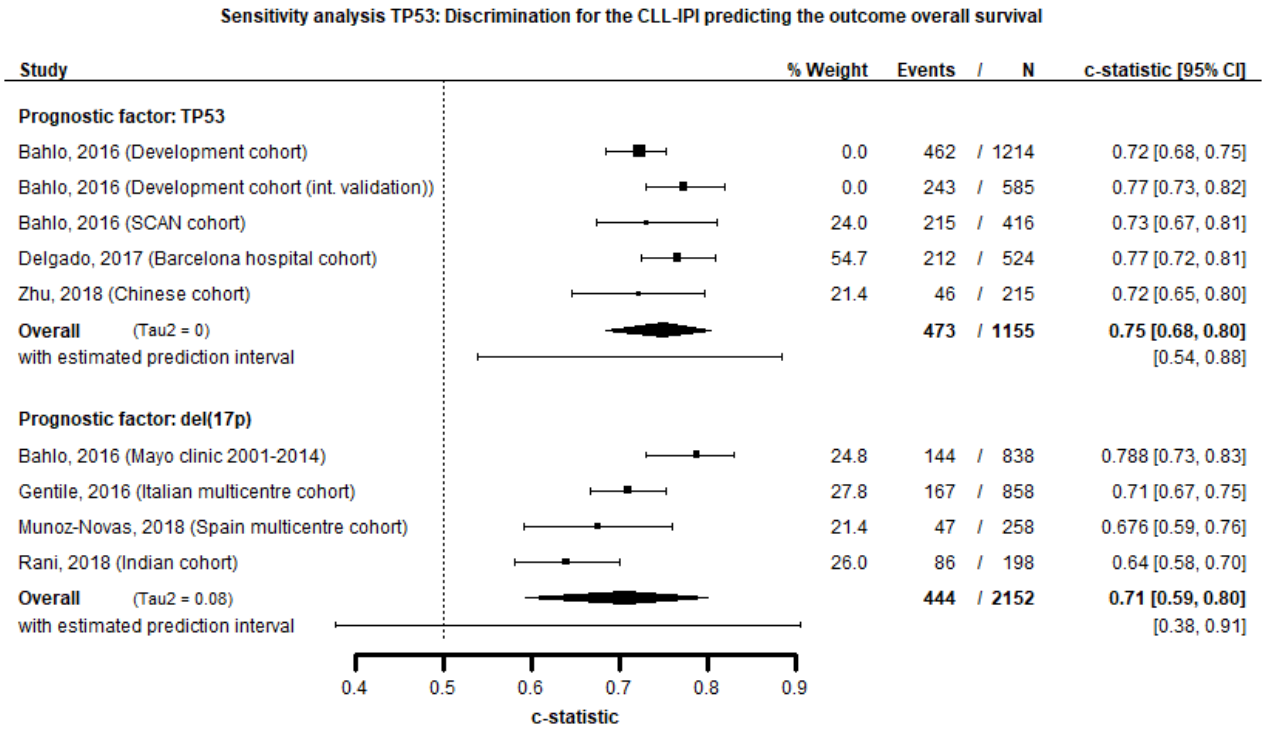


Figure 26. Sensitivity analysis of the c-statistic for the CLL-IPI regarding the availability of the predictor TP53 and its proxy del(17p)



Barcelona-Brno model *(Barcelona-Brno D - Delgado 2017 (Barcelona cohort))*

The Barcelona-Brno score is a prognostic score constructed by comparing the best combinations of the five factors included in another model for CLL (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)), the CLL-IPI by Bahlo 2016 CLL-IPI D - Bahlo 2016 (development cohort)). Incomplete data were most likely part of the exclusion criteria, and therefore persons with missing data were excluded from the analysis in the primary publication. The score classified individuals into three risk groups: persons without del(11q) or del(17p) and mutated IgHV status were classified as low risk; persons with del(11q) or del(17p) and unmutated IgHV status were classified as high risk; all other persons were classified as intermediate risk.

Calibration

None of the included studies in reference to the Barcelona-Brno model reported a measure of calibration in their publication.

We included three external validation studies of this score with a total of 1974 individuals to calculate the pooled survival frequencies per risk group at five years (Figure 20). The pooled observed survival frequencies of all validation studies were 90.5% (95% CI 85.1% to 94.0%) for the low-risk group, 79.7% (95% CI 70.7% to 86.5%) for the intermediate-risk group and 62.5% (95% CI 49.3% to 74.1%) for the high-risk group. For the low and intermediate-risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the high-risk group, survival according to the development study would have been higher than overall in the external validation studies. For the low-risk group, the 95% PI was relatively small. However, the intermediate and high-risk groups showed a wide 95% PI, indicating uncertainty (Table 4). We judged heterogeneity (visual inspection) between external validation cohorts to be low for the calibration of the Barcelona-Brno score.

Table 4. Survival per Barcelona-Brno model risk group

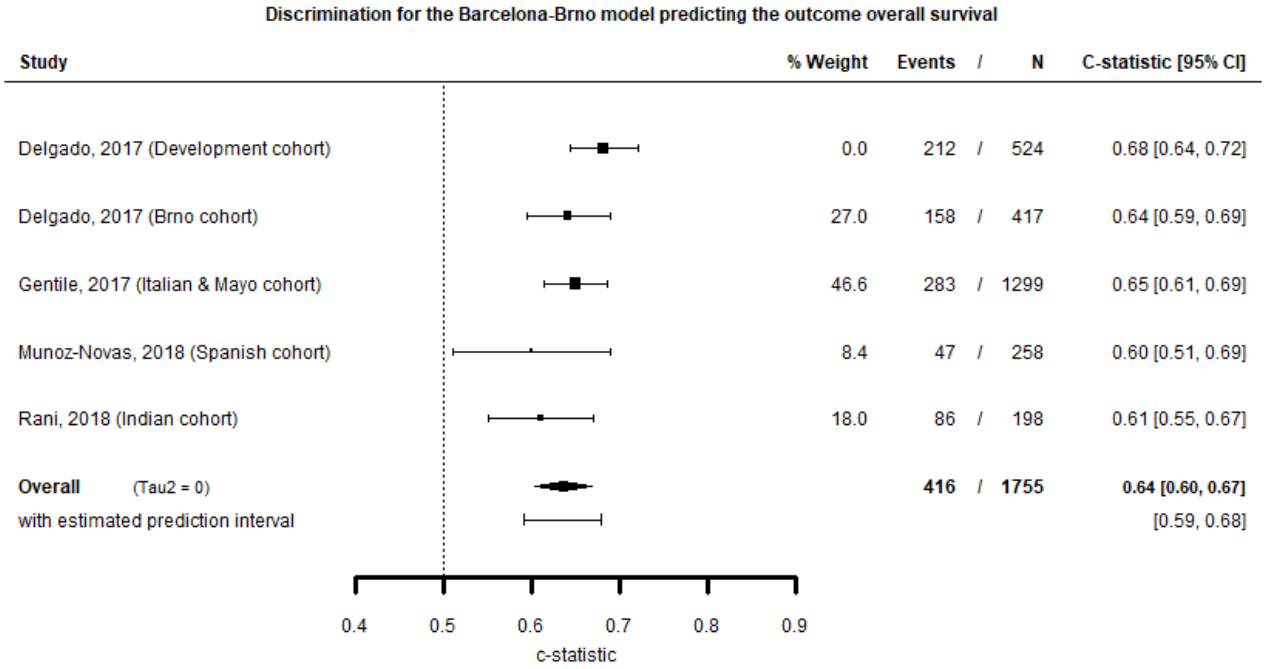
Risk group	Risk group	Development study: percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	No del(11q) or del(17p) and mutated IgHV status	93.4% (90.4% to 96.5%)	90.5% (85.1% to 94.0%)	80.4% to 95.7%
Intermediate risk	All other	83.8% (78.4% to 89.6%)	79.7% (70.7% to 86.5%)	63.1% to 90.0%
High risk	Del(11q) or del(17p) and unmutated IgHV status	70.0% (59.3% to 82.7%)	62.5% (49.3% to 74.1%)	41.3% to 79.7%

Notes: IgHV: immunoglobulin heavy chain variable region genes

Discrimination

For our meta-analysis of the c-statistic, we included four external validation studies with a total of 1755 individuals and 416 deaths observed during the overall observation time. Two studies that presented a measure for the c-statistic did not report a measure of uncertainty, which we estimated using the Newcombe method (Barcelona-Brno V - Gentile 2017 (Italian & Mayo); Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.)). The pooled c-statistic of the four studies was 0.64 (95% CI 0.60 to 0.67), with no heterogeneity between studies (tau = 0.00). The 95% PI ranged from 0.59 to 0.68 (Figure 27).

Figure 27. Meta-analysis of the c-statistic for the Barcelona-Brno model (Delgado 2017)



The c-statistic of the model development study was 0.68 (95% CI 0.64 to 0.72); it was unclear whether bootstrapping techniques were used to obtain this estimate.

MDACC 2007 (Wierda 2007) (MDACC 2007 D - Wierda 2007 (MDACC))

The MDACC 2007 model is a prognostic model based on univariable selection of predictors that were, when significant, entered into a proportional-hazards Cox regression model (MDACC 2007 D - Wierda 2007 (MDACC)). Bootstrapping was used to account for optimism in the c-statistic. The authors provided the precise formula to calculate a point score which could be translated graphically by a nomogram into individual survival probabilities at five or 10 years. In addition, for simplified use in clinical practice, the authors created an index score in a weighted manner (Table 5).

Table 5. Scoring of the MDACC 2007 model

	Point distribution			
Prognostic factor	**0**	**1**	**2**	**3**
Age (years)		< 50	50-65	> 65
ß-2 microglobulin (mg/L)	< ULN	1-2 × ULN	> 2 × ULN	
ALC (× 10^9/L)	< 20	20-50	> 50	
Gender	Female	Male		
Rai stage	0-II	III-IV		
No. of involved nodal groups	≤ 2	3		

Notes: ULN: upper limit of normal; ALC: acute lymphocyte count

Calibration

The development study provided a calibration plot based on bootstrap sampling. The data points representing the observed versus expected survival probabilities are close to the ideal line. At the higher frequencies, the nomogram slightly underestimates survival. Measures of calibration were not reported in any of the external validation studies.

We included five external validation studies of this model with a total of 3786 individuals for calculating the pooled survival frequencies per risk group at 5 years (graphically represented in Figure 21). The pooled observed survival frequencies of all validation studies was 97.0% (95% CI 94.3% to 98.4%) for the low-risk group (score 1-3), 82.3% (74.6% to 88.0%) for the intermediate-risk group (score 4-7) and 45.6% (31.3% to 60.5%) for the high-risk group (score ≥ 8). For the low and intermediate-risk group, the pooled result of the external validation studies approximated the survival frequencies of the model development study. In the high-risk group, survival according to the development study would have been higher than overall in the external validation studies. For the low-risk group, the 95% PI was relatively small. The intermediate and high-risk group showed wide 95% PIs, indicating uncertainty (Table 6). We judged heterogeneity (visual inspection) of calibration between external validation cohorts to be low for the MDACC 2007 index score.

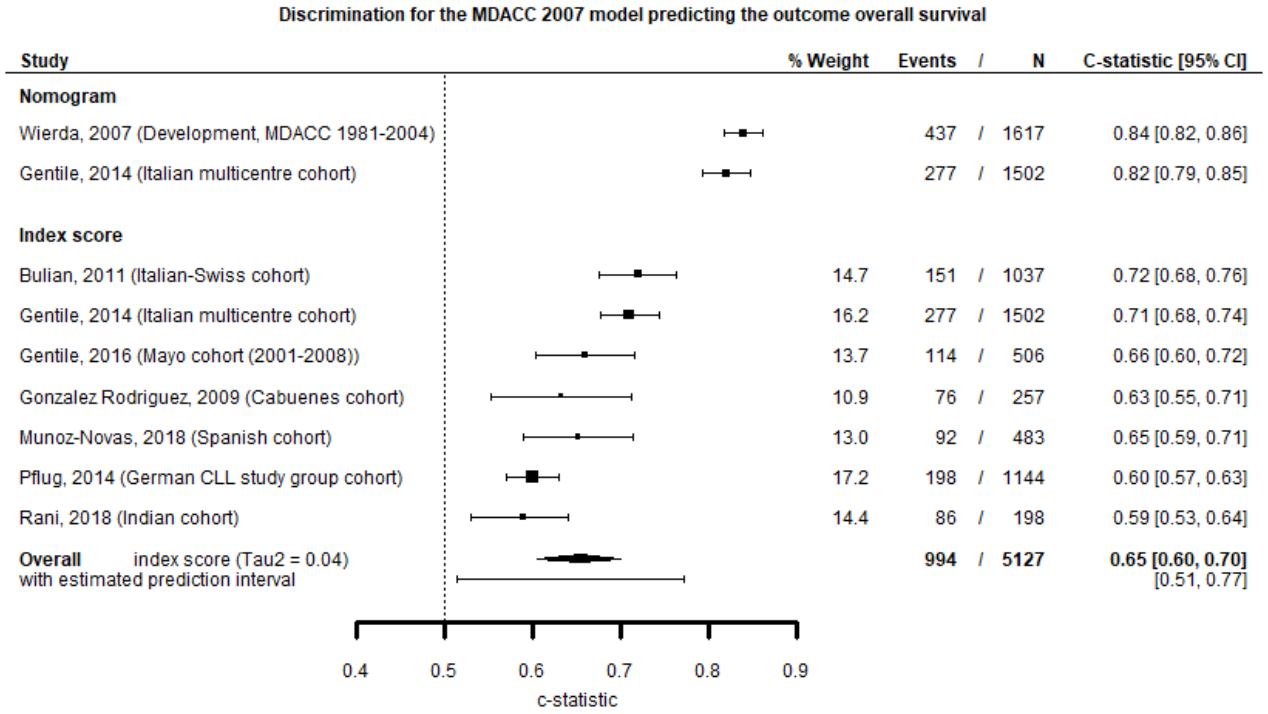
Table 6. Survival per MDACC 2007 model risk group

Risk group	Index score	Development study: Percentage of persons surviving at 5 years (95% CI)	Pooled percentage of persons surviving at 5 years (95% CI)	95% prediction interval (PI)
Low risk	1-3	97.0% (95.0% to 99.0%)	97.0% (94.3% to 98.4%)	90.9% to 99.0%
Intermediate risk	4-7	80.0% (78.0% to 82.0%)	82.3% (74.6% to 88.0%)	61.5% to 93.1%
High risk	≥ 8	55.0% (47.2% to 62.8%)	45.6% (31.3% to 60.5%)	21.2% to 72.3%

Discrimination

We included seven external validation studies with a total of 5127 individuals and 994 deaths observed during follow-up for meta-analysis of the c-statistic in the case of the index score. Only one external validation study provided sufficient data on performance for the nomogram, which is presented together with the development study discrimination on the upper part of the forest plot (Figure 28). Four studies that reported a measure for the c-statistic did not report a measure of uncertainty, which we estimated using the Newcombe method (MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort); MDACC 2007 V - Gentile 2016 (Mayo cohort); MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort)). For one study (MDACC 2007 V - González Rodríguez (Cabueñes), we included the AUC as measure of discrimination instead of the c-statistic. The pooled estimate of the seven studies was 0.65 (95% CI 0.60 to 0.70), with some degree of heterogeneity (between-study standard deviation, tau = 0.21). The prediction interval, which describes a range for the predicted model discrimination in a new validation study of the model, ranged from 0.51 to 0.77.

Figure 28. Meta-analysis of the c-statistic for the MDACC 2007 model (Wierda 2007)



The bootstrap-corrected c-statistic of the model development study was 0.84 (95% CI 0.82 to 0.86). The validation study for the nomogram showed a similar discrimination of 0.82 (95% CI 0.79 to 0.85). Other studies that used the nomogram provided the point score of the formula. However, due to the poor graphical quality of the presented nomogram, authors did not attempt to estimate the expected individual survival probabilities per individual (MDACC 2007 V - Bulian 2011 (Italian-Swiss); MDACC 2007 V - Gentile 2014 (Italian cohort)).

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

GCLLSG model (Pflug 2014) (GCLLSG D - Pflug 2014 (GCLLSG))

The GCLLSG model was derived from univariable selection of predictors which were entered into a forward and backward stepwise proportional-hazards Cox regression model (GCLLSG D - Pflug 2014 (GCLLSG)). The authors controlled for several factors specific to the RCT data that were used for model-building (e.g. study, treatment etc). Bootstrapping techniques were used to test the robustness of the Cox regression model. Persons with missing data were excluded from analysis. Factors were categorised based on published thresholds and quartiles. The authors assigned risk scores to factors, which proved to be independent in the final multivariable model, in a weighted manner to construct the prognostic index presented in Table 7.

Table 7. Scoring of the GCLLSG model

Prognostic factor	Point distribution
FISH category del(17p)	6
Serum thymidine kinase (> 10.0 U/L)	2
B2-microglobulin (> 3.5 mg/L)	2
B2-microglobulin (> 1.7 and ≤ 3.5 mg/L)	1

IgHV mutational status (unmutated)	1
ECOG PS (> 0)	1
FISH category del(11q)	1
Gender (male)	1
Age (> 60 years)	1

Notes: FISH: fluorescence in situ hybridisation; IgHV = immunoglobulin heavy chain variable region genes; ECOG PS: Eastern Cooperative Oncology Group Performance Status

Four risk groups were determined: low risk (0-2 points), intermediate risk (3-5 points), high risk (6-10 points) and very high risk (> 10 points). Five- and six-year OS was reported for each point score.

Calibration

Calibration was not reported for the development and external validation study (GCLLSG V - Pflug 2014 (Mayo cohort)).

In the development study, survival was 95.2%, 86.9%, 67.6% and 18.7% at five years for the low, intermediate, high and very high-risk group, respectively. In the validation study, survival was 95.2%, 91.1%, 71.7% and 13.6% at five years for the low, intermediate, high and very high-risk group, respectively.

Discrimination

The c-statistic of the prognostic score in the model development study was 0.75 (95% CI 0.70 to 0.78), discrimination in the external validation cohort was 0.77 (95% CI 0.70 to 0.83).

Rossi (Rossi 2013) (Rossi D - Rossi 2013 (Italian cohort))

The prognostic score by Rossi and colleagues was developed in several steps (Rossi D - Rossi 2013 (Italian cohort)). Firstly, the authors identified factors that were independently associated with OS by univariable and multivariable Cox regression analysis. Bootstrapping techniques were used to test the robustness of the model. In the next step, the factors were entered into a decision tree algorithm to divide individuals in subgroups. To test the stability of the decision tree, the random survival forest method with amalgamation algorithm was used. It was unclear how missing data were handled.

Calibration

No study reported a measure of calibration for this model.

In the development study, OS at five years was 86.9% in the very low-risk group (del13q14 only), 77.6% in the low-risk group (normal/+12), 65.9% in the intermediate-risk group (NOTCH1 and/or SF3B1 mutations and/or del11q22-q23 in the absence of TP53 and BIRC3 abnormalities) and 50.9% in the high-risk group (TP53 disruption and/or BIRC3 disruption independent of co-occurring lesions). One external validation study reported median survival per risk group (not reached for the very low-risk group, 13.8 years for the low-risk group, 11.2 years for the intermediate-risk group and 7.7 years for the high-risk group) (Rossi V - Rossi 2013 (unclear)). In the other validation study, OS at five years was 91% in the very low-risk group, 90% in the low-risk group, 75.2% in the intermediate-risk group and 62.1% in the high-risk group (Rossi V - Jeromin 2014 (Munich cohort)).

Discrimination

The c-statistic of the prognostic score in the development cohort was 0.64. Only one of the two validation studies reported a measure of discrimination; the c-statistic was 0.66 (95% CI not reported).

Stephens model (Stephens 2015) (Stephens OS D - Stephens 2015 (Ohio cohort))

To develop prognostic models for OS, the authors identified significant prognostic factors in multivariable proportional hazards models using backwards selection. Based on the final model, they defined a simplified risk score based on the strength of the association. Missing data were accounted for by multiple imputation techniques. The model and score for OS included the factors ECOG performance status, age and lactate dehydrogenase (LDH).

Calibration

No study reported a measure of calibration for the prognostic models.

In the development cohort, the percentage of alive persons at two years was 89% (95% CI 74% to 96%) for score 0, 66% (95% CI 41% to 82%) for score 2 and 0% for score 4. In the validation cohort, the percentage of alive persons at two years was 95% (95% CI 83% to 99%) for score 0, 80% (95% CI 55% to 92%) for score 2 and 20% (95% CI 1% to 58%) for score 3.

Discrimination

The discrimination of the exact model for OS was 0.76 ($P < 0.03$) and 0.73 ($P < 0.0001$) for the development cohort. The validation cohort showed a c-statistic of 0.68 for the simplified score for OS.

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

Baliakas model (Baliakas 2019) (Baliakas D - Baliakas 2019 (multicentre))

The model presented by Baliakas and colleagues was derived from univariable selection of predictors which were entered into a proportional-hazards Cox regression model (Baliakas D - Baliakas 2019 (multicentre)), that was internally validated by bootstrapping and further confirmed by recursive partitioning based on conditional inference trees and merging of terminal nodes by an amalgamation algorithm. First, the authors split individuals in two groups, mutated and unmutated IgHV. The risk groups for mutated cases were as follows: (1) low risk: non-TP53 abnormality/+12/subset #2 membership and Binet stage A; (2) intermediate risk: Binet A with one of the following: TP53 abnormality and/or +12 and/or subset #2 membership, (3) high risk: Binet B, (4) very high risk: Binet C. The risk groups for unmutated cases were as follows: (1) very low risk: non-TP53abn/SF3B1mut/del(11q) female Binet A; (2) low risk: non-TP53abn/SF3B1mut/del(11q) male Binet A; (3) intermediate risk: Binet A with one of the following: TP53abn and/or SF3B1mut and/or del(11q); (4) high risk: Binet B; (5) very high risk: Binet C.

Calibration

No measures of calibration were reported for this model.

In the development study, the five-year treatment probability for mutated cases was 12% for group 1, 40% for group 2, 64% for group 3 and 92% for group 4. For unmutated cases, the five-year treatment probability was 45% for group 1, 65% for group 2, 78% for group 3, 90% for group 4 and 100% for group 5.

In the external validation cohort, the five-year treatment probability was only reported for the intermediate-risk groups (for mutated cases 43%, group 2; and for unmutated cases 74%, group 3).

Discrimination

In the development cohort, the discrimination was 0.745 (SE = 0.013) and 0.753 (SE = 0.013) for the cases with mutated IgHV and unmutated IgHV status, respectively. Discrimination was not reported for the external validation cohort.

GIMEMA model (Molica 2005) (GIMEMA D - Molica 2005 (GIMEMA cohort))

The GIMEMA model was derived from univariable selection of predictors which were entered into a proportional-hazards Cox regression model (GIMEMA D - Molica 2005 (GIMEMA cohort)). No internal validation was reported. Negative risk factors that were independent in multivariable analysis (lymphocyte doubling time < 12 months, lymphocyte count >30 × 10^9 per L and Rai stage I to II) were assigned one point each (not weighted for HR) to create a score. Gender was added afterwards as relevant for individuals with score 0, but not for individuals with scores 1 to 3. Three risk groups were created: (1) females with score 0; (2) males with score 0; and (3) individuals with score 1 to 3 of either gender.

Calibration

No measure of calibration was reported for this model.

In the development study, the 10-year PFS was 76.2% for group 1, 61.4% for group 2 and 37.8% for group 3, respectively. In the validation study, the 10-year treatment-free survival was 89% for group 1, 54% for group 2 and 0% for group 3 (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)).

Discrimination

The development study did not report a measure of discrimination. The AUC for the validation study was 0.58 (95% CI 0.49 to 0.66).

MDACC 2011 (Wierda 2011) (MDACC 2011 D - Wierda 2011 (MDACC))

The MDACC 2011 model (MDACC 2011 D - Wierda 2011 (MDACC)) is a prognostic model derived from univariable selection of predictors. After univariable selection, the dataset was split into a training and test set. Significant predictors entered into a proportional-hazards Cox regression model (forward selection procedure with P < 0.1, and removed if not P < 0.5 in final model) were developed on the test set and applied to the test set. The factors of the final model were entered into Cox regression models for 1000 bootstrap samples to show their robustness. The final model formula can be used to derive a point score that can be translated graphically by a nomogram into individual survival probabilities at five or 10 years.

The formula is as follows: [I(No. of lymph node sites involved = 3) × 7.370 + I(FISH11q del) × 9.312 + I(FISH 17p del) × 11.285 + (diameter of largest cervical lymph node in cm) × 4.172 + (LDH/100) × I([IgHV gene mutated] × 5.000 + (LDH/100) × I(IgHV gene = unmutated) × 1.065] + 35.467.

Calibration

No study reported a measure of calibration for this model.

Discrimination

The development study did not report a measure of discrimination. The validation study included a total of 337 participants (91 treated) and reported a c-statistic of 0.71 (95% CI 0.60 to 0.82) (MDACC 2011 V - Molica 2016 (O-CLL1-GISL)).

Morabito model (Morabito 2009) (Morabito D - Morabito 2009 (Italian cohort))

The prognostic model developed by Morabito and colleagues (Morabito D - Morabito 2009 (Italian cohort)) was derived from univariable analysis, including dichotomisation of factors with cut-off determination by ROC curves. Significant (P > 0.05) prognostic factors were entered into proportional-hazards Cox regression models. Each unfavourable marker that remained significant was assigned one point, and the sum of these points formed the final score (CD38 positive = 1, ZAP-70 positive = 1, IgHV unmutated status = 1). Individuals with scores of two or three points were categorised together to form the high-risk group.

Calibration

No study reported a measure of calibration for this model.

The five-year PFS of the validation study for this model was 91.7% for the low-risk group (score 0), 82.9% for the intermediate-risk group (score 1) and 57.4% for the high-risk group (score 2-3) (Morabito V - Gentile 2014 (O-CLL1-GISL)).

Discrimination

Both the development and validation study did not report a measure of discrimination.

O-CLL1 model (Gentile 2016) (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL))

The prognostic score developed by Gentile and colleagues was derived from univariable selection of predictors (O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)). Before entering the factors into proportional-hazards Cox regression models (P < 0.05), continuous factors were dichotomised using established thresholds or ROC curves. Bootstrapping techniques were used to test the robustness of the Cox regression model. There were no persons with missing values in the sample. The authors assigned risk scores to each prognostic factor that was significant in multivariable analysis based on their hazard ratio (Table 8). Individuals were divided into three different risk categories for TFS, low (score 0 to 2), intermediate (score 3 to 5) and high risk (score 6 to 7).

Table 8. Scoring of the O-CLL1 model

	Point distribution		
Prognostic factor	**0**	**1**	**2**
Rai staging system	0	I-II	-
B2-microglobulin	Normal	-	Elevated
ALC (10^9/L)	< 10	-	≥ 10
IgHV mutational status	Mutated	-	Unmutated

Notes: ALC: acute lymphocyte count; IgHV: immunoglobulin heavy chain variable region genes

Calibration

No study reported a measure of calibration for this model.

In the development study, treatment-free survival at three years was 95.3% in the low risk, 74.5% in the intermediate-risk and 28.6% in the high-risk group. Outcome frequencies were also provided per risk score. One external validation study showed similar percentages of treatment-free survival at three years with 95.5%, 78.9% and 40.6% for the low, intermediate and high-risk group, respectively (O-CLL1 V - Gentile 2016 (Mayo cohort)). The other external validation study presented results in the form of a Kaplan-Meier curve (O-CLL-1 V - Rani 2018 (Indian cohort)).

Discrimination

The c-statistic of the prognostic score in the model development study was 0.75 (P < 0.001). The c-statistic in the external validation cohorts were 0.72 (P < 0.001) (O-CLL1 V - Gentile 2016 (Mayo cohort)), and 0.55 (95% CI 0.49 to 0.60) (O-CLL-1 V - Rani 2018 (Indian cohort)).

Stephens model (Stephens 2015) (Stephens TFS D - Stephens 2015 (Ohio cohort))

To develop prognostic models for TFS, the authors identified significant prognostic factors in multivariable proportional hazards models using backwards selection. Based on the final model, they defined a simplified risk score based on the strength of the association. Missing data were accounted for by multiple imputation techniques. The model and score for TFS include the factors ECOG performance status, Rai stage, age, white blood cell count (WBC) and FISH category del(11q22.3). Risk groups were summarised as 0 to 1 points, 2 to 3 points and ≥ 4 points.

Calibration

No study reported a measure of calibration for the prognostic model for TFS.

In the development cohort, the percentage of persons who were treatment-free at two years was 85% (95% CI 0.60 to 0.95) for the low-risk group, 51% (95% CI 32 to 67) for the intermediate-risk group, and 0% for the high-risk group. In the validation cohort, the percentage of persons who were treatment-free at three years was 63% (95% CI 39 to 79) for the low-risk group, 26% (95% CI 15 to 39) for the intermediate-risk group and 16% (95% CI 6 to 29) for the high-risk group.

Discrimination

The discrimination of the exact model for TFS was 0.84 (P < 0.017) for the development cohort. The validation cohort showed a c-statistic of 0.66 for the simplified score for TFS.

Models without any external validation studies

The main characteristics of prognostic models for which no external validation studies could be found are described in Appendix 7.

Subgroup differences

We did not explore subgroup differences as the number of external validation studies per model with available data were too small. Three of the four external validation studies for the Barcelona-Brno model defined CLL by the International Workshop on CLL Guideline (Hallek 2008), and one did not report diagnostic criteria. Among the seven external validations of the CLL-IPI, three studies used the International Workshop on CLL Guideline (Hallek 2008), and four did not clearly refer to the diagnostic criteria or used a combination

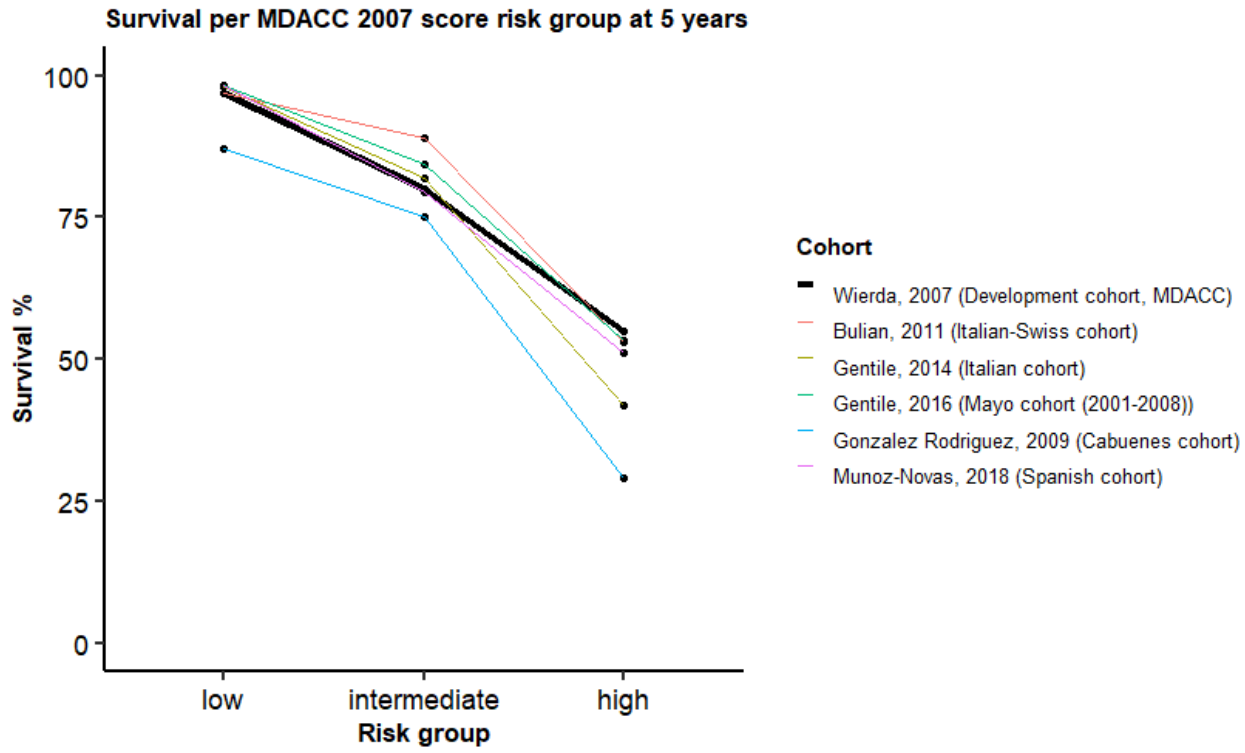
of criteria. One external validation study of the MDACC 2007 model used the International Workshop on CLL Guideline (Hallek 2008), two studies' cohorts were based on the NCI working group criteria (Cheson 1996), and four studies did not clearly report the criteria or used a combination of criteria.

Sensitivity analysis

We planned to explore the effect of risk of bias on the performance measures for the external validation studies per model. We were not able to conduct this analysis as we rated nearly all studies as high or unclear risk of bias, leaving no studies with a low risk of bias for inclusion in the sensitivity analysis.

One external validation study for the CLL-IPI reported the AUC instead of the c-statistic (CLL-IPI V - Zhu 2018 (Chinese cohort)), which we included in our meta-analysis. The pooled performance estimate for discrimination did not change (0.72, 95% CI 0.67 to 0.77 including the study; 0.72, 95% CI 0.66 to 0.78 excluding the study; see Figure 23). For the MDACC 2007 model, we also included one AUC instead of the c-statistic in the analysis. By removing this study (MDACC 2007 V - González Rodríguez (Cabueñes)), the pooled performance estimate for discrimination barely changed (from 0.65 to 0.66, 95% CI 0.60 to 0.71; for illustration, see Figure 29).

Figure 29. Sensitivity analysis of the c-statistic for the MDACC 2007 model excluding the study reporting AUC



To examine the effect of estimating the 95% CIs according to the Newcombe method, we conducted a sensitivity analysis exemplary for one included prognostic model, the CLL-IPI. In Figure 24, we limited the meta-analysis of the c-statistic to all studies that reported the 95% CI, which did not change the overall pooled estimate and 95% CIs substantially (from 0.72, 95% CI 0.67 to 0.77 for the main analysis to 0.73, 95% CI 0.66 to 0.80 for the analysis limited to studies with 95% CI reported). We calculated the 95% CIs based on Newcombe for all studies and the pooled results to compare this extreme scenario with the scenario that all 95% CIs were provided. Figure 25 shows that the estimation did not change the 95% CIs. The 95% PI changed negligibly.

For the CLL-IPI, one of the predictors was commonly missing (TP53 mutation) and replaced by a proxy prognostic factor (del(17p)). To see if the replacement limited discriminative performance, we conducted separate meta-analyses (Figure 26). The 95% CI of the pooled c-statistic overlapped, however, the number of studies in each analysis was very small (N = 3 and N = 4).

DISCUSSION

Summary of main results

We aimed to identify and describe all prognostic models for untreated CLL and their corresponding external validation studies. We identified 12 prognostic models with at least one external validation (Appendix 6). Of these, three models were validated externally more than three times, with appropriate reporting to allow us to conduct a meta-analysis of the c-statistic, which is a measure of discriminative performance of a model. Additionally, we identified 40 models or scores without any external validation study, which are further described in Appendix 7 and Characteristics of studies awaiting classification.

Results for models with more than three external validation studies per outcome for meta-analysis

Prognostic models for the prediction of OS (with meta-analysis, > 3 external validations)

The pooled c-statistic of the CLL-IPI, a point score calculated from five factors, was based on seven external validation studies with a total of 3307 individuals. The pooled estimate of its discriminative performance in seven external validation studies was 0.72 (95% CI 0.67 to 0.77). The 95% PI ranged from 0.59 to 0.83, indicating that from 100 external validation studies for this model, 95 will show a c-statistic within this range. Calibration was not reported in the publications, and thus the prognostic score was not updated based on calibration. Studies showed somewhat heterogeneous results regarding the pooled survival frequencies (observed survival frequencies) as compared to the development study survival frequencies (expected survival frequencies). Due to our diverging representation of calibration, to make assumptions regarding the calibration of the models, we assumed that the survival frequency of the development cohort can be interpreted as our expected frequency. The pooled survival per risk group was higher than observed in the development study for the very high-risk group, indicating that the model underpredicted survival in this group. This may be explained by the fact that the model was developed using data including a higher proportion of individuals with treatment indication at recruitment, who are participants with a worse prognosis than the general population with CLL. The curves showed a relatively homogeneous pattern regarding the differences between risk groups, although frequencies varied more widely for the very high-risk group, probably due to the low number of participants.

The pooled c-statistic of the Barcelona-Brno score, which contains two of the prognostic factors also included in the CLL-IPI, was derived from four external cohorts with a total of 1974 individuals. The pooled c-statistic was 0.64 (95% CI 0.60 to 0.67) and the 95% PI ranged from 0.59 and 0.68. Results between cohorts were somewhat heterogeneous, although the number of studies was relatively low to draw meaningful conclusions. Calibration was not reported in the publications, and thus the prognostic score was not updated based on calibration. The pooled survival per risk group was lower than observed in the development study, especially for the high-risk group, which means that across cohorts, the model overestimated survival as compared to the development cohort. This judgement is mainly driven by the Czech cohort. The pattern was consistent across studies.

The pooled c-statistic of the MDACC 2007 index score, which includes six prognostic factors, was derived from seven external validation studies with a total of 5127 individuals. The pooled c-statistic was 0.65 (95% CI 0.60 to 0.70; 95% PI: 0.51 to 0.77); discrimination was somewhat heterogeneous between studies. Calibration was not reported in any of the studies validating the index score, thus, no model updates were reported. The pooled survival per risk group matched the survival in the development study for the low and intermediate-risk groups. In the high-risk group, the pooled survival per risk group was lower than observed in the development study, thus the model overestimated the survival of the high-risk group. The pattern was consistent across studies.

From these three prognostic models for untreated individuals with CLL included in meta-analysis, the CLL-IPI performed best regarding the discrimination between persons with a 'as good as' compared to a worse prognosis. The score underpredicted survival for the very high-risk group, which was a small group of individuals defined by TP53 mutation or deletion.

The three models consist of unweighted (Barcelona-Brno D - Delgado 2017 (Barcelona cohort)) or weighted (CLL-IPI D - Bahlo 2016 (development cohort), MDACC 2007 D - Wierda 2007 (MDACC)) simplified point scores, which implies that information concerning the correct weighting of the prognostic factors was lost during simplification. Instead of individual outcome frequencies, the models result in a score on a limited scale, which reduced the possible outcome range compared to the original formula. Consequently, the simplification may have resulted in a reduction of the discriminative performance.

Prognostic models for the prediction of PFS or TFS (with meta-analysis, > 3 external validations)

We did not identify any prognostic model development studies that were externally validated more than three times for the outcomes PFS, TFS or TFS.

Models with one to three external validation studies

Prognostic models for the prediction of OS (without meta-analysis, 1 to 3 external validations)

In total, three models were developed to predict OS (GCLLSG D - Pflug 2014 (GCLLSG); Rossi D - Rossi 2013 (Italian cohort); Stephens OS D - Stephens 2015 (Ohio cohort)), with one, two and one external validation study, respectively. Due to the limited information on the predictive performance outside the model development setting, we would not yet consider these models ready for use in clinical practice.

Prognostic models for the prediction of PFS or TFS (without meta-analysis, 1 to 3 external validations)

In total, one model was developed to predict PFS (GIMEMA D - Molica 2005 (GIMEMA cohort)) and five were developed to predict TFS (Baliakas D - Baliakas 2019 (multicentre); MDACC 2011 D - Wierda 2011 (MDACC); Morabito D - Morabito 2009 (Italian cohort); Stephens TFS D - Stephens 2015 (Ohio cohort); O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)), with one external validation study per model for the first five and two for the last mentioned. Due to the limited information on the predictive performance outside the model development setting, we would not yet consider these models ready for use in clinical practice.

Models without any external validation studies

An abundance of prognostic models and scores without any application in external cohorts has been published, with a varying degree of reporting quality. Without any further knowledge on their performance in different settings than the development cohort, we would not recommend their use for patients.

Overall completeness, certainty of the evidence and study limitations of externally validated models

Overall completeness of the data

As expected from results of previous research (Heus 2018), reporting of the included models and their validation studies was poor. We have summarised the missing information per model for the development study and external validation studies for the primary model outcome in Appendix 8 .

For the CLL-IPI, none of the studies reported calibration (although we obtained information for five studies from corresponding authors). Reporting of discrimination was higher (8 out of 11, 6 times with 95% CI).

For the Barcelona-Brno score, none of the studies reported calibration (although we obtained information for one study from the corresponding authors). Reporting of discrimination was slightly higher (5 times out of 6, 3 times with 95% CI).

For the MDACC 2007 index score, two of the studies reported calibration. Reporting of discrimination was slightly higher (8 times out of 10, 7 times with 95% CI).

Of the other nine models with 17 validation studies that were not meta-analysed, none reported calibration (we obtained information on calibration for one study from the corresponding authors). Reporting of discrimination was higher (12 times, seven times with 95% CI).

Overall, across all categories of included studies (development studies with and without external validations, external validations), information on calibration and discrimination was lacking. Calibration was reported in six studies and discrimination in 37 out of 85 studies.

Many studies either excluded persons with missing prognostic factor data at enrolment, or excluded persons with missing data from the analysis. In most cases, we do not have information on the number of persons in the complete database or sample, and a comparison between included and excluded persons is rarely provided. Instead of removing data of individuals with missing data completely, multiple imputation would be preferable, but should be interpreted carefully (Moons 2019; Steyerberg 2014).

None of the studies explicitly mentioned reasons for censoring, a common phenomenon in time-to-event data, in their cohort which may be particularly relevant for studies based on RCT data.

Certainty of the evidence

At the date of submission of this review, no official GRADE guidance for grading the summarised results of meta-analysis of prognostic models was available. Hence, we refrained from rating the certainty of the evidence.

Study limitations of prognostic model development studies

The majority of PROBAST ratings for model development studies was high. The main reason for this was handling missing data, predictor selection based on univariable selection of predictors, failure to correct estimates for optimism, categorisation of prognostic factors and simplification of models.

Potential bias in the review process

To prevent bias in the review process, we performed all relevant steps in duplicate and solved discrepancies in group discussions. We developed a sensitive search strategy and tracked the references and citations of all included references. However, we did not search conference abstracts or trial registries, since conference abstracts do not provide sufficient information for 'Risk of bias' assessment and prognostic studies are rarely prospectively registered, but rather built on existing retrospective databases. Thus, we would not expect to find any relevant studies by searching those additional databases.

As this is a new review type, our methods evolved during the process. We decided to adapt the newest methods regardless of the protocol, which may have introduced bias.

- *Prespecification of prognostic factors*: we did not limit our inclusion to models with e.g. a minimum set of clinically relevant factors or a specific kind of factor (such as non-invasive or genetic factors only). Therefore, we made no distinction regarding which predictors were included in the retrieved models – we included all developed and validated models that were the subject of our review, regardless of which type of predictors they included. We did not aim to rate the individual factors, examine the relationship between any one factor with our outcomes or the strength of any one factor above other factors, as this would be the subject of a prognostic factor review.
- *Search*: we have slightly adapted the search to make it more inclusive, and extended it to include studies published since the inception of MEDLINE instead of using the cut-off year 1990. Based on the recommendation of the Cochrane fast-track service, we have added a search strategy for Embase. We believe that screening more references has not introduced bias.
- *Screening*: the definition of prognostic models and the reporting standards and, with that, the amount of information expected in a publication has changed substantially over time. To avoid bias, we have included all studies that explicitly stated the aim of developing a score for prognostication, although from today's point of view, we would not consider these scores a well-developed prediction model, because they did not follow the recommended steps in the model-building process. Most of these studies can be found in Appendix 7, as models without external validation studies.
- *Risk of bias*: since publication of the protocol, PROBAST was published. In our protocol, we did not specify the data analysis process. To avoid bias and use the most recently developed methods, we worked in close collaboration with the Cochrane Prognosis Methods group, and some authors of PROBAST are co-authors of this review. For transparency of our group decisions, we added the text under 'Solving disagreements'.
- *Analysis*: we planned to summarise the performance measures for calibration and discrimination. In the review we specified, according to most recent developments, that we would summarise the external validation studies per model. For some studies, we estimated the 95% CI for the c-statistic due to lack of reporting. Based on our sensitivity analysis, we assume that this estimation has not introduced bias.

We did not assess the likelihood of publication bias, i.e. a preferential publication of studies which show better model

performance, because currently, there is no established standard for this.

We would like to mention that external validation studies are ideally carried out by independent researchers (using independent samples). For the studies included in this review, this was not always the case (e.g. CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014); CLL-IPI V - Bahlo 2016 (SCAN cohort)), and may have led to additional bias to the one assessed by PROBAST.

Applicability of findings to clinical practice and policy

We identified 52 prognostic models for various outcomes. Due to the lack of external validation, we disregarded 40 of these as not (yet) applicable. Nine more prognostic models were validated only one to three times, which we would not consider to be sufficient information to make a judgement of applicability to different settings. The three models included in meta-analysis of the c-statistic offer more information concerning their performance in different populations. The CLL-IPI seems to perform best among the three models in both the development and external validation cohorts. However, due to the relatively small number of external validation studies (N < 10) and serious limitations in reporting, we cannot draw final conclusions regarding the generalisability and usefulness of this model in clinical practice.

The three models that were included in meta-analyses all predict OS. Until 2015, treatment options for individuals with CLL had not changed drastically over time. However, since the introduction of new agents, such as ibrutinib and idelalisib, survival has improved for all risk groups. For individuals with a poor prognosis, there is hope that they will benefit from new and better tailored treatments. Improvements in survival for all risk groups results in systematic underprediction of survival of prognostic models that were developed using data of patients receiving traditional treatment regimens, while the improvement of prognosis especially for high-risk individuals will lead to lowered discriminative ability of a prognostic model. The prognostic models discussed in detail in this review may thus need to be updated, taking this improved baseline risk into account. In contrast, other patient-relevant outcomes, such as time-to-treatment or progression would not be influenced by improved treatment options as long as a watch-and-wait strategy is adopted, and may be an interesting alternative outcome for prediction models. Some of the models were already externally validated for both of these outcomes, and may be included in a future update of this review.

A further limitation of prognostic models in general is that they usually consider only one outcome per model, thereby not considering a balance between different outcomes, such as survival, quality of life, and other outcomes of interest.

CLL has experienced several changes in diagnostic criteria over time (in particular, in 2008 the criterion of having an absolute lymphocyte count (ALC) more than or equal to 5.0×10^9/L[2] to a B-cell count more than or equal to 5.0×10^9/L), changing the categorisation of individuals with clinical monoclonal lymphocytosis (cMBL) or CLL (Cheson 1996; Hallek 2008). We aimed to explore this change in subgroup analysis but did not, due to the low number of included studies with clear division. However, this change is not expected to have an extensive effect on the performance of a model, as prognosis and patient characteristics are comparable in individuals with cMBL and Rai stage I.

Agreements and disagreements with other studies or reviews

To our knowledge, the current review is the first systematic review that aimed to identify all prognostic models developed for untreated individuals with CLL and their corresponding external validation studies.

We are aware of two other systematic reviews that cover one of the prognostic models that we identified, the CLL-IPI (Molica 2018a; Molica 2018b). These systematic reviews aimed at identifying all published studies that have used CLL-IPI to predict the clinical outcome of CLL; one in patients who received chemoimmunotherapy or targeted therapies (Molica 2018a), the other without restriction of therapy (Molica 2018b). There are some differences between these reviews and our systematic review. In Molica 2018a, studies were limited to individuals who received chemoimmunotherapy or targeted therapies and also included studies of relapsed/refractory CLL from conference abstracts. The authors summarised the survival frequencies per CLL-IPI risk group at two years for OS, whereas we pooled OS at five years. Molica 2018b included patients undergoing any treatment, and summarised both OS and TFS at five years. Both these reviews included data from conference proceedings, whereas we did not. Moreover, they did not report measures of discrimination such as the c-statistic (which is pivotal when assessing model performance) and did not assess risk of bias or applicability.

The included studies in Molica 2018b mostly overlapped with our identified studies for OS. The exception was two external validation studies published as abstracts (N = 2, CLL11 study population; Goede 2016 and Ferrer Lores 2016). In contrast, we identified three additional cohorts which had not yet been published at the time of publication of Molica 2018a and Molica 2018b (CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort); CLL-IPI V - Rani 2018 (Indian cohort); CLL-IPI V - Zhu 2018 (Chinese cohort)).

AUTHORS' CONCLUSIONS

Implications for practice

We identified three prognostic models that were validated more than three times for their primary outcome OS: the CLL-IPI, the Barcelona-Brno score, and the MDACC 2007 index score. Of the three models, the CLL-IPI seems to perform best regarding the discrimination between individuals with a good prognosis as compared to individuals with a worse prognosis. However, the number of external validation studies was relatively low to draw definite conclusions. Especially for the very high-risk group, the index underpredicts survival, which may be a result of the selective sample that was used to develop the index.

This review has been developed in a time of rapidly expanding treatment options that may drastically change the prognostic implications for the therapy and longevity of these patients. The external validation cohorts may not be representative anymore of the currently available, improved treatment options. Therefore, the models can be used to provide an approximate classification, but may need testing and updating before being applied to new patients.

Implications for research

This systematic review shows that for CLL, an abundance of studies of prognostic models and scores can be identified. Based on our very inclusive definition of a prognostic score, we found that these studies (including the newer studies) were not performed according to the current standards of prognostic model development. For the future, we recommend authors of model development studies incorporate the most recent checklists and tools, such as CHARMS and TRIPOD, as orientation to avoid common issues. We would also like to emphasise that reporting in more recent papers can benefit from improvements. The minimum reported information should include participant information (recruitment, study design, eligibility criteria, diagnostic criteria used, etc.), predictor assessment, outcome definition, and the relevant performance measures.

Our review also shows that the proportion of external model validation studies as compared to model development studies is very low. Before investing in the development of a new model, we recommend the validation and direct comparison of the existing models in different settings and populations. Most suitable for this aim are well-designed prospective studies that also take into account novel agents and emerging molecular makers. To improve predictive performance, a model can be updated with these markers or tailored to a specific population.

With time, newly identified prognostic factors may be added.

ACKNOWLEDGEMENTS

The research was supported by National Health Service (NHS) Blood and Transplant and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to thank Sarah Hodgkinson (Associate Editor, Cochrane Editorial and Methods Department), Helen Wakeford and Clare Dooley (Managing Editors, Cochrane Editorial and Methods Department), and Anne Lethaby (Copy Editor, Copy Edit Support) for their assistance regarding the review. We would also like to thank Robin Featherstone (Information Specialist, Cochrane Editorial and Methods Department) for suggestions to improve our search strategy.

We also thank Alexandra McAleenan (Methods), Lucinda Archer (Methods), Sherry Reisner (Health Writer), Dr Stefano Molica (Department of Hematology-Oncology, Azienda Ospedaliera Pugliese-Ciaccio, Catanzaro, Italy), Dr Mary Ann Anderson (Clinical Haematology Department of The Royal Melbourne Hospital and Peter MacCallum Cancer Centre; Blood Cells and Blood Cancer Division of the Walter and Eliza Hall Institute, Melbourne, Australia) and Dr SG Agrawal (Senior Lecturer and Honorary Consultant, Division of Haemato-Oncology, St Bartholomew's Hospital, Barts Health NHS Trust; and Centre for Immunobiology, Blizard Institute, Queen Mary, University of London) for their valuable peer review.

We would like to thank Thomas Debray from the Cochrane Prognosis Methods group for statistical advice and Natali Pflug for clinical advice.

We would like to thank Yuan Chi (Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine) for the translation of a Chinese publication (CLL-IPI V - Zhu 2018 (Chinese cohort)) and Leonardo Perales-Guerrero (medical student, Universidad de Guadalajara, Mexico) for the translation of a Spanish publication (GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)).

We thank all principal investigators who have replied to our inquiries and/or provided us with further information or data concerning their publications. Their help enabled us to include as many studies as possible in analysis. We want to thank Massimo Gentile, Caspar da Cunha-Bang, Carolina Muñoz-Novas, Lata Rani, Gian Matteo Rigolin, Huayuan Zhu, Julio Delgado, Sanja Trajkova, Natali Pflug, Jasmin Bahlo, Pietro Bulian, and Neil E Kay.

REFERENCES

References to studies included in this review

Baliakas D - Baliakas 2019 (multicentre) *{published data only}*

Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica* 2019;**104**(2):360-9.

Baliakas V - Baliakas 2019 (MLL + Scan.) *{published data only}*

Baliakas P, Moysiadis T, Hadzidimitriou A, Xochelli A, Jeromin S, Agathangelidis A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica* 2019;**104**(2):360-9.

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) *{published data only}*

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. American Journal of Hematology 2017;**92**(4):375-80.

Barcelona-Brno V - Delgado 2017 (Brno cohort) *{published data only}*

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. American Journal of Hematology 2017;**92**(4):375-80.

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) *{published data only}*

Gentile M, Shanafelt TD, Mauro FR, Laurenti L, Rossi D, Molica S, et al. Comparison between the CLL-IPI and the Barcelona-Brno prognostic model: analysis of 1299 newly diagnosed cases. American Journal of Hematology 2017;**93**(2):E35-7.

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) *{published data only}*

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. American Journal of Hematology 2017;**92**(4):375-80.

* Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. American Journal of Hematology 2017;**92**(6):E91-3.

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) *{published data only}*

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al. The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Brno biomarkers only prognostic model and the MD Anderson Cancer Center Prognostic Index. Biomed Research International 2018:9506979. [DOI: 10.1155/2018/9506979]

Barcelona-Brno V - Rani 2018 (Indian cohort) *{published data only}*

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

Barcelona-Brno V - Reda 2017 (Milan cohort) *{published data only}*

Reda G, Cassin R, Fattizzo B, Giannarelli D, Mattiello V, Barcellini W, et al. Chronic lymphocytic leukemia and prognostic models: a bridge between clinical and biological markers. *American Journal of Hematology* 2017;**92**(7):E135-7.

CLL-IPI D - Bahlo 2016 (development cohort) *{published data only}*

International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;**17**(6):779-90.

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) *{published data only}*

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

* International CLL-IPI Working Group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;**17**(6):779-90.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;**91**:1090-5.

CLL-IPI V - Bahlo 2016 (SCAN cohort) *{published data only}*

International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncology* 2016;**17**(6):779-90.

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) *{published data only}*

Da Cunha-Bang C, Christiansen I, Niemann CU. The CLL-IPI applied in a population-based cohort. *Blood* 2016;**128**(17):2181-3.

CLL-IPI V - Delgado 2017 (Barcelona cohort) *{published data only}*

Delgado J, Doubek M, Baumann T, Kotaskova J, Molica S, Mozas P, et al. Chronic lymphocytic leukemia: a prognostic model comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. American Journal of Hematology 2017;**92**(4):375-80.

CLL-IPI V - Gentile 2016 (Italian cohort) *{published data only}*

Gentile M, Shanafelt TD, Mauro FR, Laurenti L, Rossi D, Molica S, et al. Comparison between the CLL-IPI and the Barcelona-Brno prognostic model: analysis of 1299 newly diagnosed cases. American Journal of Hematology 2017;**93**(2):E35-7.

* Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

CLL-IPI V - Molica 2016 (O-CLL1-GISL) *{published data only}*

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. American Journal of Hematology 2017;**92**(6):E91-3.

* Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;**91**(11):1090-5.

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) *{published data only}*

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al. The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Brno biomarkers only prognostic model and the MD Anderson Cancer Center Prognostic Index. Biomed Research International 2018:9506979. [DOI: 10.1155/2018/9506979]

CLL-IPI V - Rani 2018 (Indian cohort) *{published data only}*

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

CLL-IPI V - Reda 2017 (Milano cohort) *{published data only}*

Reda G, Cassin R, Fattizzo B, Giannarelli D, Mattiello V, Barcellini W, et al. Chronic lymphocytic leukemia and prognostic models: a bridge between clinical and biological markers. *American Journal of Hematology* 2017;**92**(7):E135-7.

CLL-IPI V - Rigolin 2017 (Ferrera cohort) *{published data only}*

Rigolin GM, Cavallari M, Quaglia FM, Formigaro L, Lista E, Urso A, et al. In CLL, comorbidities and the complex karyotype are associated with an inferior outcome independently of CLL-IPI. *Blood* 2017;**129**(26):3495-8.

CLL-IPI V - Zhu 2018 (Chinese cohort) *{published data only}*

Zhu HY, Wang L, Qiao J, Zou YX, Xia Y, Wu W, et al. Prognostic significance of CLL-IPI for Chinese patients with chronic lymphocytic leukemia. *Chung Hua Hsueh Yeh Hsueh Tsa Chi* 2018;**39**:392-7.

GCLLSG D - Pflug 2014 (GCLLSG) *{published data only}*

* Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62.

Tam CS, Seymour JF. A new prognostic score for CLL. *Blood* 2014;**124**(1):1-2.

GCLLSG V - Molica 2015 (O-CLL1-GISL) *{published data only}*

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. The utility of two prognostic models for predicting time to first treatment in early chronic lymphocytic leukemia patients: results of a comparative analysis. *Leukemia Research* 2013;**37**(8):943-7.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

* Molica S, Giannarelli D, Mirabelli R, Levato L, Russo A, Linardi M, et al. Unavailability of thymidine kinase does not preclude the use of German comprehensive prognostic index: results of an external validation analysis in early chronic lymphocytic leukemia and comparison with MD Anderson Cancer Center model. *European Journal of Haematology* 2015;**96**(1):72-7.

GCLLSG V - Pflug 2014 (Mayo cohort) *{published data only}*

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62.

GCLLSG V - Rani 2018 (Indian cohort) *{published data only}*

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic

lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

GIMEMA D - Molica 2005 (GIMEMA cohort) *{published data only}*

Molica S, Mauro FR, Callea V, Gentile M, Giannarelli D, Lopez M, et al. A gender-based score system predicts the clinical outcome of patients with early B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2005;**46**(4):553-60.

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) *{published data only}*

Gonzalez Rodriguez AP, Gonzalez Garcia E, Fernandez Alvarez C, Gonzalez Huerta AJ, Gonzalez Rodriguez S. B-chronic lymphocytic leukemia: epidemiological study and comparison of MDACC and GIMENA prognostic indexes [Estudio epidemiológico y comparación de los índices pronósticos del MD Anderson Cancer Center y el índice del Gruppo Italiano Malattie Ematologiche Maligne dell' Adulto en pacientes con leucemia linfática crónica de células B]. *Medicina Clinica* 2009;**133**(5):161-6.

MDACC 2007 D - Wierda 2007 (MDACC) *{published data only}*

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood* 2007;**109**(11):4679-85.

MDACC 2007 V - Bulian 2011 (Italian-Swiss) *{published data only}*

Bulian P, Tarnani M, Rossi D, Forconi F, Del Poeta G, Bertoni F, et al. Multicentre validation of a prognostic index for overall survival in chronic lymphocytic leukaemia. *Hematological Oncology* 2011;**29**(2):91-9.

MDACC 2007 V - Gentile 2014 (Italian cohort) *{published data only}*

* Gentile M, Mauro FR, Rossi D, Vincelli I, Tripepi G, Recchia AG, et al. Italian external and multicentric validation of the MD Anderson Cancer Center nomogram and prognostic index for chronic lymphocytic leukaemia patients: analysis of 1502 cases. *British Journal of Haematology* 2014;**167**:224-32.

Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

MDACC 2007 V - Gentile 2016 (Mayo cohort) *{published data only}*

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

* Gentile M, Shanafelt TD, Rossi D, Laurenti L, Mauro FR, Molica S, et al. Validation of the CLL-IPI and comparison with the MDACC prognostic index in newly diagnosed patients. *Blood* 2016;**128**(16):2093-5.

Shanafelt TD, Jenkins G, Call TG, Zent CS, Slager S, Bowen DA, et al. Validation of a new prognostic index for patients with chronic lymphocytic leukemia. *Cancer* 2009;**115**:363-72.

MDACC 2007 V - González Rodríguez (Cabueñes) *{published data only}*

Gonzalez Rodriguez AP, Gonzalez Garcia E, Fernandez Alvarez C, Gonzalez Huerta AJ, Gonzalez Rodriguez S. B-chronic lymphocytic leukemia: epidemiological study and comparison of MDACC and GIMENA prognostic indexes [Estudio epidemiológico y comparación de los índices pronósticos del MD Anderson Cancer Center y el índice del Gruppo Italiano Malattie Ematologiche Maligne dell' Adulto en pacientes con leucemia linfática crónica de células B]. *Medicina Clinica* 2009;**133**(5):161-6.

MDACC 2007 V - Molica 2010 (GIMEMA cohort) *{published data only}*

Molica S, Mauro FR, Callea V, Giannarelli D, Lauria F, Rotoli B, et al. The utility of a prognostic index for predicting time to first treatment in early chronic lymphocytic leukemia: the GIMEMA experience. *Haematologica* 2010;**95**(3):464-9.

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) *{published data only}*

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

Molica S, Di Raimondo F, Cutrona G, Fabris S, Mauro F, Brugiatelli M, et al. Clinical categories identified by a new prognostic index reflect biological characteristics of patients in early chronic lymphocytic leukemia: the Gruppo Italiano Studio Linfomi (GISL) experience. *Leukemia Research* 2010;**34**(8):e217-8.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

* Molica S, Giannarelli D, Mirabelli R, Levato L, Russo A, Linardi M, et al. Unavailability of thymidine kinase does not preclude the use of German comprehensive prognostic index: results of an external validation analysis in early chronic lymphocytic leukemia and comparison with MD Anderson Cancer Center model. *European Journal of Haematology* 2015;**96**(1):72-7.

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) *{published data only}*

Munoz-Novas C, Poza-Santaella M, Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Infante MS, et al. The international prognostic index for patients with chronic lymphocytic leukemia has the higher value in predicting overall outcome compared with the Barcelona-Brno biomarkers only prognostic model and the MD Anderson Cancer Center

Prognostic Index. Biomed Research International 2018:9506979. [DOI: 10.1155/2018/9506979]

MDACC 2007 V - Pflug 2014 (3 RCTs) *{published data only}*

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62.

MDACC 2007 V - Rani 2018 (Indian cohort) *{published data only}*

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

MDACC 2007 V - Trajkova 2013 (Macedonia) *{published data only}*

Trajkova S, Cevreska L, Pivkova-Veljanovska A, Ivanovski M, Dukovski D, Popova-Simjanovska M, et al. Multivariable model consisting of clinical and biological markers for time to first treatment in CLL patients: preliminary results from single centre experience. *Makedonska Akademija na Naukite i Umetnostite Oddelenie Za Bioloshki i Meditsinski Nauki Prilozi* 2013;**34**(3):39-48.

MDACC 2011 D - Wierda 2011 (MDACC) *{published data only}*

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2011;**29**(31):4088-95.

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) *{published data only}*

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. External validation on a prospective basis of a nomogram for predicting the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer* 2013;**119**(6):1177-85.

Molica S, Giannarelli D, Gentile M, Cutrona G, Di Renzo N, Di Raimondo F, et al. The utility of two prognostic models for predicting time to first treatment in early chronic lymphocytic leukemia patients: results of a comparative analysis. *Leukemia Research* 2013;**37**(8):943-7.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Lentini M, et al. A prognostic algorithm including a modified version of MD Anderson Cancer Center (MDACC) score predicts time to first treatment of patients with clinical monoclonal lymphocytosis (cMBL)/Rai stage 0 chronic lymphocytic leukemia (CLL). *International Journal of Hematology* 2014;**100**(3):290-5.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

* Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;**91**(11):1090-5.

Morabito D - Morabito 2009 (Italian cohort) *{published data only}*

Morabito F, Cutrona G, Gentile M, Matis S, Todoerti K, Colombo M, et al. Definition of progression risk based on combinations of cellular and molecular markers in patients with Binet stage A chronic lymphocytic leukaemia. *British Journal of Haematology* 2009;**146**(1):44-53.

Morabito V - Gentile 2014 (O-CLL1-GISL) *{published data only}*

Gentile M, Cutrona G, Mosca L, Matis S, Fabris S, Lionetti M, et al. Prospective validation of a risk score based on biological markers for predicting progression free survival in Binet stage A chronic lymphocytic leukemia patients: results of the multicenter O-CLL1-GISL study. *American Journal of Hematology* 2014;**89**(7):743-50.

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) *{published data only}*

* Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

Molica S, Giannarelli D, Levato L, Mirabelli R, Gentile M, Morabito F. Assessing time to first treatment in early chronic lymphocytic leukemia (CLL): a comparative performance analysis of five prognostic models with inclusion of CLL-international prognostic index (CLL-IPI). *Leukemia & Lymphoma* 2017;**58**(7):1736-9.

Molica S, Giannarelli D, Mirabelli R, Levato L, Gentile M, Morabito F, et al. Reliability of six prognostic models to predict time-to-first-treatment in patients with chronic lymphocytic leukaemia in early phase. *American Journal of Hematology* 2017;**92**(6):E91-3.

Molica S, Shanafelt TD, Giannarelli D, Gentile M, Mirabelli R, Cutrona G, et al. The chronic lymphocytic leukemia international prognostic index predicts time to first treatment in early CLL: independent validation in a prospective cohort of early stage patients. *American Journal of Hematology* 2016;**91**(11):1090-5.

O-CLL1 V - Gentile 2016 (Mayo cohort) *{published data only}*

Gentile M, Shanafelt TD, Cutrona G, Molica S, Tripepi G, Alvarez I, et al. A progression-risk score to predict treatment-free survival for early stage chronic lymphocytic leukemia patients. *Leukemia* 2016;**30**(6):1440-3.

O-CLL-1 V - Rani 2018 (Indian cohort) *{published data only}*

Rani L, Gogia A, Singh V, Kumar L, Sharma A, Kaur G, et al. Comparative assessment of prognostic models in chronic lymphocytic leukemia: evaluation in Indian cohort. *Annals of Hematology* 2018;**98**(2):437-43.

Rossi D - Rossi 2013 (Italian cohort) *{published data only}*

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013;**121**(8):1403-12.

Rossi V - Jeromin 2014 (Munich cohort) *{published data only}*

Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* 2014;**28**(1):108-17.

Rossi V - Rossi 2013 (unclear) *{published data only}*

Rossi D, Rasi S, Spina V, Bruscaggin A, Monti S, Ciardullo C, et al. Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* 2013;**121**(8):1403-12.

Stephens OS D - Stephens 2015 (Ohio cohort) *{published data only}*

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;**90**(11):967-9.

Stephens OS V - Stephens 2015 (MDACC) *{published data only}*

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;**90**(11):967-9.

Stephens TFS D - Stephens 2015 (Ohio cohort) *{published data only}*

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients. *American Journal of Hematology* 2015;**90**(11):967-9.

Stephens TFS V - Stephens 2015 (MDACC) *{published data only}*

Stephens DM, Ruppert AS, Weirda WG, Jones JA, Woyach JA, Maddocks K, et al. Externally validated predictive clinical model for untreated del(17p13.1) chronic lymphocytic leukemia patients.. *American Journal of Hematology* 2015;**90**(11):967-9.

References to studies excluded from this review

Apelgren 2006 *{published data only}*

Apelgren P, Hasselblom S, Werlenius O, Nilsson-Ehle H, Andersson PO. Evaluation of clinical staging in chronic lymphocytic leukemia - population-based study. *Leukemia & Lymphoma* 2006;**47**(12):2505-16.

Baccarani 1982 *{published data only}*

Baccarani M, Cavo M, Gobbi M, Lauria F, Tura S. Staging of chronic lymphocytic leukemia. *Blood* 1982;**59**(6):1191-6.

Baliakas 2015 *{published data only}*

Baliakas P, Agathangelidis A, Hadzidimitriou A, Sutton LA, Minga E, Tsanousa A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood* 2015;**125**(5):856-9.

Berke 2019 *{published data only}*

Berke Mentese I, Yegin ZA, Gokcen S, Ozkurt ZN, Yagci M. Prognostic significance of serum BAFF, APRIL, TACI and BCMA levels in chronic lymphocytic leukemia. *Indian Journal of Hematology and Blood Transfusion* 2019;**35**(2):265-71.

Bettini 1986 *{published data only}*

Bettini R, Rapazzini P, Ferrari V, Steidl L. Clinical staging of chronic lymphatic leukemia. *Recenti Progressi in Medicina* 1986;**77**(2):72-6.

Binet 1977 *{published data only}*

Binet JL, Leporrier M, Dighiero G, Charron D, D'Athis P, Vaugier G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer* 1977;**40**(2):855-64.

Binet 1981 *{published data only}*

Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer 1981;**48**(1):198-206.

Bo 2014 *{published data only}*

Bo MD, Del Principe MI, Pozzo F, Ragusa D, Bulian P, Rossi D, et al. NOTCH1 mutations identify a chronic lymphocytic leukemia patient subset with worse prognosis in the setting of a rituximab-based induction and consolidation treatment. *Annals of Hematology* 2014;**93**(10):1765-74.

Bomben 2005 *{published data only}*

Bomben R, Dal Bo M, Zucchetto A, Zaina E, Nanni P, Sonego P, et al. Mutational status of IgV(H) genes in B-cell chronic lymphocytic leukemia and prognosis: percent mutations or antigen-driven selection? *Leukemia* 2005;**19**(8):1490-2.

Bomben 2009 *{published data only}*

Bomben R, Dal Bo M, Capello D, Forconi F, Maffei R, Laurenti L, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *British Journal of Haematology* 2009;**144**(4):492-506.

Bou Samra 2014 *{published data only}*

Bou Samra E, Klein B, Commes T, Moreaux J. Identification of a 20-gene expression-based risk score as a predictor of clinical outcome in chronic lymphocytic leukemia patients. *Biomed Research International* 2014;**2014**:423174.

Brejcha 2010 *{published data only}*

Brejcha M, Doubek M, Cmunt E, Schwarz J, Pospisil Z, Kozmon P, et al. New prognostic markers of chronic lymphocytic leukemia (CLL) in the everyday hematological practice. An analysis of data from four hematology departments. [Czech]. *Transfuze a Hematologie Dnes* 2010;**16**(Suppl 1):62-6.

Brugiatelli 2007 *{published data only}*

Brugiatelli M, Loteta B, Nocilli L, Mannina D. B-cell chronic lymphocytic leukemia: clinical impact of biological prognostic factors and updated treatment strategies. *Lijecnicki Vjesnik* 2007;**129 Suppl 3**:26-8.

Bulian 2014 *{published data only}*

Bulian P, Shanafelt TD, Fegan C, Zucchetto A, Cro L, Nuckel H, et al. CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2014;**32**(9):897-904.

Byrd 2006 *{published data only}*

Byrd JC, Gribben JG, Peterson BL, Grever MR, Lozanski G, Lucas DM, et al. Select high-risk genetic features predict earlier progression following chemoimmunotherapy with fludarabine and rituximab in chronic lymphocytic leukemia: justification for risk-adapted therapy. *Journal of Clinical Oncology* 2006;**24**(3):437-43.

Cailliod 2005 *{published data only}*

Cailliod R, Quantin C, Carli PM, Jooste V, Le Teuff G, Binquet C, et al. A population-based assessment of the prognostic value of the CD19 positive lymphocyte count in B-cell chronic lymphocytic leukemia using Cox and Markov models. *European Journal of Epidemiology* 2005;**20**(12):993-1001.

Callea 1999 *{published data only}*

Callea V, Morabito F, Oliva BM, Stelitano C, Levato D, Dattilo A, et al. Surface CD14 positivity in B-cell chronic lymphocytic leukaemia is related to clinical outcome. *British Journal of Haematology* 1999;**107**(2):347-52.

Catovsky 1989 *{published data only}*

Catovsky D, Fooks J, Richards S, MRC Working Party on Leukaemia in Adults. Prognostic factors in chronic lymphocytic leukaemia: the importance of age, sex and response to treatment in survival. A report from the MRC CLL 1 trial. *British Journal of Haematology* 1989;**72**(2):141-9.

Cesano 2013 *{published data only}*

Cesano A, Perbellini O, Evensen E, Chu CC, Cioffi F, Ptacek J, et al. Association between B-cell receptor responsiveness and disease progression in B-cell chronic lymphocytic leukemia: results from single cell network profiling studies. *Haematologica* 2013;**98**(4):626-34.

Chang 2003 *{published data only}*

Chang CC, Liu CZ, Cleveland RP. Relative importance of CD38 expression over myeloid-associated markers expression in predicting the clinical course of B-CLL patients. *Leukemia and Lymphoma* 2003;**44**(6):977-82.

Chastang 1985 *{published data only}*

Chastang C, Travade P, Auquier A. Critical discussion of the assessment of a three-stage prognostic classification for chronic lymphocytic leukemia. *Statistics in Medicine* 1985;**4**(3):287-93.

Chauzeix 2018 *{published data only}*

Chauzeix J, Laforet MP, Deveza M, Crowther L, Marcellaud E, Derouault P, et al. Normal serum protein electrophoresis and mutated IGHV genes detect very slowly evolving chronic lymphocytic leukemia patients. *Cancer Medicine* 2018;**7**(6):2621–8.

Chelazzi 1979 *{published data only}*

Chelazzi G, Bettini R, Michetti A, Ricci G. Prognostic value of the clinical staging of chronic lymphocytic leukemia (author's transl). *Haematologica* 1979;**64**(4):463-71.

Chen 1997 *{published data only}*

Chen PM, Lin SH, Fan SF, Chiou TJ, Hsieh RK, Yu IT, et al. Genotypic characterization and multivariate survival analysis of chronic lymphocytic leukemia in Taiwan. *Acta Haematologica* 1997;**97**(4):196-204.

Chena 2008 *{published data only}*

Chena C, Avalos JS, Bezares RF, Arrossagaray G, Turdo K, Bistmans A, et al. Biallelic deletion 13q14.3 in patients with chronic lymphocytic leukemia: cytogenetic, FISH and clinical studies. *European Journal of Haematology* 2008;**81**(2):94-9.

Chevallier 2002 *{published data only}*

Chevallier P, Penther D, Avet-Loiseau H, Robillard N, Ifrah N, Mahe B, et al. CD38 expression and secondary 17p deletion are important prognostic factors in chronic lymphocytic leukaemia. *British Journal of Haematology* 2002;**116**(1):142-50.

Chiaretti 2014 *{published data only}*

Chiaretti S, Marinelli M, Del Giudice I, Bonina S, Piciocchi A, Messina M, et al. NOTCH1, SF3B1, BIRC3 and TP53 mutations in patients with chronic lymphocytic leukemia undergoing first-line treatment: correlation with biological parameters and response to treatment. *Leukemia & Lymphoma* 2014;**55**(12):2785-92.

Christiansen 1994 *{published data only}*

Christiansen I, Gidlof C, Wallgren AC, Simonsson B, Totterman TH. Serum levels of soluble intercellular adhesion molecule 1 are increased in chronic B-lymphocytic leukemia and correlate with clinical stage and prognostic markers. *Blood* 1994;**84**(9):3010-6.

Chuang 2012 *{published data only}*

Chuang HY, Rassenti L, Salcedo M, Licon K, Kohlmann A, Haferlach T, et al. Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood* 2012;**120**(13):2639-49.

Ciccone 2012 *{published data only}*

Ciccone M, Agostinelli C, Rigolin GM, Piccaluga PP, Cavazzini F, Righi S, et al. Proliferation centers in chronic lymphocytic leukemia: correlation with cytogenetic and clinicobiological features in consecutive patients analyzed on tissue microarrays. *Leukemia* 2012;**26**(3):499-508.

Ciocoiu 1988 *{published data only}*

Ciocoiu A, Motoiu I, Niculescu R, Berceanu S. The staging of chronic lymphatic leukemia. Its prognostic value and therapeutic implications. *Revista de Medicină Internă, Neurologe, Psihiatrie, Neurochirurgie, Dermato-venerologie* 1988;**40**(1):41-5.

Claus 2012 *{published data only}*

Claus R, Lucas DM, Stilgenbauer S, Ruppert AS, Yu L, Zucknick M, et al. Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2012;**30**(20):2483-91.

Claus 2014 *{published data only}*

Claus R, Lucas DM, Ruppert AS, Williams KE, Weng D, Patterson K, et al. Validation of ZAP-70 methylation and its relative significance in predicting outcome in chronic lymphocytic leukemia. *Blood* 2014;**124**(1):42-8.

Cmunt 2002 *{published data only}*

Cmunt E, Michalova K, Sindelarova L, Karban J, Zemanova Z, Kurkova S, et al. Importance of prognostic factors in patients with chronic B-lymphocytic leukemia at the time of diagnosis. *Sbornik Lekarsky* 2002;**103**(3):359-70.

Cocco 2005 *{published data only}*

Cocco AE, Osei ES, Thut DM, Edinger AK, Powers JJ, Fu P, et al. Bimodal cell populations are common in chronic lymphocytic leukemia but do not impact overall survival. *American Journal of Clinical Pathology* 2005;**123**(6):818-25.

Corcoran 2005 *{published data only}*

Corcoran M, Parker A, Orchard J, Davis Z, Wirtz M, Schmitz OJ, et al. ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia. *Haematologica* 2005;**90**(8):1078-88.

Cordone 1998 *{published data only}*

Cordone I, Masi S, Mauro FR, Soddu S, Morsilli O, Valentini T, et al. p53 expression in B-cell chronic lymphocytic leukemia: a marker of disease progression and poor prognosis. *Blood* 1998;**91**(11):4342-9.

Cortese 2014 *{published data only}*

Cortese D, Sutton LA, Cahill N, Smedby KE, Geisler C, Gunnarsson R, et al. On the way towards a 'CLL prognostic index': focus on TP53, BIRC3, SF3B1, NOTCH1 and MYD88 in a population-based cohort. *Leukemia* 2014;**28**(3):710-3.

Coscia 2012 *{published data only}*

Coscia M, Vitale C, Peola S, Foglietta M, Rigoni M, Griggio V, et al. Dysfunctional V9V2 T cells are negative prognosticators and markers of dysregulated mevalonate pathway activity in chronic lymphocytic leukemia cells. *Blood* 2012;**120**(16):3271-9.

Crespo 2003 *{published data only}*

Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *New England Journal of Medicine* 2003;**348**(18):1764-75.

Criel 1997 *{published data only}*

Criel A, Verhoef G, Vlietinck R, Mecucci C, Billiet J, Michaux L, et al. Further characterization of morphologically defined typical and atypical CLL: a clinical, immunophenotypic, cytogenetic and prognostic study on 390 cases. *British Journal of Haematology* 1997;**97**(2):383-91.

Cro 2009 *{published data only}*

Cro L, Morabito F, Zucal N, Fabris S, Lionetti M, Cutrona G, et al. CD26 expression in mature B-cell neoplasia: its possible role as a new prognostic marker in B-CLL. *Hematological Oncology* 2009;**27**(3):140-7.

Cro 2010 *{published data only}*

Cro L, Ferrario A, Lionetti M, Bertoni F, Zucal NN, Nobili L, et al. The clinical and biological features of a series of immunophenotypic variant of B-CLL. *European Journal of Haematology* 2010;**85**(2):120-9.

Cuneo 2004 *{published data only}*

Cuneo A, Rigolin GM, Bigoni R, De Angeli C, Veronese A, Cavazzini F, et al. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. *Leukemia* 2004;**18**(3):476-83.

Cuneo 2018 *{published data only}*

Cuneo A, Follows G, Rigolin GM, Piciocchi A, Tedeschi A, Trentin L, et al. Efficacy of bendamustine and rituximab as first salvage treatment in chronic lymphocytic leukemia and indirect comparison with ibrutinib: a GIMEMA, ERIC and UK CLL FORUM study. *Haematologica* 2018;**103**(7):1209-17.

D'Arena 2001 *{published data only}*

D'Arena G, Musto P, Cascavilla N, Dell'Olio M, Di Renzo N, Perla G, et al. CD38 expression correlates with adverse biological features and predicts poor clinical outcome in B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2001;**42**(1-2):109-14.

D'Arena 2007 *{published data only}*

D'Arena G, Tarnani M, Rumi C, Vaisitti T, Aydin S, De Filippi R, et al. Prognostic significance of combined analysis of ZAP-70 and CD38 in chronic lymphocytic leukemia. *American Journal of Hematology* 2007;**82**(9):787-91.

D'Arena 2012 *{published data only}*

D'Arena G, D'Auria F, Simeon V, Laurenti L, Deaglio S, Mansueto G, et al. A shorter time to the first treatment may be predicted by the absolute number of regulatory T-cells in patients with Rai stage 0 chronic lymphocytic leukemia. *American Journal of Hematology* 2012;**87**(6):628-31.

Damle 1999 *{published data only}*

Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 1999;**94**(6):1840-7.

Dasgupta 2015 *{published data only}*

Dasgupta A, Mahapatra M, Saxena R. A study for proposal of use of regulatory T cells as a prognostic marker and establishing an optimal threshold level for their expression in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2015;**56**(6):1831-8.

Davis 2016 *{published data only}*

Davis Z, Forconi F, Parker A, Gardiner A, Thomas P, Catovsky D, et al. The outcome of Chronic lymphocytic leukaemia patients with 97% IGHV gene identity to germline is distinct from cases with < 97% identity and similar to those with 98% identity. *British Journal of Haematology* 2016;**173**(1):127-36.

DeAndres-Galiana 2016 *{published data only}*

DeAndres-Galiana EJ, Fernandez-Martinez JL, Luaces O, Del Coz JJ, Huergo-Zapico L, Acebes-Huerta A, et al. Analysis of clinical prognostic variables for chronic lymphocytic leukemia decision-making problems. *Journal of Biomedical Informatics* 2016;**60**:342-51.

De Faria 2000 *{published data only}*

De Faria JR, De Oliveira JS, Delbone de Faria RM, Silva MR, Goihman S, Yamamoto M, et al. Prognosis related to staging systems for chronic lymphocytic leukemia. *Revista Paulista de Medicina [Sao Paulo Medical Journal]* 2000;**118**(4):83-8.

Degan 2004 *{published data only}*

Degan M, Bomben R, Bo MD, Zucchetto A, Nanni P, Rupolo M, et al. Analysis of IgV gene mutations in B cell chronic lymphocytic leukaemia according to antigen-driven selection identifies subgroups with different prognosis and usage of the canonical somatic hypermutation machinery. *British Journal of Haematology* 2004;**126**(1):29-42.

Delgado 2009 *{published data only}*

Delgado J, Pratt G, Phillips N, Briones J, Fegan C, Nomdedeu J, et al. Beta2-microglobulin is a better predictor of treatment-free survival in patients with chronic lymphocytic leukaemia if adjusted according to glomerular filtration rate. *British Journal of Haematology* 2009;**145**(6):801-5.

Delgado 2014 *{published data only}*

Delgado J, Salaverria I, Baumann T, Martinez-Trillos A, Lee E, Jimenez L, et al. Genomic complexity and IGHV mutational status are key predictors of outcome of chronic lymphocytic leukemia patients with TP53 disruption. *Haematologica* 2014;**99**(11):e231-4.

Del Guidice 2011 *{published data only}*

Del Giudice I, Mauro FR, De Propris MS, Santangelo S, Marinelli M, Peragine N, et al. White blood cell count at diagnosis and immunoglobulin variable region gene mutations are independent predictors of treatment-free survival in young patients with stage A chronic lymphocytic leukemia. *Haematologica* 2011;**96**(4):626-30.

Del Poeta 2010 *{published data only}*

Del Poeta G, Del Principe MI, Maurillo L, Rossi FM, Buccisano F, Ammatuna E, et al. Spontaneous apoptosis and proliferation detected by BCL-2 and CD71 proteins are important progression indicators within ZAP-70 negative chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2010;**51**(1):95-106.

Del Principe 2004 *{published data only}*

Del Principe MI, Del Poeta G, Venditti A, Buccisano F, Maurillo L, Marini R, et al. Clinical significance of soluble p53 protein in B-cell chronic lymphocytic leukemia. *Haematologica* 2004;**89**(12):1468-75.

Del Principe 2006 *{published data only}*

Del Principe MI, Del Poeta G, Buccisano F, Maurillo L, Venditti A, Zucchetto A, et al. Clinical significance of ZAP-70 protein expression in B-cell chronic lymphocytic leukemia. *Blood* 2006;**108**(3):853-61.

De Rossi 1989 *{published data only}*

De Rossi G. Prognosis in chronic lymphocytic leukemia (CLL): experience of the Italian cooperative group (GIMEMA). *Bone Marrow Transplantation* 1989;**4 Suppl 1**:162-4.

Deslandes 2007 *{published data only}*

Deslandes E, Chevret S. Assessing surrogacy from the joint modelling of multivariate longitudinal data and survival: application to clinical trial data on chronic lymphocytic leukaemia. *Statistics in Medicine* 2007;**26**(30):5411-21.

Dimier 2018 *{published data only}*

Dimier N, Delmar P, Ward C, Morariu-Zamfir R, Fingerle-Rowson G, Bahlo J, et al. A model for predicting effect of treatment on progression-free survival using MRD as a surrogate end point in CLL. *Blood* 2018;**131**(9):955-62.

Di Raimondo 2001 *{published data only}*

Di Raimondo F, Giustolisi R, Lerner S, Cacciola E, O'Brien S, Kantarjian H, et al. Retrospective study of the prognostic role of serum thymidine kinase level in CLL patients with active disease treated with fludarabine. *Annals of Oncology* 2001;**12**(5):621-5.

Dong 2011 *{published data only}*

Dong HJ, Zhou LT, Zhu DX, Wang DM, Fang C, Zhu HY, et al. The prognostic significance of TP53 mutations in Chinese patients with chronic lymphocytic leukemia is independent of del(17p13). *Annals of Hematology* 2011;**90**(6):709-17.

Dong 2014 *{published data only}*

Dong HJ, Fang C, Wang L, Fan L, Xu J, Wu JZ, et al. TP53 Pro72 allele potentially increases the poor prognostic significance of TP53 mutation in chronic lymphocytic leukemia. *Medical Oncology* 2014;**31**(4):908.

Durak 2009 *{published data only}*

Durak B, Akay OM, Aslan V, Ozdemir M, Sahin F, Artan S, et al. Prognostic impact of chromosome alterations detected by FISH in Turkish patients with B-cell chronic lymphocytic leukemia. *Cancer Genetics and Cytogenetics* 2009;**188**(2):65-9.

El-Kinawy 2012 *{published data only}*

El-Kinawy NS, Sharaf HM, El-Hamid MA. Prognostic significance of del 17p, ZAP-70 and CD38 as independent indicators for B-CLL: correlation to response to treatment and disease outcome. *Egyptian Journal of Medical Human Genetics* 2012;**13**(2):173-81.

Fang 2019 *{published data only}*

Fang H, Reichard KK, Rabe KG, Hanson CA, Call TG, Ding W, et al. IGH translocations in chronic lymphocytic leukemia: clinicopathologic features and clinical outcomes. *American Journal of Hematology* 2019;**94**(3):338-45.

Ferrara 1981 *{published data only}*

Ferrara F, Borrelli G, Fasanaro A, Mettivier V, Rametta V, Cimino R. Prognostic evaluation of chronic lymphocytic leukemia in relation to the Rai classification. Critical analysis of 145 personal cases. *Minerva Medica* 1981;**72**(12):741-4.

Ferreira 2014 *{published data only}*

Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, et al. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Research* 2014;**24**(2):212-26.

French Cooperative Group on CLL 1988 *{published data only}*

French Cooperative Group on Chronic Lymphocytic Leukemia. Comparison of the (A, B, C) staging and the Rai's staging from a large prospective series (935 patients). *Nouvelle Revue Française d'Hématologie* 1988;**30**:363-7.

Friedman 2009 *{published data only}*

Friedman DR, Weinberg JB, Barry WT, Goodman BK, Volkheimer AD, Bond KM, et al. A genomic approach to improve prognosis and predict therapeutic response in chronic lymphocytic leukemia. *Clinical Cancer Research* 2009;**15**(22):6947-55.

Gattei 2008 *{published data only}*

Gattei V, Bulian P, Del Principe MI, Zucchetto A, Maurillo L, Buccisano F, et al. Relevance of CD49d protein expression as overall survival and progressive disease prognosticator in chronic lymphocytic leukemia. *Blood* 2008;**111**(2):865-73.

Gdynia 2018 *{published data only}*

Gdynia G, Robak T, Kopitz J, Heller A, Grekova S, Duglova K, et al. Distinct activities of glycolytic enzymes identify chronic lymphocytic leukemia patients with a more aggressive course and resistance to chemo-immunotherapy. *EBioMedicine* 2018;**32**:125-33.

Geisler 1997 *{published data only}*

Geisler CH, Philip P, Christensen BE, Hou-Jensen K, Pedersen NT, Jensen OM, et al. In B-cell chronic lymphocytic leukaemia chromosome 17 abnormalities and not trisomy 12 are the single most important cytogenetic abnormalities for the prognosis: a cytogenetic and immunophenotypic study of 480 unselected newly diagnosed patients. *Leukemia Research* 1997;**21**(11-12):1011-23.

Gentile 2016 *{published data only}*

Gentile M, Cutrona G, Molica S, Ilariucci F, Mauro FR, Di Renzo N, et al. Prospective validation of predictive value of abdominal computed tomography scan on time to first treatment in Rai 0 chronic lymphocytic leukemia patients: results of the multicenter O-CLL1-GISL study. *European Journal of Haematology* 2016;**96**(1):36-45.

Gentile 2018 *{published data only}*

Gentile M, Shanafelt TD, Mauro FR, Reda G, Rossi D, Laurenti L, et al. Predictive value of the CLL-IPI in CLL patients receiving chemo-immunotherapy as first-line treatment. *European Journal of Haematology* 2018;**101**(5):703-6.

Giles 2003 *{published data only}*

Giles FJ, Bekele BN, O'Brien S, Cortes JE, Verstovsek S, Balerdi M, et al. A prognostic model for survival in chronic lymphocytic leukaemia based on p53 expression. *British Journal of Haematology* 2003;**121**(4):578-85.

Giudice 2018 *{published data only}*

Giudice ID, Rigolin GM, Raponi S, Cafforio L, Ilari C, Wang J, et al. Refined karyotype-based prognostic stratification of chronic lymphocytic leukemia with a low- and very-low-risk genetic profile. *Leukemia* 2018;**32**(2):543-6.

Gogia 2014 *{published data only}*

Gogia A, Raina V, Gupta R, Gajendra S, Kumar L, Sharma A, et al. Prognostic and predictive significance of smudge cell percentage on routine blood smear in chronic lymphocytic leukemia. *Clinical Lymphoma, Myeloma & Leukemia* 2014;**14**(6):514-7.

Gonzalez 2013 *{published data only}*

Gonzalez D, Else M, Wren D, Usai M, Buhl AM, Parker A, et al. CLLU1 expression has prognostic value in chronic lymphocytic leukemia after first-line therapy in younger patients and in those with mutated IGHV genes. *Haematologica* 2013;**98**(2):274-8.

Gonzalez-Gascon 2015 *{published data only}*

Gonzalez-Gascon YMI, Hernandez-Sanchez M, Rodriguez-Vicente AE, Sanzo C, Aventin A, Puiggros A, et al. A high proportion of cells carrying trisomy 12 is associated with a worse outcome in patients with chronic lymphocytic leukemia. *Hematological Oncology* 2015;**17**(10):84–92.

Gonzalez-Rodriguez 2010 *{published data only}*

Gonzalez-Rodriguez AP, Contesti J, Huergo-Zapico L, Lopez-Soto A, Fernandez-Guizan A, Acebes-Huerta A, et al. Prognostic significance of CD8 and CD4 T cells in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2010;**51**(10):1829-36.

Grabowski 2005 *{published data only}*

Grabowski P, Hultdin M, Karlsson K, Tobin G, Aleskog A, Thunberg U, et al. Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. *Blood* 2005;**105**(12):4807-12.

Grever 2006 *{published data only}*

Grever MR, Dewald GW, Neuberg DS, Reed JC, Kitada S, Flinn IW, et al. Select high risk genetic features predict earlier progression following chemotherapy in chronic lymphocytic leukemia: prospective randomized trial (Intergroup E2997) to evaluate justification for risk-adapted therapy [abstract]. *Journal of Clinical Oncology: ASCO Annual Meeting Proceedings* 2006;**24**:342.

Hallek 1999 *{published data only}*

Hallek M, Langenmayer I, Nerl C, Knauf W, Dietzfelbinger H, Adorf D, et al. Elevated serum thymidine kinase levels identify a subgroup at high risk of disease progression in early, nonsmoldering chronic lymphocytic leukemia. *Blood* 1999;**93**(5):1732-7.

Han 1984 *{published data only}*

Han T, Barcos M, Emrich L, Ozer H, Gajera R, Gomez GA, et al. Bone marrow infiltration patterns and their prognostic significance in chronic lymphocytic leukemia: correlations with clinical, immunologic, phenotypic, and cytogenetic data. *Journal of Clinical Oncology* 1984;**2**(6):562-70.

Herold 2011 *{published data only}*

Herold T, Jurinovic V, Metzeler KH, Boulesteix AL, Bergmann M, Seiler T, et al. An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. *Leukemia* 2011;**25**(10):1639-45.

Hock 2010 *{published data only}*

Hock BD, McKenzie JL, McArthur L, Tansley S, Taylor KG, Fernyhough LJ. CD38 as a prognostic marker in chronic lymphocytic leukaemia at a single New Zealand centre: patient survival in comparison to age- and sex-matched population data. *Internal Medicine Journal* 2010;**40**(12):842-9.

Houldsworth 2014 *{published data only}*

Houldsworth J, Guttapalli A, Thodima V, Yan XJ, Mendiratta G, Zielonka T, et al. Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2014;**55**(4):920-8.

Hus 2006 *{published data only}*

Hus I, Podhorecka M, Bojarska-Junak A, Rolinski J, Schmitt M, Sieklucka M, et al. The clinical significance of ZAP-70 and CD38 expression in B-cell chronic lymphocytic leukaemia. *Annals of Oncology* 2006;**17**(4):683-90.

Jaksic 1981 *{published data only}*

Jaksic B, Vitale B. Total tumour mass score (TTM): a new parameter in chronic lymphocyte leukaemia. *British Journal of Haematology* 1981;**49**(3):405-13.

Jaksic 1992 *{published data only}*

Jaksic B, Kusec R. More on early chronic lymphocytic leukemia (CLL): clinical staging systems and indications for treatment. *Hematologic Pathology* 1992;**6**(4):219-21.

Jaksic 2014 *{published data only}*

Jaksic O, Vitale B, Jaksic B. An old and simple solution for a new problem - more on clinical staging and evaluation of response in B-cell chronic lymphocytic leukaemia in the era of new therapies. *British Journal of Haematology* 2014;**165**(5):737-40.

Josefsson 2007 *{published data only}*

Josefsson P, Geisler CH, Leffers H, Petersen JH, Andersen MK, Jurlander J, et al. CLLU1 expression analysis adds prognostic information to risk prediction in chronic lymphocytic leukemia. *Blood* 2007;**109**(11):4973-9.

Juliusson 1986 *{published data only}*

Juliusson G. Immunologic and cytogenetic studies improve prognosis prediction in chronic B-lymphocytic leukemia. A multivariate analysis of 24 variables. *Cancer* 1986;**58**(3):688-93.

Juliusson 1990 *{published data only}*

Juliusson G, Oscier DG, Fitchett M, Ross FM, Stockdill G, Mackie MJ, et al. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *New England Journal of Medicine* 1990;**323**(11):720-4.

Kahraman 2014 *{published data only}*

Kahraman Cetin N, Kacar Doger F, Kurtgoz S, Ruscuklu D, Yavasoglu I. Prognostic significance of the IgVH mutation status and immunohistochemical analysis of ZAP70 and CD38 in bone marrow biopsies in chronic lymphocytic leukemia. *Turkiye Klinikleri Journal of Medical Sciences* 2014;**34**(3):334-44.

Kardum-Skelin 2008 *{published data only}*

Kardum-Skelin I, Planinc-Peraica A, Ostojic Kolonic S, Radic-Kristo D, Milas M, Vrhovac R, et al. Clinical and laboratory prognostic parameters for leukemic types of chronic lymphoproliferative diseases. *Acta Medica Croatica* 2008;**62**(4):351-64.

Kardum-Skelin 2009 *{published data only}*

Kardum-Skelin I, Jaksic O, Kolonic SO, Vrhovac R, Fabijanic I, Jelic-Puskaric B, et al. New parameters of diploid histogram of image DNA cytometry and newly characterized types of nucleolar organizer region structures in defining the proliferative-kinetic index in chronic leukemic lymphoproliferative disorders. *Analytical & Quantitative Cytology & Histology* 2009;**31**(5):313-23.

Karmiris 1994 *{published data only}*

Karmiris T, Rohatiner AZ, Love S, Carter M, Ganjoo RK, Amess J, et al. The management of chronic lymphocytic leukemia at a single centre over a 24-year period: prognostic factors for survival. *Hematological Oncology* 1994;**12**(1):29-39.

Kay 2018 *{published data only}*

Kay NE, LaPlant BR, Pettinger AM, Call TG, Leis JF, Ding W, et al. Cumulative experience and long term follow-up of pentostatin-based chemoimmunotherapy trials for patients with chronic lymphocytic leukemia. *Expert Review of Haematology* 2018;**11**(4):337-49.

Keating 2000 *{published data only}*

Keating MJ, Smith TL, Lerner S, O'Brien S, Robertson LE, Kantarjian H, et al. Prediction of prognosis following fludarabine used as secondary therapy for chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2000;**37**(1-2):71-85.

Khalifa 2002 *{published data only}*

Khalifa M, Chehata S, Laatiri MA, Grira C, Gharbi O, Kortas M, et al. Epidemiologic, clinical and therapeutic aspects of chronic lymphoid leukemia: apropos of 120 cases. *Tunisie Medicale* 2002;**80**(10):584-9.

Kim 2004 *{published data only}*

Kim SZ, Chow KU, Kukoc-Zivojnov N, Boehrer S, Brieger A, Steimle-Grauer SA, et al. Expression of ZAP-70 protein correlates with disease stage in chronic lymphocytic leukemia and is associated with, but not generally restricted to, non-mutated Ig VH status. *Leukemia and Lymphoma* 2004;**45**(10):2037-45.

Kimby 1988 *{published data only}*

Kimby E, Mellstedt H, Nilsson B, Bjorkholm M, Holm G, Lindemalm C, et al. Blood lymphocyte characteristics as predictors of prognosis in chronic lymphocytic leukemia of B-cell type. *Hematological Oncology* 1988;**6**(1):47-55.

Kleinstern 2018 *{published data only}*

Kleinstern G, Camp NJ, Goldin LR, Vachon CM, Vajdic CM, De Sanjose S, et al. Association of polygenic risk score with the risk of chronic lymphocytic leukemia and monoclonal B-cell lymphocytosis. *Blood* 2018;**131**:2541-51.

Knospe 1977 *{published data only}*

Knospe WH, Gregory SA, Trobaugh FE Jr, Stedronsky JA, Schrek R. Chronic lymphocytic leukemia: correlation of clinical course and therapeutic response with in vitro testing and morphology of lymphocytes. *American Journal of Hematology* 1977;**2**(1):73-101.

Koberda 1989 *{published data only}*

Koberda J, Czyz J, Hellmann A. Leukocyte doubling time as a prognostic factor in chronic lymphocytic leukemia. *Acta Haematologica Polonica* 1989;**20**(2):195-200.

Korycka-Wolowiec 2011 *{published data only}*

Korycka-Wolowiec A, Kotkowska A, Wawrzyniak E, Blonski JZ, Robak T. Do chromosomal aberrations detected in chronic lymphocytic leukemia patients by conventional cytogenetics with DSP30 have the clinical significance? Preliminary results. *Acta Haematologica Polonica* 2011;**42**(4):669-80.

Krober 2002 *{published data only}*

Krober A, Seiler T, Benner A, Bullinger L, Bruckle E, Lichter P, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood* 2002;**100**(4):1410-6.

Krober 2006 *{published data only}*

Krober A, Bloehdorn J, Hafner S, Buhler A, Seiler T, Kienle D, et al. Additional genetic high-risk features such as 11q deletion, 17p deletion, and V3-21 usage characterize discordance of ZAP-70 and VH mutation status in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2006;**24**(6):969-75.

Kryachok 2011 *{published data only}*

Kryachok I, Abramenko I, Bilous N, Chumak A, Martina Z, Filonenko I. IGHV gene rearrangements as outcome predictors for CLL patients: experience of Ukrainian group. *Medical Oncology* 2011;**29**(2):1093-101.

Kurec 1992 *{published data only}*

Kurec AS, Threatte GA, Gottlieb AJ, Smith JR, Anderson J, Davey FR. Immunophenotypic subclassification of chronic lymphocytic leukaemia (CLL). *British Journal of Haematology* 1992;**81**(1):45-51.

Lai 2002 *{published data only}*

Lai R, O'Brien S, Maushouri T, Rogers A, Kantarjian H, Keating M, et al. Prognostic value of plasma interleukin-6 levels in patients with chronic lymphocytic leukemia. *Cancer* 2002;**95**(5):1071-5.

Lech-Maranda 2012 *{published data only}*

Lech-Maranda E, Grzybowska-Izydorczyk O, Wyka K, Mlynarski W, Borowiec M, Antosik K, et al. Serum tumor necrosis factor-alpha and interleukin-10 levels as markers to predict outcome of patients with chronic lymphocytic leukemia in different risk groups defined by the IGHV mutation status. *Archivum Immunologiae et Therapiae Experimentalis* 2012;**60**(6):477-86.

Lech-Maranda 2013 *{published data only}*

Lech-Maranda E, Mlynarski W, Grzybowska-Izydorczyk O, Borowiec M, Pastorczak A, Cebula-Obrzut B, et al. Polymorphisms of TNF and IL-10 genes and clinical outcome of patients with chronic lymphocytic leukemia. *Genes, Chromosomes & Cancer* 2013;**52**(3):287-96.

Lecouvet 1997 *{published data only}*

Lecouvet FE, Vande Berg BC, Michaux L, Schmitz PJ, Malghem J, Jamart J, et al. Early chronic lymphocytic leukemia: prognostic value of quantitative bone marrow MR imaging findings and correlation with hematologic variables. *Radiology* 1997;**204**(3):813-8.

Li 2008 *{published data only}*

Li FJ, Ding S, Pan J, Shakhmatov MA, Kashentseva E, Wu J, et al. FCRL2 expression predicts IGHV mutation status and clinical progression in chronic lymphocytic leukemia. *Blood* 2008;**112**(1):179-87.

Li 2017a *{published data only}*

Li H, Xiong W, Liu H, Yi S, Yu Z, Liu W, et al. Serum LDH level may predict outcome of chronic lymphocytic leukemia patients with a 17p deletion: a retrospective analysis of prognostic factors in China. *Chinese Journal of Cancer Research* 2017;**29**(2):156-65.

Lin 2002 *{published data only}*

Lin K, Sherrington PD, Dennis M, Matrai Z, Cawley JC, Pettitt AR. Relationship between p53 dysfunction, CD38 expression, and IgV(H) mutation in chronic lymphocytic leukemia. *Blood* 2002;**100**(4):1404-9.

Lin 2014 *{published data only}*

Lin TT, Norris K, Heppel NH, Pratt G, Allan JM, Allsup DJ, et al. Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. *British Journal of Haematology* 2014;**167**(2):214-23.

Lozano-Santos 2014 *{published data only}*

Lozano-Santos C, Martinez-Velasquez J, Fernandez-Cuevas B, Polo N, Navarro B, Millan I, et al. Vascular endothelial growth factor A (VEGFA) gene polymorphisms have an impact on survival in a subgroup of indolent patients with chronic lymphocytic leukemia. *PLOS One* 2014;**9**(6):e101063.

Lucas 2015 *{published data only}*

Lucas DM, Ruppert AS, Lozanski G, Dewald GW, Lozanski A, Claus R, et al. Cytogenetic prioritization with inclusion of molecular markers predicts outcome in previously untreated patients with chronic lymphocytic leukemia treated with fludarabine or fludarabine plus cyclophosphamide: a long-

term follow-up study of the US intergroup phase III trial E2997. *Leukemia and Lymphoma* 2015;**56**(11):3031-7.

Maffei 2007 *{published data only}*

Maffei R, Marasca R, Martinelli S, Castelli I, Santachiara R, Morandi E, et al. Angiopoietin-2 expression in B-cell chronic lymphocytic leukemia: association with clinical outcome and immunoglobulin heavy-chain mutational status. *Leukemia* 2007;**21**(6):1312-5.

Maffei 2010 *{published data only}*

Maffei R, Martinelli S, Santachiara R, Rossi D, Guarnotta C, Sozzi E, et al. Angiopoietin-2 plasma dosage predicts time to first treatment and overall survival in chronic lymphocytic leukemia. *Blood* 2010;**116**(4):584-92.

Mandelli 1987 *{published data only}*

Mandelli F, De Rossi G, Mancini P, Alberti A, Cajozzo A, Grignani F, et al. Prognosis in chronic lymphocytic leukemia: a retrospective multicentric study from the GIMEMA group. *Journal of Clinical Oncology* 1987;**5**(3):398-406.

Mansouri 2013 *{published data only}*

Mansouri L, Grabowski P, Degerman S, Svenson U, Gunnarsson R, Cahill N, et al. Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. *American Journal of Hematology* 2013;**88**(8):647-51.

Marasca 2005 *{published data only}*

Marasca R, Maffei R, Morselli M, Zucchini P, Castelli I, Martinelli S, et al. Immunoglobulin mutational status detected through single-round amplification of partial V(H) region represents a good prognostic marker for clinical outcome in chronic lymphocytic leukemia. *Journal of Molecular Diagnostics* 2005;**7**(5):566-74.

Marasca 2013 *{published data only}*

Marasca R, Maffei R, Martinelli S, Fiorcari S, Bulgarelli J, Debbia G, et al. Clinical heterogeneity of de novo 11q deletion chronic lymphocytic leukaemia: prognostic relevance of extent of 11q deleted nuclei inside leukemic clone. *Hematological Oncology* 2013;**31**(2):348-55.

Martinelli 2008 *{published data only}*

Martinelli S, Maffei R, Castelli I, Santachiara R, Zucchini P, Fontana M, et al. Increased expression of angiopoietin-2 characterizes early B-cell chronic lymphocytic leukemia with poor prognosis. *Leukemia Research* 2008;**32**(4):593-7.

Masic 1998 *{published data only}*

Masic N, Gagro A, Rabatic S, Sabioncello A, Dasic G, Jaksic B, et al. Decision-tree approach to the immunophenotype-based prognosis of the B-cell chronic lymphocytic leukemia. *American Journal of Hematology* 1998;**59**(2):143-8.

Mateva 2001 *{published data only}*

Mateva N, Nenova I, Dimitrov I, Karnolski I, Ananoshtev N. A predictive model for survival of patients with chronic lymphocytic leukemia - results of a multivariate survival analysis. *Journal of the Balkan Union of Oncology* 2001;**6**(4):429-33.

Matthews 2006 *{published data only}*

Matthews C, Catherwood MA, Morris TC, Kettle PJ, Drake MB, Gilmore WS, et al. Serum TK levels in CLL identify Binet stage A patients within biologically defined prognostic subgroups most likely to undergo disease progression. *European Journal of Haematology* 2006;**77**(4):309-17.

Matthews 2007 *{published data only}*

Matthews C, Catherwood MA, Morris TC, Alexander HD. V(H)3-48 and V(H)3-53, as well as V(H)3-21, gene rearrangements define unique subgroups in CLL and are associated with biased lambda light chain restriction, homologous LCDR3 sequences and poor prognosis. *Leukemia Research* 2007;**31**(2):231-4.

Matutes 2013 *{published data only}*

Matutes E, Bosanquet AG, Wade R, Richards SM, Else M, Catovsky D. The use of individualized tumor response testing in treatment selection: second randomization results from the LRF CLL4 trial and the predictive value of the test at trial entry. *Leukemia* 2013;**27**(2):507-10.

Matutes 2017 *{published data only}*

Matutes E, Polliack A. Predicting time to first treatment in early stage CLL: scores, values and comparative prognostic models. *Leukemia & Lymphoma* 2017;**58**(7):1528-9.

Melo 1987 *{published data only}*

Melo JV, Catovsky D, Gregory WM, Galton DA. The relationship between chronic lymphocytic leukaemia and prolymphocytic leukaemia. IV. Analysis of survival and prognostic features. *British Journal of Haematology* 1987;**65**(1):23-9.

Miao 2018 *{published data only}*

Miao Y, Miao Y, Shi K, Sun Q, Zhao SS, Xia Y, et al. A higher percentage of cells with 13q deletion predicts worse outcome in Chinese patients with chronic lymphocytic leukemia carrying isolated 13q deletion. *Annals of Hematology* 2018;**97**:1663–9.

Molica 1984 *{published data only}*

Molica S, Puzzonia P, Iannaccaro P, Pane C, Alberti A. Prognosis of chronic lymphocytic leukemia: analysis of one hundred cases. *Tumori* 1984;**70**(5):399-402.

Molica 1986 *{published data only}*

Molica S, Alberti A. Prognostic value of "total tumor mass score" (TTM): a retrospective analysis of 130 patients with chronic lymphocytic leukemia. *Tumori* 1986;**72**(6):559-64.

Molica 1988 *{published data only}*

Molica S, Alberti A. Investigation of nuclear clefts as a prognostic parameter in chronic lymphocytic leukemia. *European Journal of Haematology* 1988;**41**(1):62-5.

Molica 1991 *{published data only}*

Molica S. Progression and survival studies in early chronic lymphocytic leukemia. *Blood* 1991;**78**(4):895-9.

Molica 1994 *{published data only}*

Molica S, Brugiatelli M, Callea V, Morabito F, Levato D, Nobile F, et al. Comparison of younger versus older B-cell chronic lymphocytic leukemia patients for clinical presentation and prognosis. A retrospective study of 53 cases. *European Journal of Haematology* 1994;**52**(4):216-21.

Molica 1998 *{published data only}*

Molica S, Levato D, Dattilo A, Mannella A. Clinico-prognostic relevance of quantitative immunophenotyping in B-cell chronic lymphocytic leukemia with emphasis on the expression of CD20 antigen and surface immunoglobulins. *European Journal of Haematology* 1998;**60**(1):47-52.

Molica 1999a *{published data only}*

Molica S, Levato D, Cascavilla N, Levato L, Musto P. Clinico-prognostic implications of simultaneous increased serum levels of soluble CD23 and beta2-microglobulin in B-cell chronic lymphocytic leukemia. *European Journal of Haematology* 1999;**62**(2):117-22.

Molica 1999b *{published data only}*

Molica S, Levato D, Dattilo A. Natural history of early chronic lymphocytic leukemia. A single institution study with emphasis on the impact of disease-progression on overall survival. *Haematologica* 1999;**84**(12):1094-9.

Molica 2008 *{published data only}*

Molica S, Vitelli G, Cutrona G, Todoerti K, Mirabelli R, Digiesi G, et al. Serum thrombopoietin compared with ZAP-70 and immunoglobulin heavy-chain gene mutation status as a predictor of time to first treatment in early chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2008;**49**(1):62-7.

Montserrat 1991 *{published data only}*

Montserrat E, Gomis F, Vallespi T, Rios A, Romero A, Soler J, et al. Presenting features and prognosis of chronic lymphocytic leukemia in younger adults. *Blood* 1991;**78**(6):1545-51.

Morabito 2001 *{published data only}*

Morabito F, Mangiola M, Oliva B, Stelitano C, Callea V, Deaglio S, et al. Peripheral blood CD38 expression predicts survival in B-cell chronic lymphocytic leukemia. *Leukemia Research* 2001;**25**(11):927-32.

Morabito 2015a *{published data only}*

Morabito F, Cutrona G, Gentile M, Fabris S, Matis S, Vigna E, et al. Is ZAP70 still a key prognostic factor in early stage chronic lymphocytic leukaemia? Results of the analysis from a prospective multicentre observational study. *British Journal of Haematology* 2015;**168**(3):455-9.

Morabito 2015b *{published data only}*

Morabito F, Cutrona G, Mosca L, D'Anca M, Matis S, Gentile M, et al. Surrogate molecular markers for IGHV mutational status in chronic lymphocytic leukemia for predicting time to first treatment. *Leukemia Research* 2015;**39**(8):840-5.

Morabito 2018c *{published data only}*

Morabito F, Shanafelt TD, Gentile M, Reda G, Mauro FR, Rossi D, et al. Immunoglobulin heavy chain variable region gene and prediction of time to first treatment in patients with chronic lymphocytic leukemia: mutational load or mutational status? Analysis of 1003 cases. *American Journal of Hematology* 2018;**93**(9):E216-9.

Moreno 2019 *{published data only}*

Moreno C, Delgado J, Byrd JC, Zvagelsky WL, Suzuki S, Hsu E, et al. Changes in clinical stage identify patients with CLL and different outcome within iwCLL partial response: RESONATE study. *British Journal of Haematology* 2019;**185**(1):148-50. [DOI: 10.1111/bjh.15397]

Morilla 2008 *{published data only}*

Morilla A, Gonzalez de Castro D, Del Giudice I, Osuji N, Else M, Morilla R, et al. Combinations of ZAP-70, CD38 and IGHV mutational status as predictors of time to first treatment in CLL. *Leukemia & Lymphoma* 2008;**49**(11):2108-15.

Nabhan 2017 *{published data only}*

Nabhan C, Mato A, Flowers CR, Grinblatt DL, Lamanna N, Weiss MA, et al. Characterizing and prognosticating chronic lymphocytic leukemia in the elderly: prospective evaluation on 455 patients treated in the United States. *BMC Cancer* 2017;**17**(1):198.

Nedeva 2018 *{published data only}*

Nedeva A, Naseva E, Kindekov I, Petkova N, Nikolov I, Raynov J. The prognostic significance of interphase cytogenetic abnormalities in chronic lymphocytic leukemia. *Archives of Hellenic Medicine* 2018;**35**(4):520-6.

Nenova 2000 *{published data only}*

Nenova I, Karnolski I, Mateva N, Ananoshtev N. A study of prognostic factors in patients with chronic lymphocytic leukemia. *Journal of Bulkan Union of Oncology* 2000;**5**(1):49-54.

Nipp 2014 *{published data only}*

Nipp RD, Volkheimer AD, Davis ED, Chen Y, Weinberg JB, Friedman DR. CD38 variation as a prognostic factor in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2014;**55**(1):191-4.

Nola 2004 *{published data only}*

Nola M, Pavletic SZ, Weisenburger DD, Smith LM, Bast MA, Vose JM, et al. Prognostic factors influencing survival in patients with B-cell small lymphocytic lymphoma. *American Journal of Hematology* 2004;**77**(1):31-5.

Nowakowski 2009 *{published data only}*

Nowakowski GS, Hoyer JD, Shanafelt TD, Zent CS, Call TG, Bone ND, et al. Percentage of smudge cells on routine blood smear predicts survival in chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2009;**27**(11):1844-9.

Nuckel 2006 *{published data only}*

Nuckel H, Huttmann A, Klein-Hitpass L, Schroers R, Fuhrer A, Sellmann L, et al. Lipoprotein lipase expression is a novel prognostic factor in B-cell chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2006;**47**(6):1053-61.

Nuckel 2009 *{published data only}*

Nuckel H, Collins CH, Frey UH, Sellmann L, Durig J, Siffert W, et al. FCRL2 mRNA expression is inversely associated with clinical progression in chronic lymphocytic leukemia. *European Journal of Haematology* 2009;**83**(6):541-9.

O'Brien 1993 *{published data only}*

O'Brien S, Kantarjian H, Beran M, Smith T, Koller C, Estey E, et al. Results of fludarabine and prednisone therapy in 264 patients with chronic lymphocytic leukemia with multivariate analysis-derived prognostic model for response to treatment. *Blood* 1993;**82**(6):1695-700.

Ocana 2007 *{published data only}*

Ocana E, Delgado-Perez L, Campos-Caro A, Munoz J, Paz A, Franco R, et al. The prognostic role of CXCR3 expression by chronic lymphocytic leukemia B cells. *Haematologica* 2007;**92**(3):349-56.

Oliveira 2011 *{published data only}*

Oliveira AC, De La Banda E, Domingo-Domenech E, Encuentra M, Mercadal S, Domingo A, et al. Prospective study of clinical and biological prognostic factors at diagnosis in patients with early stage B-cell chronic lymphocytic leukemia. *Leukemia and Lymphoma* 2011;**52**(3):429-35.

Orgueira 2019 *{published data only}*

Orgueira AM, Rodriguez BA, Vence NA, Lopez AB, Arias JA, Varela ND, et al. Time to treatment prediction in chronic lymphocytic leukemia based on new transcriptional patterns. *Frontiers in Oncology* 2019;**9**:79.

Oscier 1990 *{published data only}*

Oscier DG, Stevens J, Hamblin TJ, Pickering RM, Lambert R, Fitchett M. Correlation of chromosome abnormalities with laboratory features and clinical course in B-cell chronic lymphocytic leukaemia. *British Journal of Haematology* 1990;**76**(3):352-8.

Paolino 1984 *{published data only}*

Paolino W, Infelise V, Levis A, Marmont F, Vitolo U, Paolino F, et al. Adenosplenomegaly and prognosis in uncomplicated and complicated chronic lymphocytic leukemia. A study of 362 cases. *Cancer* 1984;**54**(2):339-46.

Plesingerova 2017 *{published data only}*

Plesingerova H, Librova Z, Plevova K, Libra A, Tichy B, Skuhrova Francova H, et al. COBLL1, LPL and ZAP70 expression defines prognostic subgroups of chronic lymphocytic leukemia patients with high accuracy and correlates with IGHV mutational status. *Leukemia & Lymphoma* 2017;**58**(1):70-9.

Prokocimer 1985 *{published data only}*

Prokocimer M, Modan M, Lusky A, Hershko C. Multivariate analysis of prognostic factors in chronic lymphocytic leukemia. *Israel Journal of Medical Sciences* 1985;**21**(6):490-8.

Qin 2017 *{published data only}*

Qin SC, Xia Y, Miao Y, Zhu HY, Wu JZ, Fan L, et al. MYD88 mutations predict unfavorable prognosis in chronic lymphocytic leukemia patients with mutated IGHV gene. *Blood Cancer Journal* 2017;**7**(12):651.

Queiros 2015 *{published data only}*

Queiros AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A, et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* 2015;**29**(3):598-605.

Rai 1975 *{published data only}*

Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. *Blood* 1975;**46**(2):219-34.

Rai 1990 *{published data only}*

Rai KR, Han T. Prognostic factors and clinical staging in chronic lymphocytic leukemia. *Hematology/Oncology Clinics of North America* 1990;**4**(2):447-56.

Raponi 2018 *{published data only}*

Raponi S, Del Giudice I, Marinelli M, Wang J, Cafforio L, Ilari C, et al. Genetic landscape of ultra-stable chronic lymphocytic leukemia patients. *Annals of Oncology* 2018;**29**(4):966-72.

Resegotti 1989 *{published data only}*

Resegotti L, Levis A, Decrescenzo A, Ficara F, Pini M. Significance and independence of the immunologic, histological, and clinical features as prognostic factors in chronic lymphatic leukemia. *Haematologica* 1989;**74**(5 Suppl):368-79.

Rissiek 2014 *{published data only}*

Rissiek A, Schulze C, Bacher U, Schieferdecker A, Thiele B, Jacholkowski A, et al. Multidimensional scaling analysis identifies pathological and prognostically relevant profiles of circulating T-cells in chronic lymphocytic leukemia. *International Journal of Cancer* 2014;**135**(10):2370-9.

Rodriguez 2007 *{published data only}*

Rodriguez A, Villuendas R, Yanez L, Gomez ME, Diaz R, Pollan M, et al. Molecular heterogeneity in chronic lymphocytic leukemia is dependent on BCR signaling: clinical correlation. *Leukemia* 2007;**21**(9):1984-91.

Ronchetti 2016 *{published data only}*

Ronchetti D, Manzoni M, Agnelli L, Vinci C, Fabris S, Cutrona G, et al. lncRNA profiling in early-stage chronic lymphocytic leukemia identifies transcriptional fingerprints with relevance in clinical outcome. *Blood Cancer Journal* 2016;**6**(9):e468.

Rossi 1986 *{published data only}*

Rossi PL, Alfano G, D'Onofrio G, Menichella G, Mango G. Depletion lymphocytapheresis in chronic lymphocytic leukemias: criteria for predicting which patients will respond to treatment. *International Journal of Artificial Organs* 1986;**9**(1):59-62.

Rossi 2010a *{published data only}*

Rossi FM, Del Principe MI, Rossi D, Irno Consalvo M, Luciano F, Zucchetto A, et al. Prognostic impact of ZAP-70 expression in chronic lymphocytic leukemia: mean fluorescence intensity T/B

ratio versus percentage of positive cells. *Journal of Translational Medicine* 2010;**8**:23.

Rossi 2010b *{published data only}*

Rossi S, Shimizu M, Barbarotto E, Nicoloso MS, Dimitri F, Sampath D, et al. MicroRNA fingerprinting of CLL patients with chromosome 17p deletion identify a miR-21 score that stratifies early survival. *Blood* 2010;**116**(6):945-52.

Rossi 2011 *{published data only}*

Rossi D, Spina V, Deambrogi C, Rasi S, Laurenti L, Stamatopoulos K, et al. The genetics of Richter syndrome reveals disease heterogeneity and predicts survival after transformation. *Blood* 2011;**117**(12):3391-401.

Rozman 1979 *{published data only}*

Rozman C, Montserrat Costa E, Morey M, Aranalde JM, Feliu E, Granena A, et al. Chronic lymphoid leukemia. Survival in relation to clinical stages. Statistical analysis of 95 cases. *Medicina Clinica* 1979;**72**(7):265-71.

Salomon-Nguyen 1995 *{published data only}*

Salomon-Nguyen F, Valensi F, Merle-Beral H, Flandrin G. A scoring system for the classification of CD5-B CLL versus CD5+ B CLL and B PLL. *Leukemia & Lymphoma* 1995;**16**(5-6):445-50.

Santoro 1979 *{published data only}*

Santoro A, Musumeci R, Rilke F, Franchi F, Valagussa P, Bajetta E, et al. Clinical classification and survival in chronic lymphocyte leukemia. *Tumori* 1979;**65**(1):39-49.

Sarmiento 2002 *{published data only}*

Sarmiento MA, Palacios MF, Scolnik MP, Ramirez FR, Stanganelli C, Cabrera J, et al. Evolution of chronic lymphocytic leukemia: predictive value of immunophenotype, soluble CD23 and morphology. *Medicina* 2002;**62**(4):305-12.

Savvopoulos 2016 *{published data only}*

Savvopoulos S, Misener R, Panoskaltsis N, Pistikopoulos EN, Mantalaris A. A personalized framework for dynamic modeling of disease trajectories in chronic lymphocytic leukemia. *IEEE Transactions on Biomedical Engineering* 2016;**63**(11):2396-404.

Scolozzi 1981 *{published data only}*

Scolozzi R, Boccafogli A, Romanini D, Ramazzina E, Careri C, Guidoboni CA. Prognosis of chronic lymphatic leukemia. Retrospective study of the survival period 47 patients. *Minerva Medica* 1981;**72**(4-5):205-11.

Shanafelt 2010 *{published data only}*

Shanafelt TD, Rabe KG, Kay NE, Zent CS, Jelinek DF, Reinalda MS, et al. Age at diagnosis and the utility of prognostic testing in patients with chronic lymphocytic leukemia. *Cancer* 2010;**116**(20):4777-87.

Shanafelt 2017 *{published data only}*

Shanafelt TD, Parikh SA, Noseworthy PA, Goede V, Chaffee KG, Bahlo J, et al. Atrial fibrillation in patients with chronic lymphocytic leukemia (CLL). *Leukemia & Lymphoma* 2017;**58**(7):1630-9.

Spacek 2009 *{published data only}*

Spacek M, Pekova S, Bezdickova L, Kozak T. Risk factors in chronic lymphocytic leukemia, validation of flow cytometric detection using RT-PCR and immunohistochemistry. *Transfuze a Hematologie Dnes* 2009;**15**(2):91-6.

Stamatopoulos 2009 *{published data only}*

Stamatopoulos B, Meuleman N, Haibe-Kains B, Saussoy P, Van Den Neste E, Michaux L, et al. MicroRNA-29c and microRNA-223 down-regulation has in vivo significance in chronic lymphocytic leukemia and improves disease risk stratification. *Blood* 2009;**113**(21):5237-45.

Stamatopoulos 2017 *{published data only}*

Stamatopoulos B, Timbs A, Bruce D, Smith T, Clifford R, Robbe P, et al. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia* 2017;**31**(4):837-45.

Strefford 2015 *{published data only}*

Strefford JC, Kadalayil L, Forster J, Rose-Zerilli MJ, Parker A, Lin TT, et al. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia* 2015;**10**(10):217.

Szymczyk 2018 *{published data only}*

Szymczyk A, Chocholska S, Macheta A, Szczepanek D, Hus M, Podhorecka M. Assessment of micro RNAs expression in leukemic cells as prognostic markers in chronic lymphocytic leukemia: micro RNAs can predict survival in a course of the disease. *Oncotarget* 2018;**9**(27):19136-46.

Tallarico 2018 *{published data only}*

Tallarico M, Foster JC, Seisler D, Lafky JM, Hurria A, Jatoi A, et al. Frequency and impact of grade three or four toxicities of novel agents on outcomes of older patients with chronic lymphocytic leukemia and non-Hodgkin lymphoma (alliance A151611). *Journal of Geriatric Oncology* 2018;**9**:321-8.

Tobin 2005a *{published data only}*

Tobin G. The immunoglobulin genes and chronic lymphocytic leukemia (CLL). *Upsala Journal of Medical Sciences* 2005;**110**(2):97-113.

Tobin 2005b *{published data only}*

Tobin G, Thunberg U, Laurell A, Karlsson K, Aleskog A, Willander K, et al. Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. *Haematologica* 2005;**90**(4):465-9.

Vallat 2013 *{published data only}*

Vallat L, Kemper CA, Jung N, Maumy-Bertrand M, Bertrand F, Meyer N, et al. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 2013;**110**(2):459-64.

Van Damme 2012 *{published data only}*

Van Damme M, Crompot E, Meuleman N, Mineur P, Bron D, Lagneaux L, et al. HDAC isoenzyme expression is deregulated

in chronic lymphocytic leukemia B-cells and has a complex prognostic significance. *Epigenetics* 2012;**7**(12):1403-12.

Velardi 1980 *{published data only}*

Velardi A, Spinozzi F, Siracusa A, Colozza M, Aversa F, Rambotti P, et al. Clinical staging of chronic lymphocytic leukaemia and its relationship to survival. *Haematologica* 1980;**65**(1):82-95.

Vojdeman 2017 *{published data only}*

Vojdeman FJ, Herman SE, Kirkby N, Wiestner A, Van T' Veer MB, Tjonnfjord GE, et al. Soluble CD52 is an indicator of disease activity in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2017;**58**(10):2356-62.

Vural 2014 *{published data only}*

Vural F, Karaca E, Soyer N, Gunduz C, Sahin F, Kosova B, et al. Comparison of CD38, ZAP70 and hTERT expression with known prognostic markers in patients with chronic lymphocytic leukemia during five-year follow- up period. *Uluslararasi Hematoloji-Onkoloji Dergisi (UHOD)* 2014;**24**(3):179-84.

Weinberg 2007 *{published data only}*

Weinberg JB, Volkheimer AD, Chen Y, Beasley BE, Jiang N, Lanasa MC, et al. Clinical and molecular predictors of disease severity and survival in chronic lymphocytic leukemia. *American Journal of Hematology* 2007;**82**(12):1063-70 [Erratum appears in American Journal of Hematology 2008;83(1):90].

Weiss 2011 *{published data only}*

Weiss L, Melchardt T, Egle A, Grabmer C, Greil R, Tinhofer I. Regulatory T cells predict the time to initial treatment in early stage chronic lymphocytic leukemia. *Cancer* 2011;**117**(10):2163-9.

Wierda 2003 *{published data only}*

Wierda WG, Johnson MM, Do KA, Manshouri T, Dey A, O'Brien S, et al. Plasma interleukin 8 level predicts for survival in chronic lymphocytic leukaemia. *British Journal of Haematology* 2003;**120**(3):452-6.

Winkler 2010 *{published data only}*

Winkler D, Schneider C, Zucknick M, Bogelein D, Schulze K, Zenz T, et al. Protein expression analysis of chronic lymphocytic leukemia defines the effect of genetic aberrations and uncovers a correlation of CDK4, P27 and P53 with hierarchical risk. *Haematologica* 2010;**95**(11):1880-8.

Wu 2010 *{published data only}*

Wu T, Li ZJ, Wang YF, Qiu LG. Applicability of different clinical staging systems in chronic lymphocytic leukemia for Chinese patients. *Journal of Leukemia and Lymphoma* 2010;**19**(3):136-9.

Zengin 1997 *{published data only}*

Zengin N, Kars A, Kansu E, Ozdemir O, Barista I, Gullu I, et al. Comparison of Rai and Binet classifications in chronic lymphocytic leukemia. *Hematology* 1997;**2**(2):125-9.

Zenz 2009 *{published data only}*

Zenz T, Benner A, Duhrsen U, Durig J, Dohner H, Siffert W, et al. BCL2-938C>A polymorphism and disease progression in chronic lymphocytic leukemia. *Leukemia & Lymphoma* 2009;**50**(11):1837-42.

Zucchetto 2006 *{published data only}*

Zucchetto A, Bomben R, Dal Bo M, Sonego P, Nanni P, Rupolo M, et al. A scoring system based on the expression of six surface molecules allows the identification of three prognostic risk groups in B-cell chronic lymphocytic leukemia. *Journal of Cellular Physiology* 2006;**207**(2):354-63.

References to studies awaiting assessment

Antic 2011 *{published data only}*

Antic D, Mihaljevic B, Cokic V, Fekete MD, Djurasevic TK, Pavlovic S, et al. Patients with early stage chronic lymphocytic leukemia: new risk stratification based on molecular profiling. *Leukemia & Lymphoma* 2011;**52**(7):1394-7.

Baumann 2014 *{published data only}*

Baumann T, Delgado J, Santacruz R, Martínez-Trillos A, Royo C, Navarro A, et al. Chronic lymphocytic leukemia in the elderly: clinico-biological features, outcomes, and proposal of a prognostic model. *Haematologica* 2014;**99**(10):1599-604.

Bulian 2011 *{published data only}*

Bulian P, Tarnani M, Rossi D, Forconi F, Del Poeta G, Bertoni F, et al. Multicentre validation of a prognostic index for overall survival in chronic lymphocytic leukaemia. *Hematological Oncology* 2011;**29**(2):91-9.

Bulian 2012 *{published data only}*

Bulian P, Rossi D, Forconi F, Del Poeta G, Bertoni F, Zucca E, et al. IGHV gene mutational status and 17p deletion are independent molecular predictors in a comprehensive clinical-biological prognostic model for overall survival prediction in chronic lymphocytic leukemia. *Journal of Translational Medicine* 2012;**10**:18.

Cavallini 2017 *{published data only}*

Cavallini C, Visco C, Putta S, Rossi D, Mimiola E, Purvis N, et al. Integration of B-cell receptor-induced ERK1/2 phosphorylation and mutations of SF3B1 gene refines prognosis in treatment-naïve chronic lymphocytic leukemia. *Haematologica* 2017;**102**(4):e144-7.

Del Guidice 2005 *{published data only}*

Del Giudice I, Morilla A, Osuji N, Matutes E, Morilla R, Burford A, et al. Zeta-chain associated protein 70 and CD38 combined predict the time to first treatment in patients with chronic lymphocytic leukemia. *Cancer* 2005;**104**(10):2124-32.

Friedrichs 2011 *{published data only}*

Friedrichs B, Siegel S, Reimer R, Barsoum A, Coggin J Jr, Kabelitz D, et al. High expression of the immature laminin receptor protein correlates with mutated IGVH status and predicts a favorable prognosis in chronic lymphocytic leukemia. *Leukemia Research* 2011;**35**(6):721-9.

Furundarena 1994 *{published data only}*

Furundarena Salsamendi JR, Martinez D, Navarro E, Aguinaco R, Vidal MJ, Lasa R, et al. Chronic lymphocytic leukemia: a

retrospective analysis of diagnostic characteristics and survival predictive factors in 150 patients. *Biologia y Clinica Hematologica* 1994;**16**(2):64-70.

Gentile 2009 *{published data only}*

Gentile M, Cutrona G, Neri A, Molica S, Ferrarini M, Morabito F. Predictive value of beta2-microglobulin (beta2-m) levels in chronic lymphocytic leukemia since Binet A stages. *Haematologica* 2009;**94**(6):887-8.

Haferlach 2010 *{published data only}*

Haferlach C, Dicker F, Weiss T, Schnittger S, Beck C, Grote-Metke A, et al. Toward a comprehensive prognostic scoring system in chronic lymphocytic leukemia based on a combination of genetic parameters. *Genes, Chromosomes & Cancer* 2010;**49**(9):851-9.

Jarque 1991 *{published data only}*

Jarque I, Sanz G, Gomis F, Martinez J, Martin G, De la Rubia J, et al. Chronic lymphatic leukemia. II. Analysis of prognostic factors and development of survival predicting models. Study of 187 patients. *Sangre* 1991;**36**(4):285-94.

Lee 1987 *{published data only}*

Lee JS, Dixon DO, Kantarjian HM, Keating MJ, Talpaz M. Prognosis of chronic lymphocytic leukemia: a multivariate regression analysis of 325 untreated patients. *Blood* 1987;**69**(3):929-36.

Leotard 2000 *{published data only}*

Leotard S, Chastang C, Travade P, Jaudon MC, Tournilhac O, Baudet S, et al. Prognostic relevance of a scoring system based on clinical and biological parameters in early chronic lymphocytic leukemia. *Hematology Journal* 2000;**1**(5):301-6.

Letestu 2010 *{published data only}*

Letestu R, Levy V, Eclache V, Baran-Marszak F, Vaur D, Naguib D, et al. Prognosis of Binet stage A chronic lymphocytic leukemia patients: the strength of routine parameters. *Blood* 2010;**116**(22):4588-90.

Li 2017b *{published data only}*

Li H, Yi SH, Xiong WJ, Liu HM, Lyu R, Wang TY, et al. Chronic lymphocytic leukemia prognostic index: a new integrated scoring system to predict the Time to First Treatment in Chinese patients with chronic lymphocytic leukemia. *Chinese Medical Journal* 2017;**130**(2):135-42.

Liang 2018 *{published data only}*

Liang JH, Gao R, Dai JC, Gale RP, Li W, Fan L, et al. The prognostic role of HBV infection in chronic lymphocytic leukemia. *Journal of Cancer Research and Clinical Oncology* 2018;**144**:1309–15.

Metze 2000 *{published data only}*

Metze K, Lobo AM, Lorand-Metze I. Nucleolus organizer regions (AgNORs) and total tumor mass are independent prognostic parameters for treatment-free period in chronic lymphocytic leukemia. *International Journal of Cancer* 2000;**89**(5):440-3.

Miao 2019 *{published data only}*

Miao Y, Zou YX, Gu DL, Zhu HC, Zhu HY, Wang L, et al. SF3B1 mutation predicts unfavorable treatment-free survival in Chinese chronic lymphocytic leukemia patients. *Annals of Translational Medicine* 2019;**7**(8):176.

Molica 1990 *{published data only}*

Molica S, Reverter JC, Alberti A, Montserrat E. Timing of diagnosis and lymphocyte accumulation patterns in chronic lymphocytic leukemia: analysis of their clinical significance. *European Journal of Haematology* 1990;**44**(5):277-81.

Molica 2010 *{published data only}*

Molica S, Digiesi G, Battaglia C, Cutrona G, Antenucci A, Molica M, et al. Baff serum level predicts time to first treatment in early chronic lymphocytic leukemia. *European Journal of Haematology* 2010;**85**(4):314-20.

Molica 2015 *{published data only}*

Molica S, Giannarelli D, Levato L, Gentile M, Mirabelli R, Morabito F. Do biologic parameters affect the time to first treatment of clinical monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia Rai stage 0? Results of a prospective analysis. *Clinical Lymphoma, Myeloma and Leukemia* 2015;**15**(3):e55-60.

Molica 2019 *{published data only}*

Molica S, Giannarelli D, Levato L, Mirabelli R, Levato D, Lentini M, et al. A simple score based on geriatric assessment predicts survival in elderly newly diagnosed chronic lymphocytic leukemia patients. *Leukemia & Lymphoma* 2019;**60**(3):845-7.

Morabito 2011 *{published data only}*

Morabito F, De Filippi R, Laurenti L, Zirlik K, Recchia AG, Gentile M, et al. The cumulative amount of serum-free light chain is a strong prognosticator in chronic lymphocytic leukemia. *Blood* 2011;**118**(24):6353-61.

Pepper 2012 *{published data only}*

Pepper C, Majid A, Lin TT, Hewamana S, Pratt G, Walewska R, et al. Defining the prognosis of early stage chronic lymphocytic leukaemia patients. *British Journal of Haematology* 2012;**156**(4):499-507.

Qin 2018 *{published data only}*

Qin S, Fan L, Liang J, Gale R, Miao Y, Wu Y, et al. Definition of disease-progression risk stratification in untreated chronic lymphocytic leukemia using combined clinical, molecular and virological variables. *Hematological Oncology* 2018;**36**(4):656-62.

Rozman 1982 *{published data only}*

Rozman C, Montserrat E, Feliu E, Granena A, Marin P, Nomdedeu B, et al. Prognosis of chronic lymphocytic leukemia: a multivariate survival analysis of 150 cases. *Blood* 1982;**59**(5):1001-5.

Rozman 1984 *{published data only}*

Rozman C, Montserrat E, Rodriguez-Fernandez JM, Ayats R, Vallespi T, Parody R, et al. Bone marrow histologic pattern - the best single prognostic parameter in chronic lymphocytic

leukemia: a multivariate survival analysis of 329 cases. *Blood* 1984;**64**(3):642-8.

Schweighofer 2011 *{published data only}*

Schweighofer CD, Coombes KR, Barron LL, Diao L, Newman RJ, Ferrajoli A, et al. A two-gene signature, SKI and SLAMF1, predicts time-to-treatment in previously untreated patients with chronic lymphocytic leukemia. *PLOS One* 2011;**6**(12):e28277.

Stamatopoulos 2010 *{published data only}*

Stamatopoulos B, Meuleman N, De Bruyn C, Pieters K, Anthoine G, Mineur P, et al. A molecular score by quantitative PCR as a new prognostic tool at diagnosis for chronic lymphocytic leukemia patients. *PLOS One* 2010;**5**(9):e12780.

Tsimberidou 2007 *{published data only}*

Tsimberidou AM, Wen S, O'Brien S, McLaughlin P, Wierda WG, Ferrajoli A, et al. Assessment of chronic lymphocytic leukemia and small lymphocytic lymphoma by absolute lymphocyte counts in 2,126 patients: 20 years of experience at the University of Texas M.D. Anderson Cancer Center. *Journal of Clinical Oncology* 2007;**25**(29):4648-56.

Vetro 2018 *{published data only}*

Vetro C, Haferlach T, Jeromin S, Stengel A, Zenger M, Nadarajah N, et al. Identification of prognostic parameters in CLL with no abnormalities detected by chromosome banding and FISH analyses. *British Journal of Haematology* 2018;**183**(1):47-59.

Visentin 2015 *{published data only}*

Visentin A, Facco M, Frezzato F, Castelli M, Trimarco V, Martini V, et al. Integrated CLL scoring system, a new and simple index to predict time to treatment and overall survival in patients with chronic lymphocytic leukemia. *Clinical Lymphoma, Myeloma & Leukemia* 2015;**15**(10):612-20.e5.

Wierda 2009 *{published data only}*

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do K-A, et al. Characteristics associated with important clinical end points in patients with chronic lymphocytic leukemia at initial treatment. Journal of Clinical Oncology 2009;**27**(10):1637-43.

References to ongoing studies

NCT00275054 *{published data only}*

NCT00275054. Rituximab, fludarabine, and cyclophosphamide or observation alone in treating patients with stage 0, stage I, or stage II chronic lymphocytic leukemia. clinicaltrials.gov/ct2/show/NCT00275054 (first received 11 January 2006).

NCT03436524 *{published data only}*

NCT03436524. A prognostic tool for early stage CLL. clinicaltrials.gov/ct2/show/NCT03436524 (first received 19 February 2018).

Additional references

Alba 2017

Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users guides to the medical literature. *JAMA* 2017;**318**(14):1377-84.

Bouwmeester 2012

Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLOS Medicine* 2012;**9**(5):1-12.

Cheson 1996

Cheson BD, Bennett JM, Grever M, Kay N, Keating MJ, O'Brien S, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood* 1996;**87**(12):4990-7.

Chiorazzi 2005

Chiorazzi N, Rai KR, Ferrarini M. Chronic lymphocytic leukemia. *New England Journal of Medicine* 2005;**352**(8):804-15. [PMID: 15728813]

Collins 2015

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Annals of Internal Medicine* 2015;**162**(1):735-6.

Collins 2016

Collins GS, Ma J, Gerry S, Ohuma E, Odondi L, Trivella M, et al. Risk prediction models in perioperative medicine: methodological considerations. *Current Anesthesiology Reports* 2016;**6**(3):267-75.

Debray 2012

Debray TP, Koffijberg H, Vergouwe Y, Moons KG, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Statistics in Medicine* 2012;**31**(23):2697-712.

Debray 2014

Debray TP, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KG. Meta-analysis and aggregation of multiple published prediction models. *Statistics in Medicine* 2014;**33**(14):2341-62.

Debray 2017

Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460.

Debray 2018a

Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Statistical Methods in Medical Research* 2018;**28**(9):2768-76.

Debray 2018b [Computer program]

R Foundation for Statistical Computing metamisc: Diagnostic and prognostic meta-Analysis. R package version 0.1.9. Debray T, de Jong V. Vienna, Austria: R Foundation for Statistical Computing, 2018. Available at www.R-project.org.

Dicker 2009

Dicker F, Herholz H, Schnittger S, Nakao A, Patten N, Wu L, et al. The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype. *Leukemia* 2009;**23**(1):117-24.

Döhner 2000

Döhner H, Stilgenbauer S, Benner A, Leupolt E, Krober A, Bullinger L, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. *New England Journal of Medicine* 2000;**343**(26):1910-6. [PMID: 11136261]

Eichhorst 2015

Eichhorst B, Robak T, Montserrat E, Ghia P, Hillmen P, Hallek M, et al, ESMO Guidelines Committee. Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2015;**26**(Suppl 5):v78-84.

El Rouby 1993

El Rouby S, Thomas A, Costin D, Rosenberg CR, Potmesil M, Silber R, et al. p53 gene mutation in B-cell chronic lymphocytic leukemia is associated with drug resistance and is independent of MDR1/MDR3 gene expression. *Blood* 1993;**82**(11):3452-9. [PMID: 8241511]

Ferrer Lores 2016

Ferrer Lores B, Navarro Cubells B, Serrano Alcala A, Castillo Martin I, Teruel Casasus A, Ballester Martinez S, et al. Prognostic impact of the new CLL-IPI Index in a single center CLL Spanish cohort. *Blood* 2016;**128**(22):2019.

Geersing 2012

Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLOS One* 2012;**7**:e32844.

Goede 2012

Goede V, Hallek M. Optimal pharmacotherapeutic management of chronic lymphocytic leukaemia: considerations in the elderly. *Drugs and Aging* 2012;**28**(3):163-76.

Goede 2016

Goede V, Bahlo J, Kutsch N, Fischer K, Fink AM, Fingerle-Rowson G, et al. Evaluation of the International Prognostic Index for Chronic Lymphocytic Leukemia (CLL-IPI) in elderly patients with comorbidities: analysis of the CLL11 study population. *Blood* 2016;**128**(22):4401.

Hallek 2008

Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al, International Workshop on Chronic Lymphocytic Leukemia. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. *Blood* 2008;**111**(12):5446-56.

Hallek 2017

Hallek M. Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology* 2017;**92**(9):946-65.

Herishanu 2013

Herishanu Y, Katz BZ, Lipsky A, Wiestner A. Biology of chronic lymphocytic leukemia in different microenvironments: clinical and therapeutic implications. *Hematology/Oncology Clinics of North America* 2013;**27**(2):173-206.

Heus 2018

Heus P, Damen JA, Pajouheshnia R, Scholten RJ, Reitsma JB, Collins GS, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Medicine* 2018;**16**(1):120.

Howlader 2019

Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al, National Cancer Institute. SEER Cancer Statistics Review, 1975-2016. www.seer.cancer.gov/archive/csr/1975_2016 (based on November 2018 SEER data submission, posted to the SEER web site, April 2019) (accessed May 2019).

Kay 2007

Kay NE, O'Brien SM, Pettitt AR, Stilgenbauer S. The role of prognostic factors in assessing 'high-risk' subgroups of patients with chronic lymphocytic leukemia. *Leukemia* 2007;**21**(9):1885-91. [PMID: 17568813]

Lefebvre 2019

Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, et al. Chapter 4: Searching for and selecting studies. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). Cochrane Handbook for Systematic Reviews of Interventions *V*ersion 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology* 2009;**62**(10):1006-12.

Molica 2018a

Molica S, Giannarelli D, Mirabelli R, Levato L, Shanafelt TD. Chronic lymphocytic leukemia international prognostic index (CLL-IPI) in patients receiving chemoimmuno or targeted therapy: a systematic review and meta-analysis. *Annals of Hematology* 2018;**97**(10):2005-8.

Molica 2018b

Molica S, Giannarelli D, Mirabelli R, Levato L, Kay NE, Shanafelt TD. Chronic lymphocytic leukemia international

prognostic index: a systematic review and meta-analysis. *Blood* 2018;**131**(3):365-8.

Moons 2009

Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;**338**:b375.

Moons 2014

Moons KG, De Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Medicine* 2014;**11**:e1001744.

Moons 2015

Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine* 2015;**162**(1):W1-W73.

Moons 2019

Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Annals of Internal Medicine* 2019;**170**(1):W1-W33.

Newcombe 2006

Newcombe RG. Confidence intervals for an effect size measure based on the Mann–Whitney statistic. Part 2: asymptotic methods and evaluation. *Statistics in Medicine* 2006;**25**(4):559-73.

Pajouheshnia 2019

Pajouheshnia R, Groenwold RH, Peelen LM, Reitsma JB, Moons KG. When and how to use data from randomised trials to develop or validate prognostic models. *BMJ* 2019;**365**:l2154.

Peat 2014

Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLOS Medicine* 2014;**11**:e1001671.

Pflug 2014

Pflug N, Bahlo J, Shanafelt TD, Eichhorst BF, Bergmann MA, Elter T, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. *Blood* 2014;**124**(1):49-62. [PMID: 24797299]

Riley 2013

Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis research strategy (PROGRESS) 2: Prognostic factor research. *PLOS Medicine* 2013;**10**(2):e1001380.

Riley 2019

Riley RD, Van der Windt DA, Croft P, Moons KGM. Prognosis Research in Healthcare. Concepts, Methods, and Impact. Oxford: Oxford University Press, 2019.

Shanafelt 2004

Shanafelt TD, Geyer SM, Kay NE. Prognosis at diagnosis: integrating molecular biologic insights into clinical practice for patients with CLL. *Blood* 2004;**103**(4):1202-10. [PMID: 14576043]

Steyerberg 2013

Steyerberg EW, Moons KG, Van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLOS Medicine* 2013;**10**(2):e1001381.

Steyerberg 2014

Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European Heart Journal* 2014;**35**(29):1925-31.

Stilgenbauer 2014

Stilgenbauer S, Schnaiter A, Paschka P, Zenz T, Rossi M, Döhner K, et al. Gene mutations and treatment outcome in chronic lymphocytic leukemia: results from the CLL8 trial. *Blood* 2014;**123**(21):3247-54.

Van Calster 2019

Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. *BMC Medicine* 2019;**17**(1):230.

Viechtbauer 2010

Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 2010;**36**(3):1-48.

Watson 2008

Watson L, Wyld P, Catovsky D. Disease burden of chronic lymphocytic leukaemia within the European Union. *European Journal of Haematology* 2008;**81**(4):253-8.

Wierda 2011

Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do KA, et al. Multivariable model for time to first treatment in patients with chronic lymphocytic leukemia. *Journal of Clinical Oncology* 2011;**29**(31):4088-95. [PMID: 21969505]

Wolff 2019

Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine* 2019;**170**(1):51-8.

Zaja 2013

Zaja F, Mian M, Volpetti S, Visco C, Sissa C, Nichele I, et al. Bendamustine in chronic lymphocytic leukemia: outcome according to different clinical and biological prognostic factors in the everyday clinical practice. *American Journal of Hematology* 2013;**88**(11):955-60.

Zenz 2011

Zenz T, Mertens D, Dohner H, Stilgenbauer S. Importance of genetics in chronic lymphocytic leukemia. *Blood Reviews* 2011;**25**(3):131-7. [PMID: 21435757]

References to other published versions of this review

Skoetz 2016

Skoetz N, Trivella M, Kreuzer KA, Collins G, Köhler N, Wolff M, et al. Prognostic models for chronic lymphocytic leukaemia: an exemplar systematic review and meta-analysis. *Cochrane Database of Systematic Reviews* 2016, Issue 1. Art. No: CD012022. [DOI: 10.1002/14651858.CD012022]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Baliakas D - Baliakas 2019 (multicentre)

Study characteristics

General information	Model and type of study • Baliakas 2019 (development cohort) • Development study Secondary citations • not applicable Language of publication • English Study design • multicentre, retrospective cohort Follow-up time • median 7.1 years (range 0.1 to 33.1 years)
Participants	Number of included persons in the cohort • 2366 persons* • *analysis included only 1900 persons with Binet A, divided into M-CLL (n = 1224) and U-CLL (n = 676) Setting • 10 European institutions, multicentre study Recruitment period • not reported Age (in years) • median 64.3 years (range: 22-92 years) Sex • 61% male Stages of disease • Binet A: 80%; Binet B: 12%; Binet C: 8% Treatment • not reported Inclusion criteria

Baliakas D - Baliakas 2019 (multicentre) *(Continued)*

	• diagnosed CLL, available immunogenetic data Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: 15 • multivariable analysis: 15 List of predictors in final model (including cut-points for dichotomised factors) • M-CLL: TP53 abnormality, trisomy12, subset #2 (IGHV3-21/IGLV3-21 BcR IG) • U-CLL: TP53 abnormality, del11q, gender Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • TTFT: evaluated from the diagnostic date until the date of initial treatment; untreated cases were censored at the time of last follow-up. Additional outcome(s) • NA
Missing data	Participants with any missing data? • yes • M-CLL: 306, U-CLL: 288 If yes, how was missing data handled? • excluded from multivariable analysis if data were missing for one of the factors included in the model
Analysis	Number of participants and number of events (specific time points where reported) • TTFT: 1900 persons, number of events not reported Predictor selection method • selection of prognostic factors based on univariable analysis Statistical method • decision tree based on binary recursive partitioning and subsequent application of an amalgamation algorithm Simplification of model? • yes Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes

Baliakas D - Baliakas 2019 (multicentre) *(Continued)*

- M-CLL: 4 risk groups; i) very high risk: Binet C with identical 5- and 10-year TP of 92%; ii) high risk: Binet B, 5y-TP and 10y-TP: 64% and 84%, respectively; iii) intermediate risk: Binet A with one of the following: TP53abn and/or +12 and/or subset #2 membership, 5y-TP and 10y-TP: 40% and 55%, respectively (of note, among 18 non-censored cases with no treatment indication for more than 10 years after diagnosis, 5 (30%) carried TP53abn); and iv) low risk: non-TP53abn/+12/subset #2 Binet A, 5y-TP and 10y-TP: 12% and 25%, respectively
- U-CLL: 5 risk groups; i) very high risk: Binet C with 5- and 10-year TP of 100%; ii) high risk: Binet B, 5y-TP and 10y-TP: 90% and 100%, respectively; iii) intermediate risk: Binet A with one of the following: TP53abn and/or SF3B1mut and/or del(11q), 5y-TP and 10y-TP: 78% and 98%, respectively; iv) low risk: nonTP53abn/SF3B1mut/del(11q) male Binet A, 5y-TP and 10yTP: 65% and 85%, respectively; and v) very low risk: non-TP53abn/SF3B1mut/del(11q) female Binet A, 5y-TP and 10y-TP: 45% and 65%, respectively

PROBAST: Applicability

Domain 1: Participant selection

- high - sample not representative of entire CLL population as only participants with disease stage Binet A were analysed

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding and conflict of interest

- "... the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Stockholm, the Lion's Cancer Research Foundation, Uppsala, the Marcus Borgström Foundation and Selander's Foundation, Uppsala; H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe" by the EU; H2020 "MEDGENET, Medical Genomics and Epigenomics Network" (No.692298) by the EU; H2020 "CLLassify, Innovative risk assessment for individualizing treatment in chronic lymphocytic leukemia" (No.702714) by the EU; Associazione Italiana per la Ricerca sul Cancro AIRC Investigator grants #20246, and Special Program Molecular Clinical Oncology AIRC 5 per mille #9965; Progetti di Rilevante Interesse Nazionale (PRIN) #2015ZMRFEA, MIUR, Rome,Italy; TRANSCAN-179 NOVEL JTC 2016; project CEITEC 2020 (LQ1601) by MEYS-CZ, project AZV-MH-CZ 15-30015A-4/2015; JCS was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund; Special Program Molecular Clinical Oncology 5 x 1000 No. 10007, Associazione Italiana per la Ricerca sul Cancro Foundation Milan, Italy; Progetto Ricerca Finaliz"
- "KS and PG received research support from Janssen Pharmaceuticals and Gilead Sciences. KS, PG and AH received research support from Novartis SA, Abbvie and Roche Hellas. The other authors declare no conflict of interests in relation to the present study."

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Missing data used as a reason for exclusion: only participants: (quote) "for whom immunogenetic data was available were included in this multicentre retrospective study".
Domain 2: Predictors	Yes	Detailed description of predictor assessment in the appendix
Domain 3: Outcome	Yes	Predefined outcome definition
Domain 4: Analysis	No	Participants with missing data were excluded from multivariable analyses:

Baliakas D - Baliakas 2019 (multicentre) *(Continued)*

	(quote) "we considered only those cases with available data for all the factors included in the model (n = 918 for M-CLL and n = 384 for U-CLL)"; univariable selection of predictors; assumptions of Cox proportional hazards model were checked.
Overall judgement	No

Baliakas V - Baliakas 2019 (MLL + Scan.)

Study characteristics	
General information	Model and type of study • Baliakas 2019 (MLL + Scan cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • not reported
Participants	Number of included persons in the cohort • 649 persons Setting • Munich Leukemia Laboratory (n = 508) • Scandinavian population-based study (n = 141) Recruitment period • not reported Age (in years) • median 63.6 years (range: 29-89 years) Sex • 62% male Stages of disease • Binet A: 100% Treatment • not reported Inclusion criteria

Baliakas V - Baliakas 2019 (MLL + Scan.) *(Continued)*

	• CLL, disease stage Binet A Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • M-CLL: TP53 abnormality, trisomy12, subset #2 • U-CLL: TP53 abnormality and/or SF3B1 mutation, del(11q), gender Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • TTFT was evaluated from the diagnostic date until the date of initial treatment, untreated cases were censored at the time of last follow-up. Additional outcome(s) • NA Outcome in model development • TTFT
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants and number of events (specific time points where reported) • time-to-first-treatment: 649 persons, number of events not reported Which model was used? • classification into risk groups Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • no
PROBAST: Applicability	Domain 1: Participant selection • unclear - neither eligibility criteria nor recruitment period reported Domain 2: Predictors • low

Baliakas V - Baliakas 2019 (MLL + Scan.) *(Continued)*

	Domain 3: Outcome • low
Notes	Funding and conflict of interest • "... the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Stockholm, the Lion's Cancer Research Foundation, Uppsala, the Marcus Borgström Foundation and Selander's Foundation, Uppsala; H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe" by the EU; H2020 "MEDGENET, Medical Genomics and Epigenomics Network" (No.692298) by the EU; H2020 "CLLassify, Innovative risk assessment for individualizing treatment in chronic lymphocytic leukemia" (No.702714) by the EU; Associazione Italiana per la Ricerca sul Cancro AIRC Investigator grants #20246, and Special Program Molecular Clinical Oncology AIRC 5 per mille #9965; Progetti di Rilevante Interesse Nazionale (PRIN) #2015ZMRFEA, MIUR, Rome,Italy; TRANSCAN-179 NOVEL JTC 2016; project CEITEC 2020 (LQ1601) by MEYS-CZ, project AZV-MH-CZ 15-30015A-4/2015; JCS was funded by Bloodwise (11052, 12036), the Kay Kendall Leukaemia Fund (873), Cancer Research UK (C34999/A18087, ECMC C24563/A15581), Wessex Medical Research and the Bournemouth Leukaemia Fund; Special Program Molecular Clinical Oncology 5 x 1000 No. 10007, Associazione Italiana per la Ricerca sul Cancro Foundation Milan, Italy; Progetto Ricerca Finaliz" • "KS and PG received research support from Janssen Pharmaceuticals and Gilead Sciences. KS, PG and AH received research support from Novartis SA, Abbvie and Roche Hellas. The other authors declare no conflict of interests in relation to the present study."

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Eligibility criteria not reported; not stated whether participants were excluded based on missing values; recruitment period not reported
Domain 2: Predictors	Unclear	Lack of information about predictor assessment, especially since the validation cohort consisted of two separate cohorts which were merged for analysis
Domain 3: Outcome	Unclear	Outcome definition not reported
Domain 4: Analysis	No	Performance measures (calibration or discrimination) not reported
Overall judgement	No	

Barcelona-Brno D - Delgado 2017 (Barcelona cohort)

Study characteristics	
General information	Model and type of study • Delgado 2017 (development cohort (Barcelona)) • Development study Secondary citations • not applicable Language of publication • English Study design

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) *(Continued)*

- retrospective cohort

Follow-up time

- median 99.6 months (range 1-456 months)

Participants	Number of included persons in the cohort • 524 persons Setting • Spain, single-centre study Recruitment period • not reported Age (in years) • median 62 years (range: 22-93 years) Sex • 60% male Stages of disease • Rai 0: 62%; Rai I-IV: 38% • Binet A: 83%; Binet B/C: 17% Treatment • 83 received FCR or similar • 82 received purine analogs without MoAbs • 86 received alkylating agents • 23 received others Inclusion criteria • Information at diagnosis included age, clinical stage (Rai and Binet), IgHV mutational status, beta2-microglobulin (B2M), and FISH-detected cytogenetic abnormalities Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: 5 • multivariable analysis: 5 List of predictors in final model (including cut-points for dichotomised factors) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • OS: calculated from the date of diagnosis to the date of death or last follow-up

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) *(Continued)*

	Additional outcome(s) • TTFT: calculated from the date of diagnosis to the date of first treatment or last follow-up, considering disease-unrelated deaths as competing events
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants and number of events (specific time points where reported) • overall survival: 524 persons, 212 events • TTFT: 524 persons, 292 events Predictor selection method • aim was the simplification of the CLL-IPI. Therefore, the 5 factors of this model were combined and the most discriminatory combination was chosen. The exact procedure was unclear. Statistical method • Identification of combination of biomarkers with significant discriminatory value (based on CLL-IPI model) Simplification of model? • no Performance measures reported? • Calibration: reported upon request • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)) • at 10 years • percentage of group without event at time point (low risk: 82%, intermediate risk: 52%, high risk: 27%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - no clear eligibility criteria and recruitment period reported Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding and conflict of interest • "Red Tematica de Investigacion Cooperativa en Cancer RT, Grant No.: 06/0020/002051 and RD12/0036/0023; Instituto de Salud Carlos III (ISCIII), Grant No.: FISS PI080304; ICGC-CLL Genome Project, Generalitat de Catalunya, Grant No.: 2009SGR1008; “Emili Letang” (T.B.)" • "The authors declare no conflict of interest." Other comments

Barcelona-Brno D - Delgado 2017 (Barcelona cohort) *(Continued)*

- It seemed that the model was built for OS, but also applied to TTFT, no clear description of this process provided.
- The authors have provided additional information (calibration and discrimination, number of events) upon request via email.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Retrospective retrieval of individuals with available information for their model - individuals without the necessary data may have been left out. Unclear recruitment period
Domain 2: Predictors	Yes	No explicit statement on predictor measurement, however, there are indications that in this single-centre study, lab procedures remained similar. This was a retrospective cohort, possibility to look forward
Domain 3: Outcome	Yes	No clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar within this single-centre study and therefore did not rate as high risk (no information on e.g. registry, frequency of follow-up)
Domain 4: Analysis	No	Dichotomisation of predictors. Although the choice of predictors was based on a previous model (CLL-IPI), no formal factor selection procedure or comparison of tested factor combinations were reported. One point assigned per factor, no formally established factor weights
Overall judgement	No	

Barcelona-Brno V - Delgado 2017 (Brno cohort)

Study characteristics	
General information	Model and type of study • Delgado 2017 (Brno cohort, Czech Republic) • Validation study Secondary citations • not applicable Language of publication • English Study design • not reported Follow-up time • median 59 months (range 3-330 months)
Participants	Number of included persons in the cohort • 417 persons

Barcelona-Brno V - Delgado 2017 (Brno cohort) *(Continued)*

	Setting • Czech Republic (Brno), single-centre study Recruitment period • not reported Age (in years) • median 62 years (range: 32-85 years) Sex • 66% male Stages of disease • Rai 0: 41%; Rai I-II: 43%; Rai III-IV: 16% • Binet A: 72%; Binet B: 12%; Binet C: 16% Treatment • 125 received FCR or similar • 41 received purine analogs without MoAbs • 53 received alkylating agents • 57 received others Inclusion criteria • not reported Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) • no differences between development and validation predictors Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: OS was calculated from the date of diagnosis to the date of death or last follow-up. Additional outcome(s) • TTFT: calculated from the date of diagnosis to the date of first treatment or last follow-up, considering disease-unrelated deaths as competing events Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled?

Barcelona-Brno V - Delgado 2017 (Brno cohort) *(Continued)*

- not applicable

Analysis	Number of participants and number of events (specific time points where reported) • OS: 417 persons, 158 events • TTFT: 417 persons, 276 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: provided upon request (calibration plot) • Discrimination: provided upon request (c-statistic) Creation of risk groups? • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV) • at 5 years • percentage of group without event at time point (low risk: 88.4% (83.0 to 94.1%), intermediate risk: 72.5% (65.4 to 80.4%), high risk: 53.9% (44.2 to 65.8%))
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period and eligibility criteria unclear Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding and conflict of interest • "Ministry of Health of the Czech Republic, Grant No.: AZV 15-31834A/ 2015 and AZV 15-30015A/2015; Ministry of Education, Youth and Sports of the Czech Republic project NPUII - CEITEC 2020, Grant No.: LQ1601." • "The authors declare no conflict of interest." Other comments • The authors have provided additional information (calibration and discrimination, number of events) upon request via email.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No recruitment period and study design reported No clear inclusion and exclusion criteria (we do not know if missing values have been part of the exclusion criteria)
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way

Barcelona-Brno V - Delgado 2017 (Brno cohort) *(Continued)*

Domain 3: Outcome	Yes	No clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar within this single-centre study and therefore did not rate as high risk (no information on e.g. registry, frequency of follow-up)
Domain 4: Analysis	Unclear	No information on missing values; information on calibration and discrimination was provided upon request

Barcelona-Brno V - Gentile 2017 (Italian & Mayo)

Study characteristics	
General information	Model and type of study • Gentile 2017 (Italian multicentre and Mayo clinic cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 82 months (range 3-330 months)
Participants	Number of included persons in the cohort • 1299 persons Setting • Italy, USA, multicentre study Recruitment period • 1985 - 2015 Age (in years) • median 63 years (range: 27-92 years) Sex • 61.3% male Stages of disease • Rai 0: 57.9%; Rai I: 28.6%; Rai II: 8.6%; Rai III: 1.5%; Rai IV: 3.5% • Binet A/B/C: not reported Treatment • 194 received chemotherapy

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) *(Continued)*

	• 316 received chemoimmunotherapy Inclusion criteria • newly-diagnosed CLL persons Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) • no differences between development and validation predictors Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. Additional outcome(s) • TTFT: measured as the day of CLL diagnosis until the start of therapy or last follow-up Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants and number of events (specific time points where reported) • OS: 1299 persons, 283 events • TTFT: 1299 persons, 510 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: reported (explained variation) • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV) • at 5 years

Barcelona-Brno V - Gentile 2017 (Italian & Mayo) *(Continued)*

	• percentage of group without event at time point (low risk: 92.2%, intermediate risk: 83.6%, high risk: 68.2%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period spanned a broad time period, wherein CLL definitions and treatment options have changed Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding and conflict of interest • Funding not reported • "The authors have no conflict of interest to disclose." Other comments • This publication referred to the same cohort as in Gentile 2016 • The authors provided no additional information on calibration upon request.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used. Many missing values due to retrospective design, rated in domain 4
Domain 2: Predictors	Yes	Quote: "*IgHV* mutation analysis and FISH were performed at the reference laboratory of each participating center. The *IgHV* mutation status was tested on tumour DNA collected at diagnosis, and was assessed according to the ERIC guidelines."
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment. We assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	> 70% of individuals were excluded based on missing values (quote: "because of the absence of the required laboratory data in the cohort initially considered. However, the individuals included were representative of the whole cohort and were similar for age, sex, and Rai stage distribution").
Overall judgement	No	

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Molica 2017 (O-CLL1-GISL) • Validation study

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) *(Continued)*

Secondary citations

- Delgado 2017

Language of publication

- English

Study design

- prospective cohort

Follow-up time

- median 42 months (range 1 - 82 months)

Participants	Number of included persons in the cohort • 337 persons Setting • Italy, multicentre study Recruitment period • 2007 - not reported Age (in years) • median 61 years (range: 33 - 70 years) Sex • 57.2% male Stages of disease • Rai 0: 77.8%; Rai I-II: 22.2% • Binet A: 100%; Binet B/C: not applicable Treatment • not reported Inclusion criteria • Newly-diagnosed individuals with CLL from several Italian Institutions who were seen within 12 months of diagnosis (confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation antigen 5 (CD5), 19 (CD19), 23 (CD23)); established diagnosis of B-CLL by NCI criteria, performed by local haematologist; age above 18 years and below 70 years; Binet stage A; NCI watch-and-wait policy Exclusion criteria • diagnosis more than 12 months ago, aged above 70 years, patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis, Binet B or C
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberrations (del17p, del11q) (binary) • TP53 and del17p were used as composite predictor in place of del17p only. Timing of predictor measurement

Barcelona-Brno V - Molica 2017 (O-CLL1-GISL) *(Continued)*

	• at study entry, within 12 months of diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of registration and the date of initiation of first CLL treatment. (Molica 2016) Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 persons, 91 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (AUC) Creation of risk groups? • no
PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model
Notes	Funding & conflict of interest • "This work was supported by funding from Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN). We thank AIL Cosenza-Fondazione Amelia Scorza' onlus, Cosenza, Italy, and Brigida Gulino for precious secretarial assistance." (Gentile 2016) • "The authors have no conflict of interest to disclose." Other comments • This publication validated the Barcelona-Brno score in the same population as the validation cohort in the development study. • The outcome did not correspond to the primary outcome of the model, and was therefore not included for analysis and PROBAST rating.

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.)

Study characteristics	
General information	Model and type of study • Munoz-Novas 2018 (Spanish cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 68 months (range 3 - 277 months)
Participants	Number of included persons in the cohort • 696 persons Setting • Spain, multicentre study Recruitment period • 1989 - 2013 Age (in years) • median 65.7 years (IQR: 55.2 - 73.5) Sex • 62.8% male Stages of disease • Rai 0: 58.9%; Rai I-IV: 41.1% • Binet A/B/C: not reported Treatment • 137 received treatment • unclear which treatment Inclusion criteria • all persons diagnosed with CLL Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) *(Continued)*

	• IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) • no differences between development and validation predictors Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: Overall survival (OS) was calculated from the time of diagnosis to death or last follow-up. Additional outcome(s) • time-to-first-treatment: Time-to-first therapy (TTFT) was defined from the date of diagnosis to first treatment or last follow-up. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (438 persons) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 258 persons, 47 events • time-to-first-treatment: 258 persons, 113 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV) • at 5 years • percentage of group without event at time point (low risk: 90.7%, intermediate risk: 81.4%, high risk: 66.2%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome

Barcelona-Brno V - Muñoz-Novas 2018 (Spanish coh.) *(Continued)*

	• low
Notes	Funding & conflict of interest • Funding not reported • "The authors declare that there are no conflicts of interest regarding the publication of this article." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Although individuals with incomplete data were excluded, there was a comparison and the sample seemed representative of the cohort.
Domain 2: Predictors	Yes	The type of biomarkers used seem to be standard, and although there was no explicit information we think that the methods were consistent across studies and models.
Domain 3: Outcome	Yes	Objective standard outcome - no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Consecutive sampling of individuals in hospital routine, therefore around 2/3 of individuals with missing data for this model validation were excluded; no calibration reported.
Overall judgement	No	

Barcelona-Brno V - Rani 2018 (Indian cohort)

Study characteristics	
General information	Model and type of study • Rani 2018 (Indian cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • unclear Follow-up time • median 40.5 months (range 1 - 215 months)
Participants	Number of included persons in the cohort

Barcelona-Brno V - Rani 2018 (Indian cohort) *(Continued)*

- 198 persons

Setting

- India, unclear if single or multicentre study

Recruitment period

- not reported

Age (in years)

- median 60 years (range: not reported)

Sex

- 77% male

Stages of disease

- Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%
- Binet A/B/C: not reported

Treatment

- 62 received chlorambucil-based therapy
- 56 received rituximab-based therapy
- 20 received other therapies

Inclusion criteria

- treatment-naive CLL persons according to Hallek 2008

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) • no differences between development and validation predictors Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study • overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up. Additional outcome(s) • time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy. Outcome in model development • overall survival
Missing data	Participants with any missing data? • no

Barcelona-Brno V - Rani 2018 (Indian cohort) *(Continued)*

	If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 198 persons, 86 events • time-to-first-treatment: 140 persons, 89 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV) • median survival (low risk: 138 years, intermediate risk: 63 years, high risk: 56 years)
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period and eligibility criteria unclear Domain 2: Predictors • unclear - no information on timing of predictor because the study design is unclear Domain 3: Outcome • low
Notes	Funding & conflict of interest • "... Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG to carry out this work." • "The authors declare that they have no conflict of interest." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Inclusion and exclusion criteria unclear; it seemed to us that only individuals with complete information on all predictors were included. Recruitment period unclear
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre)

Barcelona-Brno V - Rani 2018 (Indian cohort) *(Continued)*

Domain 3: Outcome	Unclear	Observation time to observe survival was short.
Domain 4: Analysis	No	Low number of events, no information on handling of missing values; calibration was not reported.
Overall judgement	No	

Barcelona-Brno V - Reda 2017 (Milan cohort)

Study characteristics	
General information	Model and type of study • Reda 2017 (Milan cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 144 months (range 0 - 360 months)
Participants	Number of included persons in the cohort • 698 persons Setting • Italy, single-centre study Recruitment period • 1983 - 2016 Age (in years) • median 65 years (range: 32 - 91 years) Sex • 57% male Stages of disease • Rai 0-I: 74%; Rai II-IV: not reported • Binet A/B/C: not reported Treatment • 190 received Chlorambucil (CHL) • 108 received fludarabinecyclophosphamide-rituximab (FCR)

Barcelona-Brno V - Reda 2017 (Milan cohort) *(Continued)*

	• 12 received ofatumumabbendamustine (O-Benda) • 44 received bendamustine-rituximab (BR) • 65 received alemtuzumab • 29 received Ibrutinib or Idelalisib Inclusion criteria • all patients diagnosed with CLL, reclassified by 2008 criteria Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV mutational status (continuous: categorical/binary transformation; binary, cut-off not given but usually at 98%); FISH-detected genetic abberations (del17p, del11q) (binary) • no differences between development and validation predictors Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: overall survival (OS) was calculated from CLL diagnosis to death. Additional outcome(s) • time-to-first-treatment: Time-To-First-Treatment (TTFT) was evaluated from the time of diagnosis to firstline therapy start. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (369 participants) If yes, how was missing data handled? • unclear
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 329 persons, number of events not reported • time-to-first-treatment: 329 persons, number of events not reported Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups?

Barcelona-Brno V - Reda 2017 (Milan cohort) *(Continued)*

- yes, 3 risk groups (no 11q or 17p and mutated IgHV); (all others); (11q or 17p and unmutated IgHV)
- at 15 years
- percentage of group without event at time point (low risk: 80%, intermediate risk: 50%, high risk: 30%)

PROBAST: Applicability	Domain 1: Participant selection • unclear - no clear information on eligibility criteria and study design Domain 2: Predictors • unclear - no information on the timing of the predictor Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding not reported • "Authors have no affiliations that they consider to be relevant and important with any organization that to any author's knowledge has a direct interest, particularly a financial interest, in the subject matter discussed." Other comments • The authors did not reply to our request via email for additional information.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria, unclear study design
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	No clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	No information on number of events; no performance measures reported; patients with missing data were left out from analysis.
Overall judgement	No	

CLL-IPI D - Bahlo 2016 (development cohort)

Study characteristics	
General information	Model and type of study • Bahlo 2016 (development cohort) • Development study Secondary citations • not applicable Language of publication

CLL-IPI D - Bahlo 2016 (development cohort) *(Continued)*

- English

Study design

- combined data (e.g. IPD or combination of cohorts); data from 8 phase-three RCTs

Follow-up time

- median 79.9 months, IQR: 79.9 - 101.4

Participants	Number of included persons in the cohort • 3472 persons Setting • multicentre study • France, Germany, Poland, UK, USA, and possibly more Recruitment period • 1997 - 2009 Age (in years) • median 61 years (range: 27 - 86 years) Sex • 70% male Stages of disease • Rai 0: 14%; Rai I/II: 50%; Rai III/IV: 36% • Binet A: 32%; Binet B: 41%; Binet C: 27% Treatment • 690 received fluradabine (F) • 1017 received F + cyclophosphamide (C) • 113 received C + cladribine • 475 received FC + rituximab • 81 received FC + alemtuzumab • 479 received chlorambucil Inclusion criteria • untreated CLL - possibly more exclusion criteria in the individual RCTs; FR: between the ages of 18 and 65 years; absence of del17p, life expectancy longer than 6 months; DE - CLL4: 18 and 65 years; life expectancy more than 6 months; Binet stage C and stage B if they had rapid disease progression or symptoms as evidenced by enlarged lymph nodes and organs or if they had severe B symptoms, stage A with severe B-symptoms; DE - CLL5: 65 and 80 years; life expectancy more than 6 months; Binet stage C and in stage B or A if they had rapid disease progression (lymphocyte doubling time 3 months) or symptoms from enlarged lymph nodes and organs, or if they had severe B-symptoms; PL: more than 18 years, progressive disease; UK: stages B, C, and A-progressive who needed treatment; USA: 18 years or older with a diagnosis of progressive CLL Exclusion criteria • FR: comorbidities (CIRS more than 5, ECOG more than 2; autoimmune cytopenia, HIV, HBV, HCV, active second malignancy, transformation to aggressive B-cell malignancy, abnormal renal function, bilirubin, transaminase more than 2 times the upper limit; DE - CLL 4: ECOG more than 3; organ dysfunction, concomitant or previous neoplasms, autoimmune haemolytic anaemia or thrombocytopenia; DE - CLL5: ECOG more than 3; organ dysfunction, concomitant or previous neoplasias, autoimmune

CLL-IPI D - Bahlo 2016 (development cohort) *(Continued)*

	haemolytic anaemia, thrombocytopenia; DE - CLL8: ECOG more than 2; CIRS more than 5; autoimmune cytopenia, active second disease; PL: poor performance status (WHO grade 4), autoimmune haemolytic anaemia (AIHA), autoimmune thrombocytopenia, active infections, abnormal liver or renal function, Richter's syndrome, or concomitant neoplasm; USA: second malignancy, serum creatinine more than 2 mg/dL; performance status more than 2; bilirubin more than 1 mg/dL; autoimmune haemolytic anaemia, thrombocytopenia; history of steroid treatment; pregnancy and lactating women; active infection
Predictors	Number of candidate predictors • univariable analysis: 27 • multivariable analysis: 25 List of predictors in final model (including cut-points for dichotomised factors) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L based on literature); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off: 65 years) Timing of predictor measurement • before the start of first-line treatment
Outcome(s)	Primary outcome in study • overall survival: overall survival was calculated from study entry to date of death. Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • yes (1673 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 1214 persons, 462 events for training set; 585 persons, 243 events for internal validation set Predictor selection method • univariable analysis, during multivariable modelling: a stepwise modelling based on categories of missing values: step 1: N = 1908, age, gender, Binet, LDH, haemoglobin [ex: WBC, HGB, LYM]; step 2: N = 1563 (585 events), age, gender, clinical stage (Rai OR Binet), LDH, haemoglobin, ECOG, hierarchical type (del17p, del11q); step 3: N = 1192 (452 events); age, clinical stage (Rai or Binet), IgHV, B2M, hierarchical type (del17p); step 4: more MV analyses (table S11 P value or CI) Statistical method • Cox proportional hazard model Simplification of model? • yes Performance measures reported? • Calibration: not reported

CLL-IPI D - Bahlo 2016 (development cohort) *(Continued)*

- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 93.2%, intermediate risk: 79.3%, high risk: 63.3%, very high risk: 23.3%)

PROBAST: Applicability

Domain 1: Participant selection

- high - most individuals had a previous treatment indication, in total only 17.8% of individuals were on watch-and-wait strategy at enrolment, therefore the population was rather selective.

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "José Carreras Leukaemia Foundation"
- KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacyclics and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacyclics Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacyclics, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report."

Other comments

- The authors provided additional information upon request (cohort information, measures of calibration).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used, missing values were excluded from analysis. This was rated in domain 4.
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way, according to RCT protocols.
Domain 3: Outcome	Yes	Objective standard outcome, follow-up in the context of each individual RCT Median observation time seemed on the lower limit (around 5 years), but reasonable for patients with treatment indication.
Domain 4: Analysis	No	It was unclear for which model the performance was assessed (score or formula). Predictors were categorised and dichotomised. Patients with missing values were dropped from the analysis - exclusion of approximately half of all patients. Univariable analysis was used for predictor selection.

CLL-IPI D - Bahlo 2016 (development cohort) *(Continued)*

	Upon request, the authors provided information on the calibration of their model.
Overall judgement	No

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014)

Study characteristics	
General information	Model and type of study • Bahlo 2016 (Mayo clinic 2001 - 2014) • Validation study Secondary citations • Gentile 2016; Molica 2016 Language of publication • English Study design • prospective cohort Follow-up time • median 63.2 months, IQR 30.2 to 91.8
Participants	Number of included persons in the cohort • 838 persons Setting • USA, single-centre study Recruitment period • 2001 - 2014 Age (in years) • ≤ 65 years: 533 persons • > 65 years: 305 persons Sex • 67.8% male Stages of disease • Rai 0: 57.4%; Rai I-II: 38.5%; Rai III-IV: 4.1% • Binet A/B/C: not reported Treatment • 163 received purine analogue chemotherapy • 5 received purine analogue monotherapy or combination without monoclonal antibody • 37 received alkylator-based chemoimmunotherapy

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) *(Continued)*

	• 11 received alkylator monotherapy or combination without monoclonal antibody • 97 received antibody only • 7 received RTK inhibitors • 6 received other treatments Inclusion criteria • complete baseline data Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); Age (continuous: categorical/binary transformation; cut-off 65 years) • missing data for TP53, therefore replaced by del(17p) Timing of predictor measurement • not reported, presumably at diagnosis
Outcome(s)	Primary outcome in study • overall survival: OS calculated from diagnosis to death. Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 838 persons, 144 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups?

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) *(Continued)*

- yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10))
- at 5 years
- percentage of group without event at time point (low risk: 96.6%, intermediate risk: 92%, high risk: 68.5%, very high risk: 21.2%)

PROBAST: Applicability

Domain 1: Participant selection

- unclear - insufficient information on CLL diagnostic criteria and inclusion criteria (i.e. availability of baseline data).

Domain 2: Predictors

- unclear - predictor TP53 not available and replaced by a proxy, del(17p)

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "José Carreras Leukaemia Foundation"
- KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacyclics and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacyclics Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacyclics, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No information on inclusion criteria, definition of CLL and if/how many patients were excluded due to missing baseline data
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk (no info on e.g. registry, frequency of follow-up etc.)
Domain 4: Analysis	Yes	Number of events sufficient Patients with missing values possibly dropped at enrolment, not clearly stated: 'There were no missing values for the MAYO cohort.'. This was rated in domain 1. Upon request, the authors provided information on the calibration of their model.

CLL-IPI V - Bahlo 2016 (Mayo clinic 2001-2014) *(Continued)*

Overall judgement	No

CLL-IPI V - Bahlo 2016 (SCAN cohort)

Study characteristics	
General information	Model and type of study • Bahlo 2016 (SCAN cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • other: cohort of CLL patients within a so-called case-control study, which was actually a register; uncertain whether prospective or retrospective study design Follow-up time • median 151 months, IQR 124.8 to 163.0
Participants	Number of included persons in the cohort • 416 persons Setting • Denmark and Sweden, multicentre study Recruitment period • 1999 - 2002 Age (in years) • ≤ 65 years: 242 persons • > 65 years: 174 persons Sex • 62.7% male Stages of disease • Rai: not reported • Binet A: 78.1%; Binet B: 16.6%; Binet C: 5.3% Treatment • not reported Inclusion criteria • complete baseline data

CLL-IPI V - Bahlo 2016 (SCAN cohort) *(Continued)*

	Exclusion criteria • The source population for the SCALE study was restricted to subjects with sufficient knowledge of the Danish or Swedish language to answer questions in a telephone interview and without a history of organ transplantation, human immunodeficiency virus infection, or other hematopoietic malignancy (Smedby 2005).
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • 4% of ß2-microglobulin imputed (multiple imputation) Timing of predictor measurement • not reported, presumably at diagnosis
Outcome(s)	Primary outcome in study • overall survival: OS was calculated from diagnosis to death. Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (17 participants) If yes, how was missing data handled? • multiple imputation
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 416 persons, 215 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • at 5 years • percentage of group without event at time point (low risk: 92.1%, intermediate risk: 75%, high risk: 64.3%, very high risk: 28.6%)

CLL-IPI V - Bahlo 2016 (SCAN cohort) *(Continued)*

PROBAST: Applicability	Domain 1: Participant selection • unclear - the SCALE study database encompassed the entire population in Denmark and Sweden in a specified time frame. For this model validation, only the CLL cohort was relevant. We did not know exactly how many participants were excluded due to missing information. No eligibility criteria were defined. We were unsure if the population can reflect the general CLL population. Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "José Carreras Leukaemia Foundation" • KB: BMBF grant. BFE: honorarium for advisory boards, honorarium and/or scientific grants from Janssen, Gilead, Mundipharma, GlaxoSmithKline (GSK) and Roche. CG: personal fees from Roche, Janssen, Gilead, Celgene, Novartis, and AbbVie, outside the submitted work. MG: National Cancer Institute during the conduct of the study; grants, personal fees, and non-financial support from Pharmacyclics and Acerta, outside the submitted work. TDS: grants from Genentech, Pharmacyclics Janssen, GSK, Celgene, Cephalon, Hospira, and Polyphenon E International, outside the submitted work. SS: grants, personal fees, and other from AbbVie, Amgen, Boehringer Ingelheim, Celgene, Genentech, Genzyme, Gilead, GSK, Janssen, Mundipharma, Novartis, Pharmacyclics, Hoffmann La-Roche, and Sanofi, during the conduct of the study. "JBa, NK, MAB, JBy, KGC, HD, ME, SL, DN, DO, TR, RR, and MH declare no competing interests." "The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report." Other comments • Upon request, the authors provided information on the calibration of their model.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	The SCALE study database encompassed the entire population in Denmark and Sweden in a specified time frame. For this model validation, only the CLL cohort was relevant. We did not know exactly how many participants were excluded. No eligibility criteria were defined.
Domain 2: Predictors	Yes	Patient data collection between 1999 and 2002; we assumed that within this time frame, predictor assessment remained relatively homogenous.
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. However, this was a national cohort study, therefore we assumed standardised outcome assessment.
Domain 4: Analysis	Yes	Sufficient number of events Missing data was handled appropriately: 'we analysed missing values of the SCAN cohort using Little's MCAR test and imputed these using linear regression.' Upon request, the authors provided information on the calibration of their model.

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort)

Study characteristics	
General information	Model and type of study • Da Cunha-Bang 2016 (Danish National CLL Registry) • Validation study Secondary citations • not applicable Language of publication • English Study design • prospective cohort Follow-up time • median 38.4 months, range not reported
Participants	Number of included persons in the cohort • 3023 persons Setting • Denmark, multicentre study Recruitment period • 2008 - 2015 Age (in years) • median 69 years (range: not reported) Sex • 60% male Stages of disease • Rai: not reported • Binet A: 80%; Binet B/C: 20% Treatment • 295 received treatment • unclear which treatment Inclusion criteria • all patients from national registry Exclusion criteria • missing data
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) *(Continued)*

	• TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • no differences between development and validation predictors reported Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • progression-free survival: TTE (time-to-event), event defined as treatment or death. Additional outcome(s) • overall survival: OS defined as time to death Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • progression-free survival: 1514 persons, 544 events • overall survival: 1514 persons, 249 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • no time point and numerical results reported, see figure 1 instead
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • unclear - predictors not described, therefore it was unclear if TP53 or a proxy was used Domain 3: Outcome • low

CLL-IPI V - Da Cunha-Bang 2016 (Danish cohort) *(Continued)*

Notes	Funding & conflict of interest • Funding not reported • "During the study, C.U.N. received grants from the Danish Cancer Society, consultancy fees (from Janssen, Roche, Abbvie, and Gilead), and grants (from Novartis and Roche) outside the submitted work and is the principal investigator for clinical trials sponsored by Roche. The remaining authors declare no competing financial interests." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Individuals with missing data previously excluded; comparison of included and excluded individuals showed similar baseline characteristics.
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way. Relatively recent and prospectively followed cohort
Domain 3: Outcome	Unclear	For mortality, the observation time was too short (our clinician recommended observation time to exceed 5 years, median observation time here: 3.2 years)
Domain 4: Analysis	Unclear	Patients with missing data previously excluded, although authors stated that they were comparable. Rated in domain 1. No information on model performance measures, however, a validation in our sense was not planned.

CLL-IPI V - Delgado 2017 (Barcelona cohort)

Study characteristics	
General information	Model and type of study • Delgado 2017 (Barcelona hospital cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 99.6 months, range not reported
Participants	Number of included persons in the cohort

CLL-IPI V - Delgado 2017 (Barcelona cohort) *(Continued)*

- 524 persons

Setting

- Spain (Barcelona), single-centre study

Recruitment period

- not reported

Age (in years)

- median 62 years (range: 22 - 93 years)

Sex

- 60% male

Stages of disease

- Rai 0: 62%; Rai I-IV: 38%
- Binet A: 83%; Binet or C: 17%

Treatment

- 83 received FCR or similar
- 82 received purine analogs w/o MoAbs
- 86 received alkylating agents
- 23 received others

Inclusion criteria

- Availability of information on age, clinical stage, IgHV status, ß2m, FISH

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • no changes Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: OS was calculated from the date of diagnosis to the date of death or last follow-up. Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data?

CLL-IPI V - Delgado 2017 (Barcelona cohort) *(Continued)*

	• no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 524 persons, number of events not reported Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10), using Rai or Binet) • at 5 years • percentage of group without event at time point (low risk: 95.7%, intermediate risk: 87.1%, high risk: 63.5%, very high risk: 51.9%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - no clear eligibility criteria and recruitment period reported Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Red Tematica de Investigacion Cooperativa en Cancer RT, Grant No.: 06/0020/002051 and RD12/0036/0023; Instituto de Salud Carlos III (ISCIII), Grant No.: FISS PI080304; ICGC-CLL Genome Project, Generalitat de Catalunya, Grant No.: 2009SGR1008; “Emili Letang” (T.B.)." • "The authors declare no conflict of interest." Other comments • Upon request, the authors provided additional information on the discrimination and calibration of the CLL-IPI model in their cohort.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Retrospective retrieval of persons with available information for their model - individuals without the necessary data may have been left out. Unclear recruitment period

CLL-IPI V - Delgado 2017 (Barcelona cohort) *(Continued)*

Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assumed standardised outcome assessment (single-centre study).
Domain 4: Analysis	Yes	Sufficient number of events. No missing values, however, possibly an exclusion reason. This was rated in domain 1. Upon request, the authors provided information on the calibration of the CLL-IPI in their cohort.
Overall judgement	No	

CLL-IPI V - Gentile 2016 (Italian cohort)

Study characteristics	
General information	Model and type of study • Gentile 2016 (Italian multicentre cohort) • Validation study Secondary citations • Gentile 2017 Language of publication • English Study design • retrospective cohort Follow-up time • median 69.6 months, range not reported
Participants	Number of included persons in the cohort • 858 persons Setting • Italy, multicentre study Recruitment period • 1985 - 2015 Age (in years) • median 65.5 years (range: not reported) Sex • 56.2% male

CLL-IPI V - Gentile 2016 (Italian cohort) *(Continued)*

Stages of disease

- Rai 0: 58.4%; Rai I: 19.8%; Rai II: 15.2%; Rai III: 1.5%; Rai IV: 5.1%
- Binet A: 79.7%; Binet B: 18.8%; Binet C: 1.5%

Treatment

- 130 received chemotherapy
- 174 received chemoimmunotherapy

Inclusion criteria

- All patients with CLL in one of the five centres that were evaluable for CLL-IPI (i.e. fully available data on age, staging, IgHV, ß2m and del17p/TP53)

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • TP53 was replaced by the proxy del(17p). Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. Additional outcome(s) • time-to-first-treatment: For TTFT from the day of CLL diagnosis until the start of therapy or last follow-up Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 858 persons, 174 events • time-to-first-treatment: 858 persons, 304 events Which model was used? • simplified model Was the model updated?

CLL-IPI V - Gentile 2016 (Italian cohort) *(Continued)*

	• no Performance measures reported? • Calibration: not reported, survival per risk group and explained variation on outcome available • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • at 5 years • percentage of group without event at time point (low risk: 94.5%, intermediate risk: 86.4%, high risk: 74.4%, very high risk: 42.9%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - data was retrospectively collected for a large time period (1985 - 2015). During this time, the definition of CLL and treatment options have changed. Therefore, the cohort may be very heterogeneous and discrimination for OS may not reflect performance of the model for individuals diagnosed nowadays. Domain 2: Predictors • unclear - predictor TP53 not available and replaced by a proxy, del(17p) Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding source not reported • "S.A.P. received funding from Pharmacyclics. All other authors have no conflict of interest to disclose." Other comments • identical centres as presented in Gentile 2017

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: consecutive sampling, but exclusion of individuals without baseline characteristics and FISH analysis from analysis (in total, only 22.5% of individuals were evaluable). This is rated in domain 4.
Domain 2: Predictors	Unclear	Due to the broad time range and multiple participating centres, we doubt that predictor assessments were homogeneous.
Domain 3: Outcome	Yes	Objective standard outcome: '... death, which were abstracted from clinical records at the time of inclusion and updated on an ongoing basis.'
Domain 4: Analysis	No	Consecutive sampling, but exclusion of individuals without baseline characteristics and FISH analysis from analysis (in total, only 22.5% of patients were evaluable). Missing values emerged due to the long time span. Some predictors were not assessed by default at diagnosis until recently. Comparison of baseline sample and included sample similar. Calibration not reported, however, survival per risk group available
Overall judgement	No	

CLL-IPI V - Molica 2016 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Molica 2016 (O-CLL1-GISL) • Validation study Secondary citations • Molica 2017; Molica 2017; Molica 2016 Language of publication • English Study design • prospective cohort Follow-up time • median 42 months, 2038 person-years
Participants	Number of included persons in the cohort • 337 persons Setting • Italy, multicentre study Recruitment period • 2007 - not reported Age (in years) • median 61 years (range: 33 - 70 years) Sex • 57.2% male Stages of disease • Rai 0: 77.8%; Rai ≥I: 22.2% • Binet A: 100%; Binet B/C: 0% Treatment • 91 received therapy • unclear which therapy Inclusion criteria • Newly-diagnosed CLL patients from several Italian Institutions who were seen within 12 months of diagnosis were prospectively enrolled into the OCLL1- GISL protocol. Exclusion criteria • not reported

CLL-IPI V - Molica 2016 (O-CLL1-GISL) *(Continued)*

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • TP53 and del17p as composite Timing of predictor measurement • at study entry, within 12 months of diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of registration and the date of initiation of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI sponsored Working Group guidelines was satisfied. Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 persons, 91 events Which model was used? • simplified model Was the model updated? • no • new cut-off score for risk groups tested Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • at 5 years

CLL-IPI V - Molica 2016 (O-CLL1-GISL) *(Continued)*

	• percentage of group without event at time point (low risk: 76%, intermediate risk: 45%, high risk: 41%, very high risk: not available)
PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model.
Notes	Funding & conflict of interest • "Contract grant sponsor: NIH; Contract grant no.: 1R01CA197120-01."; "Contract grant sponsors: AIRC (the Italian Association for Cancer Research), a non-profit organization to FM. “Special Program Molecular Clinical Oncology - 5 per mille” n. 9980, 2010/15 and AIRC “Innovative immunotherapeutic treatments of human cancer” n.16695, 2015/18." • "Conflict of interest: Nothing to report" Other comments • The same model was validated on the same cohort in other publications.

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort)

Study characteristics	
General information	Model and type of study • Munoz-Novas, 2018 (Spanish cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 68 months, range not reported
Participants	Number of included persons in the cohort • 696 persons Setting • Spain, multicentre study Recruitment period • 1989 - 2013 Age (in years) • median 65.7 years (range: IQR 55.2 - 73.5) Sex • 62.8% male

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) *(Continued)*

	Stages of disease • Rai 0: 58.9%; Rai I-IV: 41.1% • Binet A: not reported Treatment • 137 received treatment • unclear which treatment Inclusion criteria • all patients diagnosed with CLL Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); Age (continuous: categorical/binary transformation; cut-off 65 years) • TP53 not assessed, only del(17p) available Timing of predictor measurement • not reported, presumably at diagnosis
Outcome(s)	Primary outcome in study • overall survival: Overall survival (OS) was calculated fromthe time of diagnosis to death or last follow-up. Additional outcome(s) • time-to-first-treatment: Time to first therapy (TTFT) from the date of diagnosis to first treatment or last follow-up. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (438 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 258 persons, 47 events • time-to-first-treatment: 258 persons, 113 events Which model was used? • simplified model Was the model updated?

CLL-IPI V - Muñoz-Novas 2018 (Spanish cohort) *(Continued)*

	• no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) and ROC curve (S2) Creation of risk groups? • yes, 4 risk groups, (cut points at (0-1);(2-3);(4-6); (7-10)) • at 5 years • percentage of group without event at time point (low risk: 93.6%, intermediate risk: 87.6%, high risk: 67.8%, very high risk: 28.6%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • unclear - Predictor TP53 not available and replaced by a proxy, del(17p) Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding not reported • "The authors declare that there are no conflicts of interest regarding the publication of this article." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data source used: 'A total of 696 unselected CLL patients newly diagnosed and previously untreated from different institutions of the central region of Spain were included in this study.'
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assume standardised outcome assessment.
Domain 4: Analysis	No	Insufficient number of events Participants with missing values not included in model - consecutive sampling of routine patients, therefore around 2/3 of patients with missing data Calibration not reported
Overall judgement	No	

CLL-IPI V - Rani 2018 (Indian cohort)

Study characteristics	
General information	Model and type of study • Rani 2018 (Indian cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • unclear Follow-up time • median 40.5 months (range: 1 - 215 months)
Participants	Number of included persons in the cohort • 198 persons Setting • India, unclear if single or multicentre study Recruitment period • not reported Age (in years) • median 60 years (range: not reported) Sex • 77% male Stages of disease • Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.8%; Rai III: 14.1%; Rai IV: 15.2% • Binet A: not reported Treatment • 62 received chlorambucil-based therapy • 56 received rituximab-based therapy • 20 received other therapies Inclusion criteria • treatment-naive CLL persons according to Hallek 2008 Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

CLL-IPI V - Rani 2018 (Indian cohort) *(Continued)*

	• TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • only TP53 deletions were used to document TP53 aberrations because of unavailability of TP53 mutation status for all the patients. Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study • overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up. Additional outcome(s) • time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy. Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 198 persons, 86 events • time-to-first-treatment: 140 persons, 89 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups (cut-points at (0-1);(2-3);(4-6); (7-10)) • median survival (low risk: not reached, intermediate risk: 190 years, high risk: 62 years, very high risk: 28 years)
PROBAST: Applicability	Domain 1: Participant selection • unclear - eligibility criteria were not reported, possibly patients with missing values excluded. We are unsure if the sample is representative for CLL patients. Domain 2: Predictors

CLL-IPI V - Rani 2018 (Indian cohort) *(Continued)*

- unclear - TP53 was used as deletion only instead of deletion or mutation.

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work."
- "The authors declare that they have no conflict of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Inclusion and exclusion criteria unclear, it seemed like only patients with complete datasets were included. The recruitment period unclear
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study).
Domain 3: Outcome	Unclear	Observation time was short.
Domain 4: Analysis	No	Low number of events; no information on handling of missing values; calibration was not reported.
Overall judgement	No	

CLL-IPI V - Reda 2017 (Milano cohort)

Study characteristics

General information

Model and type of study

- Reda 2017 (Milano cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- English

Study design

- retrospective cohort

Follow-up time

CLL-IPI V - Reda 2017 (Milano cohort) *(Continued)*

- median 144 months, range not reported

Participants	Number of included persons in the cohort • 698 persons Setting • Italy, single-centre study Recruitment period • 1983 - 2016 Age (in years) • median 65 years (range: 32 - 91 years) Sex • 57% male Stages of disease • Rai 0-I: 74%; Rai II-IV: not reported • Binet A: 82%; Binet B/C: not reported Treatment • 190 received Chlorambucil (CHL) • 108 received fludarabinecyclophosphamide-rituximab (FCR) • 12 received ofatumumabbendamustine (O-Benda) • 44 received bendamustine-rituximab (BR) • 65 received alemtuzumab • 29 received Ibrutinib or Idelalisib Inclusion criteria • all patients diagnosed with CLL reclassified by 2008 criteria Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • It was not clear if TP53 was used, or its proxy del(17p) instead. Timing of predictor measurement • not reported, probably at diagnosis (medical records)
Outcome(s)	Primary outcome in study • overall survival: OS was defined as time from CLL diagnosis to death. Additional outcome(s) • time-to-first-treatment: TTFT was evaluated from the time of diagnosis to first-line therapy start.

CLL-IPI V - Reda 2017 (Milano cohort) *(Continued)*

	Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (369 participants) If yes, how was missing data handled? • unclear
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 329 persons, number of events not reported • time-to-first-treatment: 329 persons, number of events not reported Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • no time point and numerical results reported, see figure 1
PROBAST: Applicability	Domain 1: Participant selection • unclear - long observation period (1983 - 2016), no clear eligibility criteria Domain 2: Predictors • unclear - not much information on predictors reported, mentioning of FISH data but unclear if TP53 or del(17p) was used as predictor Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding not reported • "Authors have no affiliations that they consider to be relevant and important with any organization that to any author's knowledge has a direct interest, particularly a financial interest, in the subject matter discussed." Other comments • Hematology Unit, IRCCS Ca' Granda Ospedale Maggiore Policlinico of Milan, Milano, Italy

Item	Authors' judgement	Support for judgement

CLL-IPI V - Reda 2017 (Milano cohort) *(Continued)*

Domain 1: Participant selection	Unclear	No eligibility criteria reported
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment in this publication. We assumed standardised outcome assessment
Domain 4: Analysis	No	Unclear how many events and how many patients etc. were included for validation Patients with missing values were excluded from analysis. No performance measures and survival per group reported (figure 1 not sharp)
Overall judgement	No	

CLL-IPI V - Rigolin 2017 (Ferrera cohort)

Study characteristics	
General information	Model and type of study • Rigolin 2017 (Ferrera cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • not reported
Participants	Number of included persons in the cohort • 335 persons Setting • Italy, single-centre study Recruitment period • 2006 - 2016 Age (in years)rri • median 68.7 years (range: 33 - 96 years) Sex

CLL-IPI V - Rigolin 2017 (Ferrera cohort) *(Continued)*

	• 57.9% male Stages of disease • Rai: not reported • Binet A: 77.9%; Binet B: 14.3%; Binet C: 7.8% Treatment • 114 received chemotherapy Inclusion criteria • all patients diagnosed with CLL at 'our center' Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • no changes Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study • overall survival: OS was calculated from the date of diagnosis until death due to any cause or until the last patient follow-up. Additional outcome(s) • time-to-first-treatment: TTFT was calculated as the interval between diagnosis and the start of first-line treatment. Outcome in model development • overall survival
Missing data	Participants with any missing data? • unclear or not reported If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • 335 participants, number of events not reported Which model was used? • simplified model Was the model updated? • no

CLL-IPI V - Rigolin 2017 (Ferrera cohort) *(Continued)*

	Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • no time point and numerical results reported, see figure 1
PROBAST: Applicability	Domain 1: Participant selection • unclear - no eligibility criteria Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "... Fondo di Ateneo per la Ricerca 2013, 2014, 2016 of the University of Ferrara (G.M.R., A.C., M.N.), Programma Ricerca Regione Università 2007-2009 University of Ferrara (G.M.R., A.C.), Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (A.C.), Ricerca Finalizzata (A.C.; project RF-2011-02349712), Ministero ell'Istruzione, dell'Università e della Ricerca PRIN 2015 (A.C.; project 2015ZMRFEA), and AIL Ferrara. E.V. and E.S. are supported by AIL Ferrara." • "The authors declare no competing financial interests." Other comments • The authors provided their primary data and, in addition, a screen capture of the Stata output of their analysis.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way (single-centre study)
Domain 3: Outcome	Unclear	Objective standard outcome, however, observation time not reported
Domain 4: Analysis	No	No information on number of events, missing data and missing data handling; no performance measures reported
Overall judgement	No	

CLL-IPI V - Zhu 2018 (Chinese cohort)

Study characteristics	
General information	Model and type of study

CLL-IPI V - Zhu 2018 (Chinese cohort) *(Continued)*

- Zhu 2018 (Chinese cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

- Chinese

Study design

- retrospective cohort

Follow-up time

- median 48 months, (range: 1 - 192 months)

Participants	Number of included persons in the cohort • 215 persons Setting • China, single-centre study Recruitment period • 2002 - 2017 Age (in years) • median 60 years (range 16 - 85 years) Sex • 66.5% male Stages of disease • Rai 0: 16.7%; Rai I-II: 54.0%; Rai III-IV: 29.3% • Binet A: 44.2%; Binet B: 27.9%; Binet C: 27.9% Treatment • 147 received any of the following: • benzodiazepine • FCR (fludarabine, cyclophosphamide, rituximab), • bendamoxetine • rituximab • R+FFP+HDMP (rituximab + fresh frozen plasma + high-dose methylprednisolone) Inclusion criteria • newly-diagnosed CLL patients treated in our hospital between 2002 and 2017 Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

CLL-IPI V - Zhu 2018 (Chinese cohort) *(Continued)*

	• TP53 status (binary; deleted or mutated); IgHV mutational status (continuous: categorical/binary transformation; cut-off not given but usually 98%); ß2-microglobulin concentration (continuous: categorical/binary transformation; binary, cut-off 3.5 mg/L (based on literature)); clinical stage (Rai or Binet) (categorical); age (continuous: categorical/binary transformation; cut-off 65 years) • no changes Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: OS time was defined as the interval from a definitive diagnosis to the end point of death or follow-up for any cause. Additional outcome(s) • time-to-first-treatment: TTFT was defined as the interval from the definitive diagnosis to the initiation of treatment. Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 215 persons, 46 events • time-to-first-treatment: 215 persons, 147 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: was provided upon request (observed-expected ratio per risk group, calibration plot, Hosmer and Lemeshow test) • Discrimination: reported (AUC) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-1);(2-3);(4-6); (7-10)) • at 5 years • percentage of group without event at time point (low risk: 97.6%, intermediate risk: 83.7%, high risk: 67.8%, very high risk: 55.2%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - no eligibility criteria reported. Domain 2: Predictors

CLL-IPI V - Zhu 2018 (Chinese cohort) *(Continued)*

	• low Domain 3: Outcome • low
Notes	Funding & conflict of interest • not reported Other comments • original study translated by Yuan Chi (Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, China) • the authors provided additional data upon request (calibration plot, O:E ratios per risk group, AUC).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	No eligibility criteria reported, unclear if patients with missing values were excluded
Domain 2: Predictors	Yes	Well-established predictors used - not explicitly stated, but predictors probably assessed in a similar way
Domain 3: Outcome	Yes	Objective standard outcome
Domain 4: Analysis	No	Number of events low (46) No information on missing values The authors provided the calibration and discrimination upon request.
Overall judgement	No	

GCLLSG D - Pflug 2014 (GCLLSG)

Study characteristics	
General information	Model and type of study • Pflug 2014 (development cohort) • Development study Secondary citations • Tam 2014 Language of publication • English Study design • combined data (e.g. IPD or combination of cohorts); data from 3 RCTs Follow-up time

GCLLSG D - Pflug 2014 (GCLLSG) *(Continued)*

- median 63.4 months (range not reported)

Participants	Number of included persons in the cohort • 2007 persons Setting • multicentre study • Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain Recruitment period • 1997 - 2006 Age (in years) • median 60 years (range: 30 - 81 years) Sex • 68.4% male Stages of disease • Rai 0: 20.8%; Rai I: 18.4%; Rai II: 37.9%; Rai III: 8.6%; Rai IV: 14.3% • Binet A: 42.6%; Binet B: 37.7%; Binet C: 19.8% Treatment • participants were treated with: • early fludarabine (F) • F plus cyclophosphamide (FC) • FC plus rituximab (FCR) Inclusion criteria • different inclusion criteria per clinical trial (CLL1, CLL4 and CLL8 trials); common criteria: confirmed diagnosis of B-cell chronic lymphocytic leukaemia, untreated patient, life expectancy above 6 months Exclusion criteria • different exclusion criteria per clinical trial (CLL1, CLL4 and CLL8 trials)
Predictors	Number of candidate predictors • univariable analysis: 20 • multivariable analysis: 18 List of predictors in final model (including cut-points for dichotomised factors) • sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) Timing of predictor measurement • not reported, presumably at study entry
Outcome(s)	Primary outcome in study

GCLLSG D - Pflug 2014 (GCLLSG) *(Continued)*

- overall survival (the main end point of statistical analyses was OS defined as the time between registration/randomisation and death)

Additional outcome(s)

- not applicable

Missing data	Participants with any missing data? • yes (784 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 1223 participants, number of events included for analysis unclear Predictor selection method • univariable analysis, during multivariable modelling: stepwise regression, (P value or CI) Statistical method • Cox proportional hazard model Simplification of model? • yes Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14)) • at 5 years • percentage of group without event at time point (low risk: 95.2 %, intermediate risk: 86.9 %, high risk: 67.6 %, very high risk: 18.7 %)
PROBAST: Applicability	Domain 1: Participant selection • unclear - the study used RCT data with specific inclusion and exclusion criteria, patient population younger than average Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society." • "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or

GCLLSG D - Pflug 2014 (GCLLSG) *(Continued)*

holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: combination of three RCT cohorts
Domain 2: Predictors	Yes	Not explicitly stated in this publication, but we assumed standard assessment in the context of each RCT
Domain 3: Outcome	Yes	Objective standard outcome, follow-up in the context of each individual RCT
Domain 4: Analysis	No	Univariable selection of predictors Bootstrapping, however only for multivariable modeling and not used for correction of optimism; complete case analysis, background sample similar but many missing values; calibration not reported; simplification of the model, thus the assigned weights did not correspond to the results of multivariable modeling.
Overall judgement	No	

GCLLSG V - Molica 2015 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Molica 2015 (O-CLL1-GISL) • Validation study Secondary citations • Molica 2013; Molica 2017; Molica 2017 Language of publication • English Study design • prospective cohort Follow-up time • median 42 months (range: 1 - 82 months; 2038 person-years)
Participants	Number of included participants in the cohort

GCLLSG V - Molica 2015 (O-CLL1-GISL) *(Continued)*

- 337 participants

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- median 61 years (range: 33 - 70 years)

Sex

- 57.1% male

Stages of disease

- Rai 0: 77.8%; Rai I-IV: not reported
- Binet A: 100 %

Treatment

- 91 received treatment
- unclear which treatment

Inclusion criteria

- newly-diagnosed CLL patients from several Italian Institutions who were seen within 12 months of diagnosis were prospectively enrolled into the OCLL1-GISL protocol. The inclusion criteria for CLL diagnosis, employed at the time of study design, were those of the National Cancer Institute (NCI)-sponsored Working Group guidelines (NCI-WG). Patients enrolled did not require therapy according to NCI guidelines (i.e. asymptomatic Binet stage A).

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) • TK was omitted. Timing of predictor measurement • at study entry, within 12 months of diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment (TTFT was defined as the interval between the date of database registration and the date of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied).

GCLLSG V - Molica 2015 (O-CLL1-GISL) *(Continued)*

	Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 participants, 91 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • no
PROBAST: Applicability	Not applicable; outcome did not match primary outcome of the model.
Notes	Funding & conflict of interest • "Contract grant sponsor: NIH; Contract grant number: 1R01CA197120-01."; "Contract grant sponsors: AIRC (the Italian Association for Cancer Research), a non-profit organization to FM. “Special Program Molecular Clinical Oncology - 5 per mille” n. 9980, 2010/15 and AIRC “Innovative immunotherapeutic treatments of human cancer” n.16695, 2015/18." (Molica 2016) • "Conflict of interests: nothing to report." (Molica 2016) Other comments • The validation of this model in the O-CLL1-GISL cohort appeared in several publications. The c-statistic was reported differently in each of them. • The authors did not reply to our request for additional information.

GCLLSG V - Pflug 2014 (Mayo cohort)

Study characteristics	
General information	Model and type of study • Pflug 2014 (Mayo clinic cohort, undefined time range) • Validation study

GCLLSG V - Pflug 2014 (Mayo cohort) *(Continued)*

	Secondary citations • not applicable Language of publication • English Study design • prospective cohort Follow-up time • median 57 months (range not reported)
Participants	Number of included participants in the cohort • 676 participants Setting • USA, multicentre study Recruitment period • not reported Age (in years) • median 61.5 years (range: 32 - 89 years) Sex • 67% male Stages of disease • Rai 0: 57.1%; Rai I: 34%; Rai II: 5.9%; Rai III: 1%; Rai VI: 1.9% • not reported Treatment • 190 received treatment • unclear which treatment Inclusion criteria • consecutive series of 676 newly-diagnosed, prospectively-followed CLL patients cared for at Mayo Clinic who had baseline data on all considered variables except s-TK and/or s-B2M available and who had stored serum collected less than 36 months (median, 1 month) of diagnosis available for s-TK and s-B2M analysis. Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation

GCLLSG V - Pflug 2014 (Mayo cohort) *(Continued)*

	after univariable analysis); serum ß2-microglobulin (continuous, dichotomisation after univariable analysis) • TK from frozen blood samples within 36 months of diagnosis; different assessment methods for beta2m than in development study Timing of predictor measurement • at study entry
Outcome(s)	Primary outcome in study • overall survival: the main end point of statistical analyses was OS defined as the time between registration/randomisation and death. Additional outcome(s) • time-to-first-treatment: TFS was calculated from the date of diagnosis to the start of the first CLL treatment. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (3 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 673 participants, 85 events • time-to-first-treatment: 673 participants, 190 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14)) · • at 5 years • percentage of group without event at time point (low risk: 95.2%, intermediate risk: 91.4%, high risk: 71.7%, very high risk: 13.6%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - unclear recruitment period, data availability as inclusion criterion Domain 2: Predictors • low - differences in the measurement of one predictor, which is usually not available in the USA as compared to Germany (serum thymidine kinase): interassay calibration

GCLLSG V - Pflug 2014 (Mayo cohort) *(Continued)*

	Domain 3: Outcome • low
Notes	Funding & conflict of interest • "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society." • "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and-Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Only patients with available data included: "... who had baseline data on all considered variables except s-TK and/or s-b2m available and who had stored serum collected > 36 months ..."
Domain 2: Predictors	Yes	Standard laboratory measures used. Where s-TK and s-B2M were not available, blood samples were shipped to Germany and analysed.
Domain 3: Outcome	Yes	Objective standard outcome - no clear description of outcome assessment - we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk (no info on e.g. registry, frequency of follow-up, etc.)
Domain 4: Analysis	No	Low number of events; patients with missing data were not enrolled in the study, rated in domain 1; calibration not reported
Overall judgement	No	

GCLLSG V - Rani 2018 (Indian cohort)

Study characteristics	
General information	Model and type of study • Rani 2018 (Indian cohort) • Validation study Secondary citations • not applicable Language of publication

GCLLSG V - Rani 2018 (Indian cohort) *(Continued)*

- English

Study design

- unclear

Follow-up time

- median 40.5 months (range: 1 - 215 months)

Participants	Number of included participants in the cohort • 198 participants Setting • India, unclear if single or multicentre study Recruitment period • not reported Age (in years) • median 60 years (range not reported) Sex • 77% male Stages of disease • Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2% • Binet A/B/C: not reported Treatment • 62 received chlorambucil-based therapy • 56 received rituximab-based therapy • 20 received other therapies Inclusion criteria • treatment-naive CLL patients according to Hallek 2008 Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • sex (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, dichotomisation after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study

GCLLSG V - Rani 2018 (Indian cohort) *(Continued)*

	• time-to-first-treatment (TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy) Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 140 participants, 89 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at (0-2); (3-5); (6-10); (11-14) • median survival (low risk: not reached, intermediate risk: 36 years, high risk: 4 years, very high risk: 0.01 years)
PROBAST: Applicability	Not applicable, outcome did not match primary outcome of the model
Notes	Funding & conflict of interest • "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work." • "The authors declare that they have no conflict of interest." Other comments • Graphs were provided to obtain survival at a specific time point, however, these graphs were printed very small.

GIMEMA D - Molica 2005 (GIMEMA cohort)

Study characteristics

GIMEMA D - Molica 2005 (GIMEMA cohort) *(Continued)*

General information	Model and type of study • Molica 2005 (GIMEMA cohort) • Development study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 54 months, range 4 - 309 months
Participants	Number of included participants in the cohort • 1138 participants Setting • Italy, multicentre study Recruitment period • 1991 - 2000 Age (in years) • median 65 years (range: 27 - 100 years) Sex • 54% male Stages of disease • Rai 0: 77.2 %; Rai I-III: 22.7% • Binet A: 100% Treatment • 133 received chlorambucil • 20 received fludarabine alone or in association with cyclophosphamide • 7 received CHOP-like regimens • 40 received other therapies (CVP, cyclophosphamide with or without prednisone) Inclusion criteria • All previously untreated B-cell CLL patients in Binet stage A whose diagnosis was immunologically confirmed (CD5 +/ SmIg weak) and who were observed at different GIMEMA centres during the period 1991 – 2000 were considered eligible for this study. Exclusion criteria • inadequate follow-up
Predictors	Number of candidate predictors

GIMEMA D - Molica 2005 (GIMEMA cohort) *(Continued)*

- univariable analysis: 10 or more
- multivariable analysis: 6

List of predictors in final model (including cut-points for dichotomised factors)

- Lymphocyte doubling time (LDT, continuous: categorical/binary transformation; dichotomised to above and below 12 months); absolute peripheral blood lymphocytosis (continuous: categorical/binary transformation; two categories, above and below 30x10^9/l); Rai stage (categorical; 0 vs. I-III); gender (binary)

Timing of predictor measurement

- at diagnosis

Outcome(s)	Primary outcome in study • progression-free survival (progressive disease was defined on the basis of a shift to a more advanced clinical stage (i.e. from A to B or C) and/or LDT larger than 12 months. For those individuals who did not change Binet stage nor doubled their lymphocyte count, disease progression was considered when one of the following conditions was met: decrease in haemoglobin larger than 3 g/dL, splenomegaly larger than 6 cm below the left costal margin, massive lymphadenopathy (i.e. larger than 8 cm) or progressive lymphadenopathy and/or peripheral blood lymphocytosis more than 100x10^9 /l) Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • yes (545 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • progression-free survival: 593 participants, 235 events in overall sample, unclear number of events in sample used for modeling Predictor selection method • univariable analysis, during multivariable modelling: unclear if stepwise regression was used (criteria were P value or CI) Statistical method • Cox proportional hazard model, stratification of the resulting score by gender Simplification of model? • yes Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 3 risk groups (risk groups were: no risk factors; one risk factor; more than one risk factor) • at 10 years

GIMEMA D - Molica 2005 (GIMEMA cohort) *(Continued)*

	• percentage of group without event at time point (low risk: 67.8%, intermediate risk: 41%, high risk: 24.8%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • not reported Other comments • The model development characteristics were not well reported.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used (retrospective cohort); 1.7% of patients were excluded due to inadequate follow-up. We rated this as low proportion of exclusions.
Domain 2: Predictors	Yes	The predictors used seem to be routine demographic and laboratory information, and although there was no explicit information, we think that the methods were consistent.
Domain 3: Outcome	No	Knowledge of the predictor was used both in the definition of progression and the model (peripheral blood lymphocytes).
Domain 4: Analysis	No	Individuals with missing data were excluded: "The outcome of the 593 patients with complete data was different from the outcome of the 545 patients with incomplete data (10 year PFS 55.5% vs. 70.2%). The reasons for this difference rely on the higher number of patients in Rai stages I – III (P = 0.01) and increased serum levels of b2-m (P = 0.003) found in the subgroup with complete data"; univariable selection of predictors; no weighting of predictors: "These differences in the RR did not prevent us from constructing a risk score simply by adding the negative factors present in a single patient at the time of diagnosis."; no model performance measures reported
Overall judgement	No	

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.)

Study characteristics	
General information	Model and type of study • Gonzalez Rodriguez 2009 (Hospital de Cabuenes) • Validation study

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) *(Continued)*

	Secondary citations • not applicable Language of publication • Spanish Study design • retrospective cohort Follow-up time • not reported
Participants	Number of included participants in the cohort • 265 participants Setting • Spain, single-centre study Recruitment period • 1997 - 2007 Age (in years) • median 71.7 years (range: 42 - 94 years) Sex • 58.5% male Stages of disease • modified Rai low: 61.2%; Rai intermediate: 29.3%; Rai high: 9.5% • Binet A: 76.8%; Binet B: 14.8%; Binet C: 8.4% Treatment • not reported Inclusion criteria • Only cases with an increase of more than 5*10^9/l lymphocytes with morphology and immunophenotype of B-CLL and with Royal Mardsen Scoring System greater than 3, and cases of CLL confirmed by lymph node biopsy have been included. The inclusion criteria were based on the initial data at the time of diagnosis. Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Lymphocyte doubling time (LDT, continuous: categorical/binary transformation; dichotomised to above and below 12 months); absolute peripheral blood lymphocytosis (continuous: categorical/binary transformation; two categories, above and below 30x10^9/l); Rai stage (categorical; 0 vs. I-III); gender (binary) • no changes Timing of predictor measurement

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) *(Continued)*

- at diagnosis

Outcome(s)	Primary outcome in study • overall survival: to estimate the overall survival (OS), the date of diagnosis and the date of death for any cause or the final date of the study were considered. Additional outcome(s) • time-to-first-treatment: treatment-free survival was considered as the time elapsed from diagnosis until the start date of the treatment. Outcome in model development • progression-free survival
Missing data	Participants with any missing data? • yes (8 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 257 participants, 76 events • time-to-first-treatment: 257 participants, 67 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (AUC) Creation of risk groups? • yes, 3 risk groups • at 5 years • percentage of group surviving at 5 years (low risk: 79%, intermediate risk: 67%, high risk: 79%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • This work was partially funded through research project FIS06/ 0841 from the Health Research Fund.

GIMEMA V - González Rodríguez 2009 (Cabueñes coh.) *(Continued)*

Other comments

- The original study was translated by Leonardo Perales-Guerrero (medical student, Universidad de Guadalajara, Mexico).
- The authors could not be contacted since the email address of the contact author as stated in the publication was no longer valid.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: retrospective cohort
Domain 2: Predictors	Yes	The predictors used seem to be routine demographic and laboratory information, and although there was no explicit information, we think that the methods are consistent.
Domain 3: Outcome	Unclear	No median observation time; standard outcome, but treatment indication not described
Domain 4: Analysis	No	Only small amount of missing outcome data (n = 8); discrimination reported in form of AUC, however, time point not clear (probably 5 years); calibration not reported
Overall judgement	No	

MDACC 2007 D - Wierda 2007 (MDACC)

Study characteristics	
General information	Model and type of study • Wierda 2007 (development, MDACC 1981-2004) • Development study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 58.8 months, 95% CI: 55.2 - 61.2
Participants	Number of included participants in the cohort • 1674 participants Setting • US (Texas), single-centre study

MDACC 2007 D - Wierda 2007 (MDACC) *(Continued)*

	Recruitment period • 1981 - 2004 Age (in years) • median 58 years (range: 0 - 90 years) Sex • 1029 male participants Stages of disease • Rai 0: 469 individuals; Rai I: 739 individuals; Rai II: 235 individuals; Rai III: 93 individuals; Rai IV: 127 individuals • Binet A: 1019 individuals; Binet B: 470 individuals; Binet C: 168 individuals Treatment • 902 received treatment • unclear which treatment Inclusion criteria • Previously untreated patients who presented for initial evaluation to MDACC from August 1981 through August 2004 were included in this analysis. Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: 18 • multivariable analysis: 16 List of predictors in final model (including cut-points for dichotomised factors) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); sex (binary) Timing of predictor measurement • at study entry
Outcome(s)	Primary outcome in study • overall survival (the time interval was measured from the day of presentation to MDACC until death or last follow-up. Death from all causes was included). Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • yes (57 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported)

MDACC 2007 D - Wierda 2007 (MDACC) *(Continued)*

- overall survival: 1617 participants, 437 events

Predictor selection method

- univariable analysis, during multivariable modelling: backward elimination (P value or CI)

Statistical method

- Cox proportional hazard model

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: reported (c-statistic)

Creation of risk groups?

- yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 8))
- at 5 years
- percentage of group without event at time point (low risk: 0.97%, intermediate risk: 0.8%, high risk: 0.55%)

PROBAST: Applicability

Domain 1: Participant selection

- low

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- Funding not reported
- "The authors declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, consecutive sampling
Domain 2: Predictors	Yes	Routine clinical data used: although the recruitment period was very broad, standard clinical data (gender, age, Rai stage, ALC and beta-2 microglobulin) have been included for modelling
Domain 3: Outcome	Yes	Objective standard outcome: no clear description of outcome assessment; we assumed that assessment of the objective outcome OS was similar and therefore did not rate as high risk.

MDACC 2007 D - Wierda 2007 (MDACC) *(Continued)*

Domain 4: Analysis	Yes	Number of missing values and therefore the number of patients excluded from analysis was low; continuous factors handled appropriately; performance measures of original model reported; bootstrap correction applied; original formula and a simplified score provided; univariable selection of factors, however, nearly all factors were included in multivariable modelling as well.
Overall judgement	Yes	

MDACC 2007 V - Bulian 2011 (Italian-Swiss)

Study characteristics	
General information	Model and type of study • Bulian 2011 (Italian-Swiss cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 63.6 months, range not reported
Participants	Number of included participants in the cohort • 1480 participants Setting • Italy, Switzerland, multicentre study Recruitment period • 1996 - 2011 Age (in years) • median 65 years (range: 21 - 94 years) Sex • 58.2% male Stages of disease • Rai 0-II: not reported; Rai III/IV: 7.2% • Binet A: 74.6%; Binet B: 18.4%; Binet C: 6.9% Treatment • 306 received chemotherapy

MDACC 2007 V - Bulian 2011 (Italian-Swiss) *(Continued)*

- 129 received Chemoimmunotherapy
- 42 received FCR
- 49 received FR
- 38 received R-other
- 40 with missing data

Inclusion criteria

- 'The series included all cases observed at each centre over a given time period, although each centre allegedly failed to observe all the incident CLL cases.'

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) • no changes Timing of predictor measurement • not explicitly reported, but probably at diagnosis, as this was a retrospective cohort using standard demographic and clinical information
Outcome(s)	Primary outcome in study • overall survival Additional outcome(s) • time-to-treatment ('Patient treatment was started according to NCI-WG indications.') Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 1037 participants, 151 events Which model was used? • The authors tried to use the nomogram, but had problems due to the low quality of the graph in the development publication; additionally, they used the index score. Was the model updated? • yes • 'An additional model was developed by replacing the dichotomous variables Rai staging (0–I–II vs. III–IV) with the three-level Binet staging.' Performance measures reported?

MDACC 2007 V - Bulian 2011 (Italian-Swiss) *(Continued)*

	• Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3), (4-7), (>7)) • at 5 years • percentage of group without event at time point (low risk: 97%, intermediate risk: 89%, high risk: 53%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Ministero della Salute (Ricerca Finalizzata I.R.C.C.S. and 'Alleanza Contro il Cancro'), Rome; Associazione Italiana contro le Leucemie, linfomi e mielomi (A.I.L.), Venezia Section, Pramaggiore Group; Ricerca Scientifica Applicata, Regione Friuli Venezia Giulia, Trieste ('Linfonet' project); Associazione Italiana per la Ricerca sul Cancro (Investigator Grant IG-8701), Milan, Italy; Helmut Horten Foundation, San Salvatore Foundation, Fondazione per la Ricerca e la Cura sui Linfomi (Lugano, Switzerland); Swiss Cancer League (Krebsliga Schweiz) grant KLS-01690-03-2005." • "The authors declare no competing financial interests." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	"Clinical and biological data of 1480 untreated CLL patients were retrospectively collected from 8 centres. The series included all cases observed at each centre over a given time period, ..."
Domain 2: Predictors	Yes	Measurements were taken in 8 different centres. However, the authors standardised the predictors so this should not be of influence.
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Exclusion of individuals with missing values: "Some CLL cases with missing values were excluded from multivariate models. The death rate in this group was higher. The lack of inclusion of a number of individuals with possible worse prognosis could partly justify the observed bias in nomogram prediction."; expected OS estimated, but only point score: no discrimination calculated, because estimation based on nomogram not possible (quality of graph)
Overall judgement	No	

MDACC 2007 V - Gentile 2014 (Italian cohort)

Study characteristics	
General information	Model and type of study • Gentile 2014 (Italian multicentre cohort) • Validation study Secondary citations • Gentile 2016 Language of publication • English Study design • retrospective cohort Follow-up time • median 68.4 months, range not reported
Participants	Number of included participants in the cohort • 1502 indviduals Setting • Italy, multicentre study Recruitment period • 1983 - 2013 Age (in years) • median 67 years (range: not reported) Sex • 55.7% male Stages of disease • Rai 0: 56.5%; Rai I: 21.4%; Rai II: 15%; Rai III: 2.6%; Rai IV: 4.5% • Binet A: 82.9%; Binet B: 11.6%; Binet C: 5.5% Treatment • 337 received chemotherapy • 142 received chemoimmunotherapy Inclusion criteria • The CLL databases of four Italian centres, which included all patients diagnosed with CLL since 1983, were utilised for research purposes. Exclusion criteria • Missing data, ß2m
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

MDACC 2007 V - Gentile 2014 (Italian cohort) *(Continued)*

	• Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) Timing of predictor measurement • at study entry
Outcome(s)	Primary outcome in study • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. Additional outcome(s) • time-to-first-treatment: for TFS, it was from the day of CLL diagnosis until therapy start or last follow-up. Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 1502 participants, 277 events • time-to-first-treatment: 1502 participants, 479 events Which model was used? • Authors used both nomogram and the index score, and also compared with previous validations. Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) • at 5 years • percentage of group without event at time point (low risk: 89%, intermediate risk: 82%, high risk: 42%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - very long period of retrospective enrolment: patients were probably treated differently. Inclusion of cMBL patients previously classified as CLL - sensitivity analysis showed no difference in discrimination. Domain 2: Predictors • low

MDACC 2007 V - Gentile 2014 (Italian cohort) *(Continued)*

	Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding not reported • "The authors indicated no potential conflicts of interest." Other comments • 4 of the 5 Italian centres overlapped with the cohort in the secondary citation. This paper was chosen as primary because more information was provided.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort: "... which included all patients diagnosed with CLL since 1983, were utilized for research purposes."
Domain 2: Predictors	Yes	Routine clinical data used: although the recruitment period was very broad, standard clinical data (gender, age, Rai stage, ALC and beta-2 microglobulin) have been included for modelling.
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Individuals with missing values were excluded. Although background characteristics were similar, we rated this as high risk because nearly half of all patients were excluded from analysis; measure for calibration not reported, however, reporting of survival per risk group
Overall judgement	No	

MDACC 2007 V - Gentile 2016 (Mayo cohort)

Study characteristics	
General information	Model and type of study • Gentile 2016 (Mayo clinic cohort) • Validation study Secondary citations • Gentile 2016; Shanafelt 2009 Language of publication • English Study design • prospective cohort Follow-up time

MDACC 2007 V - Gentile 2016 (Mayo cohort) *(Continued)*

- median 86.4 months, range not reported

Participants	Number of included participants in the cohort • 506 participants Setting • USA, single-centre study Recruitment period • 2001 - 2008 (correspondence with authors) Age (in years) • median 62.5 years (range: 36 - 89 years) Sex • 68.6% male Stages of disease • Rai 0: 62.5%; Rai I: 28.9%; Rai II: 5.1%; Rai III: 1.2%; Rai IV: 2.4% • Binet A: not reported Treatment • 213 received treatment • unclear which treatment Inclusion criteria • not reported Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: for OS, the time interval was measured from the day of CLL diagnosis until death from all causes or last follow-up. Additional outcome(s) • time-to-first-treatment: for TTFT from the day of CLL diagnosis until the start of therapy or last follow-up Outcome in model development • overall survival

MDACC 2007 V - Gentile 2016 (Mayo cohort) *(Continued)*

Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 506 participants, 114 events • time-to-first-treatment: 506 participants, 213 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) • at 5 years • percentage of group without event at time point (low risk: 98.3%, intermediate risk: 84.4%, high risk: 15.6%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - no clear eligibility criteria Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • Funding not reported • "S.A.P. received funding from Pharmacyclics. All other authors have no conflict of interest to disclose." Other comments • The corresponding author provided additional information to describe the cohort (recruitment period, overlap of cohort with other publications)

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Missing values as exclusion reason: 'The validation data set consisted of a consecutive series of 506 newly-diagnosed CLL patients, prospectively followed

MDACC 2007 V - Gentile 2016 (Mayo cohort) *(Continued)*

		at the Mayo Clinic, with baseline data available for all variables considered for both CLL-IPI and the MDACC score.'
Domain 2: Predictors	Yes	Routine clinical data used, prospective follow-up, relatively recent cohort (2001-2008)
Domain 3: Outcome	Yes	Objective standard outcome, clear definition and assessment: 'death, which were abstracted from clinical records at the time of inclusion and updated on an ongoing basis.'
Domain 4: Analysis	Unclear	Individuals with missing data were excluded from the study, rated in domain 1. Calibration not reported; survival per risk groups was available.
Overall judgement	No	

MDACC 2007 V - González Rodríguez (Cabueñes)

Study characteristics	
General information	Model and type of study • Gonzalez Rodriguez 2009 (Cabuenes cohort) • Validation study Secondary citations • not applicable Language of publication • Spanish Study design • retrospective cohort Follow-up time • not reported
Participants	Number of included participants in the cohort • 265 participants Setting • Spain, single-centre study Recruitment period • 1997 - 2007 Age (in years) • median 71.7 years (range: 42 - 94 years) Sex • not reported

MDACC 2007 V - González Rodríguez (Cabueñes) *(Continued)*

Stages of disease

- reported as Rai modified stages; low risk: 61.2%; intermediate: 29.3%; high risk: 9.5%
- Binet A: 76.8%; Binet B: 14.8%; Binet C: 8.4%

Treatment

- 67 received treatment
- unclear which treatment

Inclusion criteria

- Only cases with an increase of more than 5 × 10^9/l lymphocytes with morphology and immunophenotype of B-CLL and with Royal Mardsen Scoring System greater than 3, and cases of CLL confirmed by lymph node biopsy have been included. The inclusion criteria were based on the initial data at the time of diagnosis.

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) • no changes Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival: to estimate overall survival, the diagnosis date was considered and the date of death by any cause or the final date of the study Additional outcome(s) • time-to-first-treatment: treatment-free survival was considered as the time elapsed from diagnosis until the date of treatment initiation. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (8 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 257 participants, 76 events • time-to-first-treatment: 257 participants, 67 events Which model was used? • index score (simplified model)

MDACC 2007 V - González Rodríguez (Cabueñes) *(Continued)*

	Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (AUC) Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) • at 5 years • percentage of group without event at time point (low risk: 87%, intermediate risk: 75%, high risk: 29%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • This work was partially funded through research project FIS06/0841 from the Health Research Fund. • Conflict of interest not reported Other comments • original study translated by Leonardo Perales-Guerrero (Medical student, Universidad de Guadalajara, Mexico) • data extraction from both the original and translated study

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, consecutive sampling: 'Un total de 265 pacientes se diagnosticaron de LLCB en el periodo de 10 anos.'
Domain 2: Predictors	Yes	Routine clinical data assessed at diagnosis
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Low number of events; individuals with missing values excluded from modelling, however this number was small; calibration not reported
Overall judgement	No	

MDACC 2007 V - Molica 2010 (GIMEMA cohort)

Study characteristics

General information	Model and type of study • Molica 2010 (GIMEMA cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 41.5 months, range not reported
Participants	Number of included participants in the cohort • 1158 participants • characteristics only reported for 310 participants with complete data Setting • Italy, multicentre study Recruitment period • 1991 - 2000 Age (in years) • median 64 years (range: not reported) Sex • 56.1% male Stages of disease • Rai 0: 67%; Rai I-III: 32.9%; Rai III/IV: 0% • Binet A: 100% Treatment • 51 received therapy • unclear which treatment Inclusion criteria • previously untreated CLL participants in Binet stage A whose diagnosis was immunologically confirmed (CD5+/SmIg weak) and who were observed at different GIMEMA primary haematology centres during the period 1991–2000 Exclusion criteria • not reported

MDACC 2007 V - Molica 2010 (GIMEMA cohort) *(Continued)*

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); sex (binary) • cut-off for Rai was adapted in the score Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment: active, therapy-requiring disease was defined by very detailed criteria. Additional outcome(s) • overall survival: definition not reported Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 310 participants, number of events not reported • time-to-first-treatment: 310 participants, 51 events Which model was used? • Authors used both nomogram and index score, and also compared with previous validations. Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) • at 5 years • percentage of group without event at time point (low risk: 99.3%, intermediate risk: 88.2%, high risk: no individuals in this category)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors

MDACC 2007 V - Molica 2010 (GIMEMA cohort) *(Continued)*

	• high - the authors changed the cut-points for the predictor 'Rai stage'. Instead of a binary variable stage 0-II vs II-IV, Molica 2010 used Rai 0 vs. Rai I-II Domain 3: Outcome • High - outcome did not match primary outcome of the model. Survival per risk group reported
Notes	Funding & conflict of interest • Funding not reported • "The authors reported no potential conflicts of interest." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort, inclusion of complete database: "The GIMEMA CLL database includes information on previously untreated CLL patients in Binet stage A whose diagnosis was immunologically confirmed".
Domain 2: Predictors	Yes	Although individuals came from 25 different centres, predictors were routine clinical data.
Domain 3: Outcome	Unclear	The median observation time was below 5 years (41.5 months); outcome time-to-first-treatment was not the primary outcome of the validated model.
Domain 4: Analysis	No	Individuals with missing values were excluded from the analysis ("The outcome of the 593 patients with complete data was different from the outcome of the 545 patients with incomplete data"); no performance measures reported
Overall judgement	No	

MDACC 2007 V - Molica 2015 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Molica 2015 (O-CLL1-GISL) • Validation study Secondary citations • Gentile 2016; Molica 2010; Molica 2017 (leukemia & lymphoma); Molica 2017 (American Journal of Hematology) Language of publication • English Study design • prospective cohort

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) *(Continued)*

	Follow-up time • median 42 months, range not reported
Participants	Number of included participants in the cohort • 337 participants Setting • Italy, multicentre study Recruitment period • 2007 - not reported Age (in years) • median 61 years (range: 33 - 70 years) Sex • 57.1% male Stages of disease • Rai 0: 77.5%; Rai I-II: 22.5% • Binet A: 100%; Binet B/C: not applicable Treatment • 91 received treatment • unclear which treatment Inclusion criteria • Established diagnosis of B-CLL by NCI criteria, performed by local haematologist (diagnosis will be confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation antigen 5 (CD5), 19 (CD19),23 (CD23)). Age above 18 years and below 70 years. Eastern Cooperative Oncology Group (ECOG) below 2. Binet stage A. Diagnosis performed within 12 months before inclusion in the study. Patients who do not necessitate therapy by NCI guidelines (watch-and-wait policy). Shipment of peripheral blood sample to centralised laboratory for biological assessment. Clinical data including baseline information on disease localisation and laboratory parameters at staging and assurance of follow-up updating for at least 3 years are requested. Written informed consent. Exclusion criteria • Patients with CLL whose diagnosis exceeded 12 months before registration. Patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis. Clinical Binet stage B or C. Patients who necessitated therapy according to NCI guidelines (no watch-and-wait policy). Age above 70 years. Without a written informed consent.
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) • no changes Timing of predictor measurement • at study entry, within 12 months of diagnosis

MDACC 2007 V - Molica 2015 (O-CLL1-GISL) *(Continued)*

Outcome(s)	Primary outcome in study • time-to-first-treatment: the primary endpoint, TTFT, was defined as the interval between the date of database registration and the date of first CLL treatment (individuals underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied). Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 participants, 91 events Which model was used? • Authors used nomogram/formula Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • no
PROBAST: Applicability	Not applicable, outcome did not match primary outcome of the model.
Notes	Funding & conflict of interest • "This work was supported by funding from Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN). We thank AIL Cosenza-Fondazione Amelia Scorza' onlus, Cosenza, Italy, and Brigida Gulino for precious secretarial assistance." (Gentile 2016) • Conflict of interest not reported Other comments • Cohort identical in several publications

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort)

Study characteristics	
General information	Model and type of study • Munoz-Novas 2018 (Spanish cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 46 months, range not reported
Participants	Number of included participants in the cohort • 696 participants Setting • Spain, multicentre study Recruitment period • 1989 - 2013 Age (in years) • median 67 years (range: 25 - 90 years) Sex • 64.2 % male Stages of disease • Rai 0-II: 93.8%; Rai III-IV: 6.2% • Binet A/B/C: not reported Treatment • 186 received treatment • unclear which treatment Inclusion criteria • all participants diagnosed with CLL Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) *(Continued)*

	• Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) Timing of predictor measurement • at diagnosis, data from medical records
Outcome(s)	Primary outcome in study • overall survival: overall survival (OS) was calculated from the time of diagnosis to death or last follow-up. Additional outcome(s) • time-to-first-treatment: time-to-first-therapy (TTFT) from the date of diagnosis to first treatment or last follow-up Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (213 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 483 participants, 92 events • time-to-first-treatment: 483 participants, 186 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups, (cut-points at (1-3);(4-7);(above 7)) • at 5 years • percentage of group without event at time point (low risk: 98.4%, intermediate risk: 79.6%, high risk: 51.1%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome

MDACC 2007 V - Muñoz-Novas 2018 (Spanish cohort) *(Continued)*

	• low
Notes	Funding & conflict of interest • Funding not reported • "The authors declare that there are no conflicts of interest regarding the publication of this article." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Retrospective cohort: "A total of 696 unselected CLL patients newly diagnosed and previously untreated from different institutions of the central region of Spain were included in this study." Patients with missing values were later excluded at analysis stage, rated in domain 4.
Domain 2: Predictors	Yes	Routine clinical data assessed at diagnosis was used (from medical records).
Domain 3: Outcome	Yes	Objective standard outcome: no definition of outcome and assessment - we assumed that definition and assessment of the objective outcome OS was similar and therefore did not rate as high risk.
Domain 4: Analysis	No	Low number of event; exclusion of patients with missing values from analysis; calibration not reported
Overall judgement	No	

MDACC 2007 V - Pflug 2014 (3 RCTs)

Study characteristics	
General information	Model and type of study • Pflug 2014 (German CLL study group cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • combined data (e.g. IPD or combination of cohorts) Follow-up time • median 63.4 months, range not reported
Participants	Number of included participants in the cohort • 2007 participants

MDACC 2007 V - Pflug 2014 (3 RCTs) *(Continued)*

Setting

- several, multicentre study (Australia, Austria, Belgium, Czech Republic, Denmark, France, Germany, Israel, Italy, New Zealand, Spain)

Recruitment period

- 1997 - 2006

Age (in years)

- median 60 years (range: 30 - 81 years)

Sex

- 68.4% male

Stages of disease

- Rai 0: 20.8%; Rai I: 18.4%; Rai II: 37.9%; Rai III: 8.6%; Rai IV: 14.3%
- Binet A: 42.6%; Binet B: 37.7%; Binet C: 19.8%

Treatment

- participants were treated with:
- early fludarabine (F)
- F plus cyclophosphamide (FC)
- FC plus rituximab (FCR)

Inclusion criteria

- CLL1 trial: 18-75 years, confirmed diagnosis of B-cell chronic lymphocytic leukaemia, Rai stage 0-II (Binet stage A), life expectancy above 6 months, no autoimmune haemolytic anaemia, no thrombocytopenia, no severe organ dysfunction, no other prior or concurrent malignancy, no prior or other concurrent chemotherapy
- CLL4 trial: < 65 years with untreated CLL in advanced stage (all Binet stage C patients; Binet stage B with symptoms, which require therapy; Binet stage A with severe B-symptoms)
- CLL8 trial: >= 18 years, diagnosed B-cell chronic lymphocytic leukaemia (CLL) defined by the National Cancer Institute (NCI) Working Group criteria, Binet stage C disease OR Binet stage B disease AND 1 of other predefined signs or symptoms, ECOG performance status 0-1, Cumulative Illness Rating Scale (CIRS) score above 6, life expectancy above 6 months, no active infection, no pregnancy or nursing, no concurrent or previous treatment of CLL by chemotherapy, radiotherapy, or immunotherapy, and more

Exclusion criteria

- CLL4 trial: age < 18 and > 65 years were excluded as well as patients with any previous treatment of CLL, life expectancy less than six months and an Eastern Cooperative Oncology Group performance status of more than two. Patients were also excluded if they had severe organ dysfunction, concomitant or previous other neoplasms or an autoimmune haemolytic anaemia or thrombocytopenia.

Predictors

List of predictors used for validation (and changes between original predictors and predictors in validation study)

- Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary)
- no changes

Timing of predictor measurement

- at study entry (< 1 year: 63.3%, > 1 year: 36.7%)

MDACC 2007 V - Pflug 2014 (3 RCTs) *(Continued)*

Outcome(s)	Primary outcome in study • overall survival: OS was defined as the time between registration/randomisation and death. Additional outcome(s) • progression-free survival: PFS was calculated from registration/randomisation to start of the first CLL treatment or disease progression or death. Outcome in model development • overall survival
Missing data	Participants with any missing data? • yes (863 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 1144 participants, number of events not reported in publication - 180 at 5 years Which model was used? • c-statistic reported for the nomogram Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • no
PROBAST: Applicability	Domain 1: Participant selection • unclear - restricted inclusion and exclusion criteria based on RCTs Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "This manuscript was written on behalf of the German CLL Study Group. Studies CLL1, CLL4, and CLL8 were planned and conducted as investigator-initiated trials by the German CLL Study Group and were supported by research grants from German Cancer Aid, Medac Schering Onkologie, and F. Hoffmann-La Roche. T.D.S. is a clinical scholar of the Leukemia Lymphoma Society." • "N.P. received Travel Grants from Roche. T.D.S. received research grants from Genentech, Celgene, Glaxo-Smith-Kline, Cephalon, Hospira, and Polyphenon E International. B.E. is a consultant and/or holds an advisory role for Celgene and Pharmacyclics and has received honoraria and research funding from Roche and Mundipharma. S.S. is a consultant and/or holds an advisory role for Roche and

MDACC 2007 V - Pflug 2014 (3 RCTs) *(Continued)*

Mundipharma, and received honoraria and research funding from both. H.D. received research grants from Roche. U.J. received honoraria and research funding from Roche. M.H. is a consultant and/or holds an advisory role and received research funding from Roche. J.B., M.A.B., T.E., K.B., G.M., K.G.R., M.J.E., G.H., R.B., A.-M.F., C.-M.W., K.F., and N.E.K. declare no competing financial interests."

Other comments

- The authors provided additional data upon request (classification tables).

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Overlap of individuals between trials prevented; only concern could be comparability of the three trials.
Domain 2: Predictors	Yes	Prospective collection of predictors within clinical trials
Domain 3: Outcome	Yes	Objective standard outcome: follow-up in frame of each RCT
Domain 4: Analysis	No	Number of events sufficient; individuals with missing data were excluded from model development. It was not clear if this was the case for the validation of the MDACC2007 model; calibration not reported
Overall judgement	No	

MDACC 2007 V - Rani 2018 (Indian cohort)

Study characteristics	
General information	Model and type of study • Rani 2018 (Indian cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • unclear Follow-up time • median 40.5 months (range not reported)
Participants	Number of included participants in the cohort • 198 participants Setting • India, unclear if single or multicentre study Recruitment period

MDACC 2007 V - Rani 2018 (Indian cohort) *(Continued)*

- not reported

Age (in years)

- median 60 years (range: not reported)

Sex

- 77% male

Stages of disease

- Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2%
- Binet A/B/C: not reported

Treatment

- 62 received chlorambucil-based therapy
- 56 received rituximab-based therapy
- 20 received other therapies

Inclusion criteria

- treatment-naive CLL patients according to Hallek 2008

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study • overall survival: OS was defined as the time from the date of diagnosis to date of death or date of last follow-up. Additional outcome(s) • not applicable for this model (TTFT was assessed in this cohort for other prognostic models) Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 198 participants, 86 events

MDACC 2007 V - Rani 2018 (Indian cohort) *(Continued)*

	Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups • survival probabilities not reported
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period and eligibility criteria unclear Domain 2: Predictors • unclear - no information on timing of predictor because the study design was unclear Domain 3: Outcome • low
Notes	Funding & conflict of interest • "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work." • "The authors declare that they have no conflict of interest." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No clear eligibility criteria; possibly missing values as exclusion reason; no recruitment period
Domain 2: Predictors	Yes	Study design unclear, "... were obtained from medical records of the patients". Routine clinical predictors were used in this model, therefore we rated this study as low.
Domain 3: Outcome	Unclear	Short follow-up time
Domain 4: Analysis	No	Low number of events; calibration not reported; survival per risk groups graphically presented
Overall judgement	No	

MDACC 2007 V - Trajkova 2013 (Macedonia)

Study characteristics	
General information	Model and type of study • Trajkova 2013 (Macedonian cohort) • Validation study Secondary citations • not applicable Language of publication • English Study design • unclear Follow-up time • not reported
Participants	Number of included participants in the cohort • 100 participants Setting • Macedonia, single-centre study Recruitment period • 2011 - 2013 Age (in years) • median 64.8 years (range: 47 - 78 years) Sex • 63% male Stages of disease • Rai 0: 54%; Rai I: 5%; Rai II: 32%; Rai III: 5%; Rai IV: 4% • Binet A/B/C: not reported Treatment • 41 received treatment • unclear which therapy Inclusion criteria • Our study incorporated 100 consecutive treatment-naive CLL patients with IgHV mutational status. Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study)

MDACC 2007 V - Trajkova 2013 (Macedonia) *(Continued)*

	• Age (continuous: no transformation; in years); ALC (absolute lymphocyte count) (continuous: no transformation; x10^9/L); serum ß2-microglobulin (continuous: no transformation); nodal groups (binary); Rai stage (categorical; binary: stage III or IV vs. 0-II); gender (binary) Timing of predictor measurement • at presentation at University Haematology Clinic
Outcome(s)	Primary outcome in study • time-to-first-treatment: time-to-event end point was defined as the time from first visit to the University Haematology Clinic to first CLL treatment. There was no restriction of time from diagnosis to presentation at the University Haematology Clinic. Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 100 participants, 41 events Which model was used? • unclear Was the model updated? • unclear Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes • No reporting of observed survival per subgroup
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest

MDACC 2007 V - Trajkova 2013 (Macedonia) *(Continued)*

- not reported

Other comments

- "estimated and projected" survival without any comparison of observed survival provided; we could not use this information for our review and therefore, we did not continue with data extraction.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	"100 consecutive treatment naïve CLL patients"
Domain 2: Predictors	Yes	Routine clinical data
Domain 3: Outcome	Unclear	No observation time reported; outcome assessment not described
Domain 4: Analysis	No	Low number of participants and events; not really validating MDACC 2007 - reporting of projected survival only and additional univariable analysis
Overall judgement	No	

MDACC 2011 D - Wierda 2011 (MDACC)

Study characteristics	
General information	Model and type of study • Wierda 2011 (development (MDACC 2004-2009)) • Development study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 26 months (range 3 - 73 months)
Participants	Number of included participants in the cohort • 930 participants Setting • US (Texas), single-centre study Recruitment period • 2004 - 2009

MDACC 2011 D - Wierda 2011 (MDACC) *(Continued)*

	Age (in years) • median 59 years (range: 30 - 89 years) Sex • 61% male Stages of disease • Rai 0: 36%; Rai I: 51%; Rai II: 8%; Rai III-IV: 5% Treatment • 232 received treatment • unclear which treatment Inclusion criteria • no previous treatment; not recommended for first-line treatment at initial visit Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: more than 25 • multivariable analysis: unclear List of predictors in final model (including cut-points for dichotomised factors) • IgHV status (continuous: categorical/binary transformation); diameter of largest palpated cervical LN, cm (continuous: no transformation); FISH category (categorical; del(11q) or del(17p) vs. none); number of involved lymph node sites (continuous: categorical/binary transformation); lactate dehydrogenase IU/L/100 (continuous: no transformation); IgHV status*LDH (interaction term) Timing of predictor measurement • at study entry (median time from diagnosis to presentation to MDACC was 3.4 months, range 0 to 428 months)
Outcome(s)	Primary outcome in study • time-to-first-treatment: TTFT was defined as the time interval between the date of presentation to MDACC and date of first CLL treatment. Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • yes (243 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 687 participants, 193 events Predictor selection method

MDACC 2011 D - Wierda 2011 (MDACC) *(Continued)*

- univariable analysis, during multivariable modelling: stepwise regression mix, (P value or CI)

Statistical method

- Cox proportional hazard model

Simplification of model?

- yes

Performance measures reported?

- Calibration: not reported
- Discrimination: not reported

Creation of risk groups?

- no

PROBAST: Applicability

Domain 1: Participant selection

- unclear - individuals with missing values were excluded. We were unsure if the sample was representative for all individuals with CLL

Domain 2: Predictors

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "Supported in part by the Leukemia and Lymphoma Society Clinical Scholar program (W.G.W.)"
- "The author(s) indicated no potential conflicts of interest."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Individuals without evaluation of clinical factors and at least one newer factor were excluded: "We identified 930 previously untreated patients who presented to MD Anderson Cancer Center (MDACC) between January 2004 and December 2009, were not recommended for first-line treatment at initial visit, and were evaluated for traditional clinical and laboratory prognostic factors and one or more of the newer prognostic factors including IgHV mutation status, chromosome abnormalities by FISH analysis, and ZAP-70 expression by flow cytometry and/or immunohistochemistry (IHC)."
Domain 2: Predictors	Yes	Predictor assessment well described; comparison of assessment tools
Domain 3: Outcome	No	Some univariably tested predictors were not excluded from the outcome definition (progressive or symptomatic splenomegaly, massive nodes, progressive or symptomatic lymphadenopathy, Rai stage; Hallek 2008); short observation time
Domain 4: Analysis	No	Many predictors compared to events, 193 events (more than 25 predictors); univariable selection of predictors; unclear how many predictors entered in-

MDACC 2011 D - Wierda 2011 (MDACC) *(Continued)*

	to MV analysis; no performance measures reported; split sample, but after the model was already developed, which means that the data were internally validated on data that it was developed with. Bootstrapping was done, however not used to correct the model; assumption of proportionality was checked
Overall judgement	No

MDACC 2011 V - Molica 2016 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Molica 2013 (O-CLL1-GISL) • Validation study Secondary citations • Molica 2013; Molica 2014; Molica 2016; Molica 2017; Molica 2017 Language of publication • English Study design • prospective cohort Follow-up time • median 30 months (range 1 - 65 months)
Participants	Number of included participants in the cohort • 328 participants Setting • Italy, multicentre study Recruitment period • 2006 - 2010 Age (in years) • median 61 years (range: 33 - 70 years) Sex • 58.2% male Stages of disease • Rai 0: 76.5%; Rai I-II: 23.5% • Binet A: 100% Treatment • 91 received treatment • unclear which treatment

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) *(Continued)*

	Inclusion criteria • The inclusion criteria for CLL diagnosis, which were used at the time of study design and initiation, followed the National Cancer Institute-sponsored Working Group guidelines, which require absolute lymphocytosis with a lower threshold of above 5000 mature-appearing lymphocytes/mL in the peripheral blood. Because the objective of this observational study was to evaluate the role of novel prognostic variables in younger patients with CLL, only those aged above 70 years were considered eligible. Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • IgHV status (continuous: categorical/binary transformation); diameter of largest palpated cervical LN, cm (continuous: no transformation); FISH category (categorical; del(11q) or del(17p) vs. None); number of involved lymph node sites (continuous: categorical/binary transformation); lactate dehydrogenase IU/L/100 (continuous: no transformation); IgHV status*LDH (interaction term) Timing of predictor measurement • at study entry, within 12 months of diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment: TTFT was defined as the interval between the date of registration and the date of initiation of first CLL treatment. Patients underwent sequential monitoring, and the frequency of follow-up visits was individualised according to patient risk; this ranged from 3 to 6 months (median 6 months). All physicians who registered patients in this observational database stated that they had used the NCI-WG guidelines as a reference criterion for starting therapy. In particular, the absolute lymphocyte count was not used as the sole indicator for treatment. Active disease, requiring therapy, was defined when at least one of the criteria set out in the NCI-sponsored Working Group guidelines was satisfied. Additional outcome(s) • not applicable Outcome in model development • time-to-first-treatment
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 participants, 91 events • in Molica 2013: time-to-first-treatment: 328 participants, 68 events Which model was used? • Cox proportional hazard model Was the model updated? • no

MDACC 2011 V - Molica 2016 (O-CLL1-GISL) *(Continued)*

	Performance measures reported? • Calibration: reported (calibration plot) • Discrimination: reported (c-statistic) Creation of risk groups? • no
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "This work was supported by grants from Associazione Italiana Ricerca sul Cancro (AIRC) (to Antonino Neri -IG4569, MFIG10492 and Fortunato Morabito -RG6432) and AIRC-Special Program Molecular Clinical Oncology-''5 per mille'', grant 9980, 2010-15 to Antonino Neri, Manlio Ferrarini and Fortunato Morabito; Ricerca Finalizzata from Italian Ministry of Health 2006 (GC, FM and MF) and 2007 (GC)." • "The authors made no disclosures." Other comments • The O-CLL1-GISL cohort was used in various publications and to validate different CLL models.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospective cohort, detailed eligibility criteria in trial registry
Domain 2: Predictors	Yes	Predictors prospectively collected
Domain 3: Outcome	No	See development of the model: some predictors in the model were not excluded from the outcome definition (number and size of lymph nodes included in model, however these factors also formed Binet stage, which was a reason for treatment indication, Hallek 2008).
Domain 4: Analysis	No	Low number of events; no information on missing data reported
Overall judgement	No	

Morabito D - Morabito 2009 (Italian cohort)

Study characteristics	
General information	Model and type of study • Morabito 2009 (Development cohort (Italian)) • Development study

Morabito D - Morabito 2009 (Italian cohort) *(Continued)*

Secondary citations

- not applicable

Language of publication

- English

Study design

- not reported

Follow-up time

- median 36 months (range 12 - 180 months)

Participants	Number of included participants in the cohort • 262 participants Setting • Italy, multicentre study Recruitment period • not reported Age (in years) • median 65 years (range not reported) Sex • 61.5% male Stages of disease • Rai 0-IV: not reported • Binet A: 100% Treatment • not reported Inclusion criteria • diagnosis of typical CLL based on the National Cancer Institute (NCI) Working Group criteria (Hallek 2008) and confirmed by flow cytometry Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: 3 • multivariable analysis: not applicable List of predictors in final model (including cut-points for dichotomised factors) • IgHV mutational status (continuous: categorical/binary transformation; 2% threshold, as established in literature); CD38 (continuous: categorical/binary transformation; 30% threshold); ZAP-70 (continuous: categorical/binary transformation; threshold established by ROC curve) Timing of predictor measurement

Morabito D - Morabito 2009 (Italian cohort) *(Continued)*

- markers were tested within two years in 70% of cases, within 4 years in 15% and from 4 to 5 years from diagnosis in the remaining 15%

Outcome(s)	Primary outcome in study • time-to-first-treatment: TTT was measured from diagnosis to first-line treatment or last follow-up. Treatment was decided uniformly in all participating centres based on documented progressive and symptomatic disease according to NCI working guidelines. Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 252 participants, 51 events Predictor selection method • univariable analysis only (P value or CI); a scoring system was designed in which one point was assigned to each unfavourable prognostic marker and the final score was determined by the sum of unfavourable markers present. Statistical method • univariable analysis: Kaplan-Meier method Simplification of model? • not applicable Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • no
PROBAST: Applicability	Domain 1: Participant selection • unclear - not sufficient information (eligibility criteria, sample characteristics, inclusions and exclusions) on the sample to make a judgement Domain 2: Predictors • high - predictors not collected at study entry or at diagnosis: "Markers were tested within two years in 70% of cases, within 4 years in 15% and from 4 to 5 years from diagnosis in the remaining 15%." Domain 3: Outcome • low
Notes	Funding & conflict of interest

Morabito D - Morabito 2009 (Italian cohort) *(Continued)*

- "Supported in part from Associazione Italiana Ricerca sul Cancro (AIRC) (to FM, AN and MF), FIRB (Grant no RBIP06LCA9, to MF) MIUR, CIPE (2006–8, CBA project, to SZ), ISS (to SZ), Progetti Strategici – Ricerca Finalizzata Ministero Italiano della Salute 'RFPS_2006_3_33_99_60' (to GC) and 'RFPS_2006_340196' (to FM and MF); progetto ordinario ricerca finalizzata Ministero Italiano della salute-2007 (to GC), progetto Compagnia San Paolo (to GC), the Fondazione Internazionale Ricerche Medicina Sperimentale (FIRMS) provided financial and administrative assistance, Fondazione 'Amelia Scorza' onlus, Cosenza, Italy; and Associazione Italiana contro le Leucemie -Milano. KT and SM are supported by fellowships from the Fondazione Italiana Ricerca sul Cancro (FIRC)".
- Conflicts of interests not reported

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Only one statement on inclusion criteria. No statement on baseline sample and possible exclusions. Very little baseline information on participants
Domain 2: Predictors	Yes	Prospective collection of predictor information, centralised laboratory
Domain 3: Outcome	Yes	Standard outcome definition used: "Treatment was decided uniformly in all participating centres based on documented progressive and symptomatic disease according to NCI working guidelines".
Domain 4: Analysis	No	Not sufficient events; continuous data not handled appropriately (ROC to find the 'best' cut-point); no information on missing data; no performance measure reported
Overall judgement	No	

Morabito V - Gentile 2014 (O-CLL1-GISL)

Study characteristics	
General information	Model and type of study • Gentile 2014 (O-CLL1-GISL) • Validation study Secondary citations • not applicable Language of publication • English Study design • prospective cohort Follow-up time • median 38.5 months (range 6 - 82 months)
Participants	Number of included participants in the cohort

Morabito V - Gentile 2014 (O-CLL1-GISL) *(Continued)*

- 480 participants

Setting

- Italy, multicentre study

Recruitment period

- 2007 - not reported

Age (in years)

- 47.4 % of individuals were aged < 60 years

Sex

- 58.5% male

Stages of disease

- Rai 0: 77.3%; Rai I-II: 22.7%
- Binet A: 100%

Treatment

- not reported

Inclusion criteria

- the inclusion criteria for CLL diagnosis, employed at the time of study start, followed the NCI/WG guidelines established in 1996, required > 5.000 lymphocytes/mL in the peripheral blood.

Exclusion criteria

- exclusion criteria were the following: (i) CD5- and/or CD23- B-lymphoproliferative disorders; (ii) clinical Binet stage B or C, and (iii) need of therapy according to NCI/WG guidelines; (iv) age > 70 years.

Predictors	List of predictors in final model (including cut-points for dichotomised factors) • IgHV mutational status (continuous: categorical/binary transformation; 2% treshold, as established in literature); CD38 (continuous: categorical/binary transformation; 30% treshold); ZAP-70 (continuous: categorical/binary transformation; threshold established by ROC curve) • no difference between development and validation cohort predictors Timing of predictor measurement • at study entry, maximum 1 year after diagnosis
Outcome(s)	Primary outcome in study • progression-free survival: PFS has been defined as time to therapy requirement. Additional outcome(s) • not applicable Outcome in model development • time-to-first-treatment
Missing data	Participants with any missing data? • no If yes, how was missing data handled?

Morabito V - Gentile 2014 (O-CLL1-GISL) *(Continued)*

- not applicable

Analysis	Number of participants & number of events (specific time points where reported) • progression-free survival: 468 participants, 121 events Which model was used? • original model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: not reported Creation of risk groups? • yes, 3 risk groups, (cut-points at 1; 2-3; 4) • at 3 years • percentage of group without event at time point (low risk: 91.7%, intermediate risk: 82.9%, high risk: 57.4%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Associazione Italiana per la Ricerca sul Cancro; Contract grant no.: 9980, IG10492, IG10136" • "Nothing to report" Other comments • The O-CLL1-GISL cohort was used in various publications and to validate different CLL models. • The corresponding author provided additional information regarding the cohort.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospectively registered study (O-CLL1-GISL) with defined inclusion and exclusion criteria
Domain 2: Predictors	Yes	Prospective collection of predictor information. Predictors assessed study entry at a maximum of 1 year after diagnosis.
Domain 3: Outcome	Yes	CLL progression criteria 2008 were used. Outcome labelled differently than model development study. However, progression-free survival (PFS) has been defined as time-to-therapy requirement, which is equivalent to time-to-treatment in Morabito 2009.

Morabito V - Gentile 2014 (O-CLL1-GISL) *(Continued)*

Domain 4: Analysis	Unclear	No relevant performance measures reported (progression per risk group was reported)

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL)

Study characteristics

General information	Model and type of study • Gentile 2016 (O-CLL1-GISL) • Development study Secondary citations • Molica 2016; Molica 2017; Molica 2017 Language of publication • English Study design • prospective cohort Follow-up time • median 42 months (range 6 - 82 months)
Participants	Number of included participants in the cohort • 480 participants Setting • Italy, multicentre study Recruitment period • 2007 - not reported Age (in years) • 46.7% of individuals were younger than 60 years Sex • 57.3% male Stages of disease • Rai 0: 77.7%; Rai I-II: 22.3% • Binet A: 100% Treatment • 84 out of 337 received treatment • unclear which treatment Inclusion criteria • Established diagnosis of B-CLL by NCI criteria, performed by local haematologist (confirmed by the biological review committee according to flow cytometry analysis (positive clusters of differentiation

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) *(Continued)*

	antigen 5 (CD5), 19 (CD19), 23 (CD23)). Age more than 18 years and less than 70 years. Eastern Cooperative Oncology Group (ECOG) less than or equal to 2. Binet stage A. Diagnosis performed within 12 months before inclusion in the study. NCI watch-and-wait policy. Exclusion criteria • Patients with CLL whose diagnosis exceeded 12 months before registration. Patients with leukaemic phase of lymphoproliferative disorders of B origin CD5- and/or CD23- according to flow cytometry analysis. Clinical Binet stage B or C. Patients who necessitated therapy according to NCI guidelines. Age > 70 years. Without a written informed consent.
Predictors	Number of candidate predictors • univariable analysis: 10 • multivariable analysis: 7 List of predictors in final model (including cut-points for dichotomised factors) • Rai stage (categorical; 0 vs. I/II); ALC (continuous: categorical/binary transformation); beta2-microglobulin (continuous: categorical/binary transformation; normal/elevated (elevated above 2 mg/L)); IgHV status (continuous: categorical/binary transformation) Timing of predictor measurement • at study entry (< 12 months from diagnosis)
Outcome(s)	Primary outcome in study • time-to-first-treatment: TFS was calculated from diagnosis to the first CLL treatment (event) or to last follow-up (censoring) Additional outcome(s) • not applicable
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 337 participants, 84 events Predictor selection method • univariable analysis, unclear if stepwise approach was used in multivariable modelling (criterion: P value or CI) Statistical method • Cox proportional hazard model Simplification of model? • yes Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic)

O-CLL-1 D - Gentile 2016 (O-CLL-1-GISL) *(Continued)*

	Creation of risk groups? • yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7)) • at 3 years • percentage of group without event at time point (low risk: 95.3%, intermediate risk: 74.5%, high risk: 28.6%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN)." • "The authors declare no conflict of interest." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Prospective cohort, clear inclusion and exclusion criteria in trial registry
Domain 2: Predictors	Yes	Prospective and standardised collection of predictor information ("CLL cell phenotype, CD38 and ZAP-70 expression, and *IgHV* mutational status were performed in a central laboratory in Genova, while all FISH and genetic analyses were performed in Milan").
Domain 3: Outcome	No	Standard outcome definition, however, predictors were not excluded from outcome definition (e.g. Rai stage).
Domain 4: Analysis	No	Continuous data not handled appropriately ("Continuous variables of prognostic importance on TFS in univariate proportional hazards Cox regression were dichotomized using published thresholds and laboratory norms").
Overall judgement	No	

O-CLL-1 V - Rani 2018 (Indian cohort)

Study characteristics	
General information	Model and type of study • Rani 2018 (Indian cohort) • Validation study

O-CLL-1 V - Rani 2018 (Indian cohort) *(Continued)*

Secondary citations

- not applicable

Language of publication

- English

Study design

- unclear

Follow-up time

- median 40.5 months (range: 1 - 215 months)

Participants	Number of included participants in the cohort • 198 participants Setting • India, unclear if single or multicentre study Recruitment period • not reported Age (in years) • median 60 years (range not reported) Sex • 77% male Stages of disease • Rai 0: 14.6%; Rai I: 21.2%; Rai II: 34.9%; Rai III: 14.1%; Rai IV: 15.2% • Binet A/B/C: not reported Treatment • 62 received chlorambucil-based therapy • 56 received rituximab-based therapy • 20 received other therapies Inclusion criteria • treatment-naive CLL participants according to Hallek 2008 Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • gender (binary); age (continuous, dichotomisation after univariable analysis); ECOG PS (categorical, dichotomisation after univariable analysis); del(17p) (continuous: categorical/binary transformation); del(11q) (continuous: categorical/binary transformation; literature-based); IgHV mutational status (continuous: categorical/binary transformation; literature-based); s-TK (continuous, i after univariable analysis); s-beta2m (continuous, dichotomisation after univariable analysis) Timing of predictor measurement

O-CLL-1 V - Rani 2018 (Indian cohort) *(Continued)*

- not reported

Outcome(s)	Primary outcome in study • time-to-first-treatment: TTFT was defined as the time from the date of diagnosis to date of commencement of first therapy. Additional outcome(s) • not applicable Outcome in model development • time-to-first-treatment
Missing data	Participants with any missing data? • no If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 140 participants, 89 events Which model was used? • simplified model Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7)) • median survival (low risk: 3 years, intermediate risk: 24 years, high risk: 79 years)
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period unclear Domain 2: Predictors • unclear - no information on timing of predictor because the study design was unclear Domain 3: Outcome • low
Notes	Funding & conflict of interest • "The financial support was provided by the Department of Biotechnology (BT/PR11106/GBD/27/145/2008, BT/PR15438/MED/30/606/2011 and T/PR8680/AGR/36/754/2013), Ministry of Science and Technology, GOI, and All India Institute of Medical Sciences, New Delhi (8-60/A060/ 2011/RS) to RG for carrying out this work." • "The authors declare that they have no conflict of interest."

O-CLL-1 V - Rani 2018 (Indian cohort) *(Continued)*

Other comments

- Graphs were provided to obtain the percentage of participants untreated at a specific time point, however, these graphs were printed very small.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Inclusion and exclusion criteria unclear; it seemed to us that only individuals with complete information on all predictors were included; recruitment period unclear; study design unclear
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre)
Domain 3: Outcome	No	See model development: Predictors were not excluded from outcome definition (e.g. Rai stage).
Domain 4: Analysis	No	Low number of events; no information on handling of missing values
Overall judgement	No	

O-CLL1 V - Gentile 2016 (Mayo cohort)

Study characteristics	
General information	Model and type of study • Gentile 2016 (Mayo cohort 2001 - 2008) • Validation study Secondary citations • not applicable Language of publication • English Study design • not reported Follow-up time • median 97 months (range not reported)
Participants	Number of included participants in the cohort • 428 participants Setting • USA, single-centre study Recruitment period • 2001 - 2008

O-CLL1 V - Gentile 2016 (Mayo cohort) *(Continued)*

Age (in years)

- 40.2% <= 60 years

Sex

- 66.4% male

Stages of disease

- Rai 0: 67.5%; Rai I-II: 32.5%
- Binet A/B/C: not reported

Treatment

- 130 were treated
- unclear which treatment

Inclusion criteria

- not reported

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • Rai stage (categorical; 0 vs. I/II); ALC (continuous: categorical/binary transformation); beta2-microglobulin (continuous: categorical/binary transformation; normal/elevated (elevated above 2 mg/L)); IgHV status (continuous: categorical/binary transformation) Timing of predictor measurement • within 18 months of diagnosis
Outcome(s)	Primary outcome in study • time-to-first-treatment: TFS was calculated from diagnosis to the first CLL treatment (event) or to last follow-up (censoring). Additional outcome(s) • not applicable Outcome in model development • time-to-first-treatment
Missing data	Participants with any missing data? • not reported If yes, how was missing data handled? • not reported
Analysis	Number of participants & number of events (specific time points where reported) • time-to-first-treatment: 428 participants, 130 events Which model was used? • simplified model

O-CLL1 V - Gentile 2016 (Mayo cohort) *(Continued)*

	Was the model updated? • no Performance measures reported? • Calibration: yes (Hosmer-Lemeshow test) • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 3 risk groups, (cut-points at (0-2); (3-5); (6-7)) • at 3 years • percentage of group without event at time point (low risk: 95.5%, intermediate risk: 78.9%, high risk: 40.6%)
PROBAST: Applicability	Domain 1: Participant selection • unclear - recruitment period unclear Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Associazione Italiana per la Ricerca sul Cancro (AIRC 5xmille grant 9980, IG10492 to MF and FM and IG10136 to AN)." • "The authors declare no conflict of interest." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	No eligibility criteria; individuals with missing values were excluded.
Domain 2: Predictors	Yes	Not explicitly stated, but predictors probably assessed in a similar way (single-centre).
Domain 3: Outcome	No	See model development: Predictors not excluded from outcome definition (e.g. Rai stage)
Domain 4: Analysis	Unclear	Calibration not reported; individuals with missing values were probably excluded from analysis, rated in domain 1.
Overall judgement	No	

Rossi D - Rossi 2013 (Italian cohort)

Study characteristics

General information	Model and type of study • Rossi 2013 • Development study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 67.2 months (range not reported)
Participants	Number of included participants in the cohort • 637 participants Setting • Italy, multicentre study Recruitment period • 1996 - 2011 Age (in years) • 51.2% of individuals were older than 65 years. Sex • 58.4% male Stages of disease • Rai 0: 74.9%; Rai I-II: 11.5% • Binet A: not reported Treatment • 122 patients received rituximab-based regimens (i.e. fludarabine-cyclophosphamide-rituximab, fludarabinerituximab, or pentostatin-cyclophosphamide-rituximab) • 64 received fludarabine-based regimens (i.e. fludarabine or fludarabinecyclophosphamide) • 80 received alkylator-based regimens (i.e. chlorambucil) Inclusion criteria • 'newly diagnosed and previously untreated CLL patients of whom 583 (91.5%) were provided with regular follow-up' Exclusion criteria • not reported
Predictors	Number of candidate predictors

Rossi D - Rossi 2013 (Italian cohort) *(Continued)*

- univariable analysis: no univariable analysis
- multivariable analysis: 18

List of predictors in final model (including cut-points for dichotomised factors)

- TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary)

Timing of predictor measurement

- at diagnosis

Outcome(s)	Primary outcome in study • overall survival: OS was measured from the date of initial presentation to the date of death from any cause (event) or last follow-up (censoring). Additional outcome(s) • Treatment-free survival: TFS was measured from date of initial presentation to date of progressive disease requiring treatment according to IWCLL-NCI guidelines (event), death or last follow-up (censoring).
Missing data	Participants with any missing data? • yes (54 participants) If yes, how was missing data handled? • not reported
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 583 participants, 178 events Predictor selection method • full model, during multivariable modelling: hierarchical clustering Statistical method • recursive partitioning Simplification of model? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at very low risk: del13q14 sole lesion; low risk: +12 or all lesions; intermediate: NOTCH1 and/or SF3B1; high risk: TP53 and/or BIRC3) • at 5 years • percentage of group without event at time point (low risk: 86.9%, intermediate risk: 77.6%, high risk: 65.9%, very high risk: 50.9%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors

Rossi D - Rossi 2013 (Italian cohort) *(Continued)*

- low

Domain 3: Outcome

- low

Notes

Funding & conflict of interest

- "This study was supported by the Associazione Italiana per la Ricerca sul Cancro Foundation, Special Program Molecular Clinical Oncology, 5 1000, number 10007, Milan, Italy (to G.G. and to R.F.); Progetto Futuro in Ricerca 2008 (to D.R.); Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (to G.G.and R.M.); PRIN 2009 (to D.R.); Progetto Futuro in Ricerca 2012 (to D.R.); Ministero dell'Istruzione, dell'Universita` e della Ricerca, Rome, Italy; Progetto Giovani Ricercatori 2008 (to D.R.); Progetto Giovani Ricercatori 2010 (to D.R.); Ricerca Sanitaria Finalizzata 2008 (to G.G.); Ministero della Salute, Rome, Italy; Novara-AIL Onlus Foundation, Novara, Italy (to G.G. and D.R.); Compagnia di San Paolo, Turin, Italy (to R.F.); Helmut Horten Foundation and San Salvatore Foundation, Lugano, Switzerland (to F.B.); Nelia et Amadeo Barletta Foundation, Lausanne (to F.B.); National Institutes of Health grant PO1-CA092625 (to R.D.-F.); and a Specialized Center of Research grant from the Leukemia and Lymphoma Society (to R.D.-F.). S.M. and S.C. are supported by fellowships from the Novara-AIL Onlus Foundation, Novara, Italy. L.P. is on leave from the University of Perugia Medical School."
- "The authors declare no competing financial interests."

Other comments

- none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Yes	Appropriate data sources used: "multicentric cohort of 637 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation at four institutions from June 1996 through June 2011"
Domain 2: Predictors	Yes	Detailed description of predictors and timing of measurements: "clinical information prospectively collected at clinically relevant time points (i.e. at diagnosis, progression, and last follow-up)"
Domain 3: Outcome	Yes	Objective standard outcome; regular clinical database update
Domain 4: Analysis	Unclear	Unclear if individuals with missing data were excluded; number of events sufficient; building of model by recursive partitioning (decision tree, tenfold cross-validation), testing for stability with random forest algorithm and Cox model with bootstrapping; correction of c-statistic; assumptions were checked

Rossi V - Jeromin 2014 (Munich cohort)

Study characteristics

General information

Model and type of study

- Jeromin 2014 (MLL Munich cohort)
- Validation study

Secondary citations

- not applicable

Language of publication

Rossi V - Jeromin 2014 (Munich cohort) *(Continued)*

	• English Study design • prospective cohort Follow-up time • median 55.2 months (range not reported)
Participants	Number of included participants in the cohort • 1160 participants Setting • probably Germany, unclear Recruitment period • 2005 - 2010 Age (in years) • median 67 years (range: 29.6 - 90.5 years) Sex • 64.6% male Stages of disease • not reported Treatment • Not reported Inclusion criteria • not reported Exclusion criteria • prior treatment
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary) • BIRC3, which is part of the original model, was not analysed in this study. Timing of predictor measurement • at diagnosis for most participants (82.6%)
Outcome(s)	Primary outcome in study • overall survival: OS was measured from the date of diagnosis until last follow-up or death. Additional outcome(s) • time to treatment: TTT was evaluated from the date of diagnosis until the date of initial treatment. Outcome in model development

Rossi V - Jeromin 2014 (Munich cohort) *(Continued)*

	• overall survival
Missing data	Participants with any missing data? • yes (160 participants) If yes, how was missing data handled? • complete-case analysis
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 930 participants, events not reported Which model was used? • original model as developed in Rossi 2013 Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups (cut-points not reported) • at 5 years • percentage of group without event at time point (low risk: 91%, intermediate risk: 90%, high risk: 75.2%, very high risk: 62.1%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Next-generation deep sequencing studies were supported in part by the IRON-II study framework (Roche Diagnostics, Penzberg, Germany)." • "SS, WK, CH, and TH are part owners of the MLL Munich Leukemia Laboratory. SJ, SW, VG, KB, FD, TA, AR and AK are employed by the MLL Munich Leukemia Laboratory." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Study design unclear

Rossi V - Jeromin 2014 (Munich cohort) *(Continued)*

Domain 2: Predictors	Yes	Relatively short recruitment period (2005 - 2010); samples were sent to a laboratory which was not involved in treatment.
Domain 3: Outcome	Yes	Objective standard outcome; no information on outcome assessment. However, we assumed that for OS, risk of bias was low.
Domain 4: Analysis	No	Individuals with missing data excluded from analysis (complete data for OS in 935/1160 cases); no performance measures, just risk groups
Overall judgement	No	

Rossi V - Rossi 2013 (unclear)

Study characteristics	
General information	Model and type of study • Rossi 2013 (unclear) • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • median 70.8 months, (range not reported)
Participants	Number of included participants in the cohort • not reported Setting • Italy, multicentre study Recruitment period • 1996 - 2011 Age (in years) • not reported Sex • 55.9% male Stages of disease • Rai 0-I: 75.4%; Rai II: 17.3%; Rai III-IV: 7.3% Treatment

Rossi V - Rossi 2013 (unclear) *(Continued)*

- not reported

Inclusion criteria

- not reported

Exclusion criteria

- not reported

Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • TP53 (binary); BIRC3 DIS (binary); SF3B1 M (binary); NOTCH1 M (binary); del11q22-q23 (binary) • same predictors Timing of predictor measurement • at diagnosis
Outcome(s)	Primary outcome in study • overall survival Additional outcome(s) • not applicable Outcome in model development • overall survival
Missing data	Participants with any missing data? • unclear If yes, how was missing data handled? • not applicable
Analysis	Number of participants & number of events (specific time points where reported) • overall survival: 370 participants, 62 events Which model was used? • recursive partitioning Was the model updated? • yes • What has been updated? Excel sheet Performance measures reported? • Calibration: reported (observed-expected ratio) • Discrimination: reported (c-statistic) Creation of risk groups? • yes, 4 risk groups, (cut-points at very low risk: del13q14 sole lesion; low risk: +12 or all lesions; intermediate: NOTCH1 and/or SF3B1; high risk: TP53 and/or BIRC3) • at 5 years

Rossi V - Rossi 2013 (unclear) *(Continued)*

	• percentage of group without event at time point (low risk: 86.9%, intermediate risk: 77.6%, high risk: 65.9%, very high risk: 50.9%)
PROBAST: Applicability	Domain 1: Participant selection • low Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "This study was supported by the Associazione Italiana per la Ricerca sul Cancro Foundation, Special Program Molecular Clinical Oncology, 5 1000, number 10007, Milan, Italy (to G.G. and to R.F.); Progetto Futuro in Ricerca 2008 (to D.R.); Programmi di Ricerca di Rilevante Interesse Nazionale (PRIN) 2008 (to G.G.and R.M.); PRIN 2009 (to D.R.); Progetto Futuro in Ricerca 2012 (to D.R.); Ministero dell'Istruzione, dell'Università e della Ricerca, Rome, Italy; Progetto Giovani Ricercatori 2008 (to D.R.); Progetto Giovani Ricercatori 2010 (to D.R.); Ricerca Sanitaria Finalizzata 2008 (to G.G.); Ministero della Salute, Rome, Italy; Novara-AIL Onlus Foundation, Novara, Italy (to G.G. and D.R.); Compagnia di San Paolo, Turin, Italy (to R.F.); Helmut Horten Foundation and San Salvatore Foundation, Lugano, Switzerland (to F.B.); Nelia et Amadeo Barletta Foundation, Lausanne (to F.B.); National Institutes of Health grant PO1-CA092625 (to R.D.-F.); and a Specialized Center of Research grant from the Leukemia and Lymphoma Society (to R.D.-F.). S.M. and S.C. are supported by fellowships from the Novara-AIL Onlus Foundation, Novara, Italy. L.P. is on leave from the University of Perugia Medical School." • "The authors declare no competing financial interests." Other comments • difference between external validation cohort and development cohort unclear

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	Unclear	Study design not clear: "The validation series was represented by a cohort of 370 newly diagnosed and previously untreated CLL patients who consecutively presented for initial evaluation from June 1996 through June 2011 (Table S1) and provided with regular follow-up (at least three visits/year)."
Domain 2: Predictors	Yes	No detailed description of predictor assessment; we assumed that it was similar to the development cohort described in the same publication.
Domain 3: Outcome	Yes	Objective standard outcome, regular clinical database update
Domain 4: Analysis	No	Small number of events; no information on missing values; calibration not reported
Overall judgement	No	

Stephens OS D - Stephens 2015 (Ohio cohort)

Study characteristics

Stephens OS D - Stephens 2015 (Ohio cohort) *(Continued)*

General information	Model and type of study • Stephens 2015 • Development study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time • not reported
Participants	Number of included participants in the cohort • 114 participants Setting • USA, single-centre study (Ohio State University) Recruitment period • 2002 - 2012 Age (in years) • median 62 years (range: 40 - 92 years) Sex • not reported Stages of disease • Rai 0: 33%; Rai I-II: 46%; Rai III-IV: 21% Treatment • 35 received purine analogue-based therapy • 18 received treatment on a clinical trial • 5 received other treatment Inclusion criteria • patients with CLL with del(17p13.1), previously untreated Exclusion criteria • not reported
Predictors	Number of candidate predictors • univariable analysis: no univariable analysis • multivariable analysis: 18 List of predictors in final model (including cut-points for dichotomised factors)

Stephens OS D - Stephens 2015 (Ohio cohort) *(Continued)*

- for OS: ECOG Performance Status (binary); age (binary); lactate dehydrogenase (LDH; binary)

Timing of predictor measurement

- at presentation, median 4.8 months after diagnosis (0 days to 19.7 years)

Outcome(s)	Primary outcome in study • overall survival: OS was calculated from date of first visit until date of death or last follow-up. Additional outcome(s) • treatment-free survival: TFS was calculated from date of first visit until date of first treatment or death, censoring patients alive and treatment-free at last follow-up.
Missing data	Participants with any missing data? • yes If yes, how was missing data handled? • multiple imputation technique
Analysis	Number of participants & number of events (specific time points where reported) • OS: 114 participants, events not reported • TFS: 114 participants, 58 events Predictor selection method • univariable analysis, during multivariable modelling: backwards selection Statistical method • proportional hazards model Simplification of model? • yes • risk score (RS) based on the variables and regression coefficients of the multivariable model • simplified risk score (SRS) to be used in clinical practice based on the strength of associations with clinical outcome when all variables had been categorised Performance measures reported? • Calibration: not reported • Discrimination: reported for OS (c-statistic), but not reported for TFS (c-statistic for multivariable model only) Creation of risk groups? • for OS: yes, 3 risk groups (SRS = 0; 2; 4) • at 2 years
PROBAST: Applicability	Domain 1: Participant selection • high - selective sample of very high risk individuals with a del(17p) deletion Domain 2: Predictors • low Domain 3: Outcome

Stephens OS D - Stephens 2015 (Ohio cohort) *(Continued)*

- low

Notes	Funding & conflict of interest • "Contract grant sponsors: Four Winds Foundation, D. Warren Brown Foundation; Contract grant sponsors: Mr. and Mrs. Michael Thomas, Harry Mangurian Foundation; Contract grant no.: P50 CA140158; Contract grant sponsor: Leukemia and Lymphoma Society". • "The authors have no competing interests to disclose." Other comments • The authors reported a multivariable model, a risk score (RS) and a simplified risk score (SRS) for each outcome. Since the former two were not validated externally, only data for the SRS were extracted.

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Yes	Mostly laboratory data, probably assessed in a similar way: "Stimulated cytogenetic and fluorescent in situ hybridisation (FISH) analyses were performed on peripheral blood or bone marrow samples, as previously described"; single-centre study
Domain 3: Outcome	Yes	Objective standard outcome: although median observation time was not reported, median OS estimates were 5.2 years, therefore, we assumed a median observation time exceeding 5 years.
Domain 4: Analysis	No	Number of events only reported for TFS, not OS - the number of persons included was too small, total n = 114; missing data handling appropriate; inconsistent reporting of c-statistics; calibration was not reported
Overall judgement	No	

Stephens OS V - Stephens 2015 (MDACC)

Study characteristics	
General information	Model and type of study • Stephens 2015 • Validation study Secondary citations • not applicable Language of publication • English Study design • retrospective cohort Follow-up time

Stephens OS V - Stephens 2015 (MDACC) *(Continued)*

- not reported

Participants	Number of included participants in the cohort • 129 participants Setting • USA, single-centre study (MDACC) Recruitment period • not reported Age (in years) • median 63 years (range: 40 - 85 years) Sex • not reported Stages of disease • Rai 0: 29%; Rai I-II: 52%; Rai III-IV: 19% Treatment • not reported Inclusion criteria • patients with CLL with del(17p13.1), previously untreated Exclusion criteria • not reported
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • for OS: ECOG Performance Status (binary); age (binary); lactate dehydrogenase (LDH; binary) • same predictors Timing of predictor measurement • not reported
Outcome(s)	Primary outcome in study • overall survival Additional outcome(s) • treatment-free survival Outcome in model development • overall survival • treatment-free survival
Missing data	Participants with any missing data? • not reported If yes, how was missing data handled?

Stephens OS V - Stephens 2015 (MDACC) *(Continued)*

- not applicable

Analysis	Number of participants & number of events (specific time points where reported) • OS: 129 participants, events not reported • TFS: 129 participants, events not reported Which model was used? • simplified risk score (SRS) Was the model updated? • no Performance measures reported? • Calibration: not reported • Discrimination: reported for OS and TFS (c-statistic) Creation of risk groups? • for OS: yes, 3 risk groups (SRS = 0; 2; 3) • at 2 years
PROBAST: Applicability	Domain 1: Participant selection • high - selective sample of very high risk individuals with a del(17p) deletion Domain 2: Predictors • low Domain 3: Outcome • low
Notes	Funding & conflict of interest • "Contract grant sponsors: Four Winds Foundation, D. Warren Brown Foundation; Contract grant sponsors: Mr. and Mrs. Michael Thomas, Harry Mangurian Foundation; Contract grant no.: P50 CA140158; Contract grant sponsor: Leukemia and Lymphoma Society". • "The authors have no competing interests to disclose." Other comments • none

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Unclear	lack of information about similarity of predictor assessment to the development study; no recruitment period
Domain 3: Outcome	Yes	Objective standard outcome: although median observation time was not reported, median OS estimates were 6.4 years, therefore we assumed a median observation time exceeding 5 years.

Stephens OS V - Stephens 2015 (MDACC) *(Continued)*

Domain 4: Analysis	No	Number of included persons too small, total N = 129; calibration not reported
Overall judgement	No	

Stephens TFS D - Stephens 2015 (Ohio cohort)

Study characteristics	
General information	Model and type of study • Stephens 2015 • Development study Different outcome but identical publication as Stephens OS D - Stephens 2015 (Ohio cohort) and Stephens OS V - Stephens 2015 (MDACC). Therefore, only relevant differences are reported in this table. For additional information, see tables above.
Participants	see Stephens OS D - Stephens 2015 (Ohio cohort)
Predictors	List of predictors in final model (including cut-points for dichotomised factors) • for TFS: ECOG Performance Status (binary); Rai stage (binary); age (binary); white blood cell (WBC; binary); del(11q22.3) (binary)
Outcome(s)	Primary outcome in study • overall survival: OS was calculated from date of first visit until date of death or last follow-up. Additional outcome(s) • treatment-free survival: TFS was calculated from date of first visit until date of first treatment or death, censoring patients alive and treatment-free at last follow-up.
Missing data	
Analysis	Number of participants & number of events (specific time points where reported) • OS: 114 participants, events not reported • TFS: 114 participants, 58 events Performance measures reported? • Calibration: not reported • Discrimination: reported for OS (c-statistic), but not reported for TFS (c-statistic for multivariable model only) Creation of risk groups? • for TFS: yes, 3 risk groups (SRS = 0/1; 2/3; >= 4) • at 2 years
PROBAST: Applicability	
Notes	

Stephens TFS D - Stephens 2015 (Ohio cohort) *(Continued)*

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Yes	Mostly laboratory data, probably assessed in a similar way: "Stimulated cytogenetic and fluorescent in situ hybridization (FISH) analyses were performed on peripheral blood or bone marrow samples, as previously described"; single-centre study
Domain 3: Outcome	No	Predictors were not excluded from the outcome definition (Rai stage; Hallek 2008)
Domain 4: Analysis	No	Number of events only reported for TFS, not OS - the number of persons included was too small, total N = 114; missing data handling appropriate; inconsistent reporting of c-statistics; calibration was not reported
Overall judgement	No	

Stephens TFS V - Stephens 2015 (MDACC)

Study characteristics	
General information	Model and type of study • Stephens 2015 • Validation study Different outcome but identical publication as Stephens OS D - Stephens 2015 (Ohio cohort) and Stephens OS V - Stephens 2015 (MDACC). Therefore, only relevant differences are reported in this table. For additional information, see tables above.
Participants	
Predictors	List of predictors used for validation (and changes between original predictors and predictors in validation study) • for TFS: ECOG Performance Status (binary); Rai stage (binary); age (binary); white blood cell (WBC; binary); del(11q22.3) (binary) • same predictors
Outcome(s)	Primary outcome in study • overall survival Additional outcome(s) • treatment-free survival Outcome in model development • overall survival • treatment-free survival
Missing data	

Stephens TFS V - Stephens 2015 (MDACC) *(Continued)*

Analysis	Number of participants & number of events (specific time points where reported) • OS: 129 participants, events not reported • TFS: 129 participants, events not reported Performance measures reported? • Calibration: not reported • Discrimination: reported for OS and TFS (c-statistic) Creation of risk groups? • for TFS: yes, 3 risk groups (SRS = 0/1; 2/3; >= 4) • at 2 years
PROBAST: Applicability	
Notes	

Item	Authors' judgement	Support for judgement
Domain 1: Participant selection	No	Possibly selective inclusion of individuals, as only patients with del(17p13.1) were considered. Individuals without assessment could not have been included. Other eligibility criteria were unclear.
Domain 2: Predictors	Unclear	Lack of information about similarity of predictor assessment to the development study, no recruitment period
Domain 3: Outcome	No	Predictors were not excluded from the outcome definition (Rai stage; Hallek 2008)
Domain 4: Analysis	No	Number of included persons too small, total N = 129; calibration not reported
Overall judgement	No	

AIHA: autoimmune haemolytic anaemia
ALC: absolute lymphocyte count
AUC: area under the curve
B2M: beta2-microglobulin
BcR IG: clonotypic B-cell receptor immunoglobulin
BIRC3: baculoviral IAP repeat containing 3
BR: bendamustine-rituximab
C: cyclophosphamide
CHL: chlorambucil
CHOP: cyclophosphamide, doxorubicin, vincristine, and prednisone
CI: confidence interval
CIRS: Cumulative Illness Rating Score
CLL: chronic lymphocytic leukaemia
CLL-IPI: Chronic Lymphocytic Leukaemia-International Prognostic Index
cMBL: clinical monoclonal B lymphocytosis
CVP: central venous pressure
DE: Germany
del11q: FISH-detected genetic abberation del11q
del17p: FISH-detected genetic abberation del17p
ECOG PS: Eastern Cooperative Oncology Group performance status
F: fludarabine
FC: fludarabine and cyclophosphamide

FCR: fludarabine, cyclophosphamide and rituximab
FFP: fresh frozen plasma
FR: France
FISH: fluorescence in situ hybridisation
HIV: human immunodeficiency virus
HBV: hepatitis B virus
HCV: hepatitis C virus
HDMP: high-dose methylprednisolone
IgHV: immunoglobulin heavy chain variable region genes
IGLV: immunoglobulin lambda variable cluster
IQR: Interquartile range
IPD: individual patient data
iwCLL: International workshop on Chronic Lymphocytic Leukemia
LDH: lactate dehydrogenase
LDT: lymphocyte doubling time
LN: lymph node
LYM: lymphocyte count
MCAR: missing completely at random
M-CLL: mutated chronic lymphocytic leukaemia
MDACC: MD Anderson Cancer Center
MLL: Munich Leukemia Laboratory
MoAbs: monoclonal antibodies
MV analyses: multivariable analysis
NA: not available
NCI: National Cancer Institute
NOTCH1: notch receptor 1
O-Benda: ofatumumab bendamustine
O:E: observed-expected ratio
OS: overall survival
PFS: progression-free survival
PL: Poland
PROBAST: Prediction model Risk Of Bias ASsessment Tool
R: rituximab
RCT: randomised controlled trial
ROC: receiver operating characteristics curve
RR: risk ratio
RS: risk score
RTK: tyrosine kinase inhibitors
s-B2M: serum beta-2 microglobulin
SCALE: Scandinavian population-based study
SCAN: Scandinavian population-based case-control study
SF3B1: Splicing Factor 3b Subunit 1
SRS: simplified risk score
s-TK: serum thymidine kinase
TFS: treatment-free survival
TP53: tumour protein P53
TP: treatment probability
TTE: Time-to-Evenr
TTFT: time-to-first-treatment
U-CLL: unmutated chronic lymphocytic leukaemia
WBC: white blood cell count
WHO: World Health Organization

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Apelgren 2006	Study focus on staging system(s)
Baccarani 1982	Study focus on staging system(s)

Study	Reason for exclusion
Baliakas 2015	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Berke 2019	Prognostic factor study
Bettini 1986	Outdated staging system, which is no longer used in today's clinical practice
Binet 1977	Study focus on staging system(s)
Binet 1981	Study focus on staging system(s)
Bo 2014	Prognostic factor study
Bomben 2005	Comment or addition referring to an excluded study
Bomben 2009	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Bou Samra 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Brejcha 2010	Prognostic factor study
Brugiatelli 2007	Prognostic factor study
Bulian 2014	Prognostic factor study
Byrd 2006	Prognostic factor study
Cailliod 2005	Prognostic factor study
Callea 1999	Prognostic factor study
Catovsky 1989	Prognostic factor study
Cesano 2013	Prognostic factor study
Chang 2003	Prognostic factor study
Chastang 1985	Outdated staging system, which is no longer used in today's clinical practice
Chauzeix 2018	Prognostic factor study
Chelazzi 1979	Study focus on staging system(s)
Chen 1997	Prognostic factor study
Chena 2008	Prognostic factor study
Chevallier 2002	Prognostic factor study
Chiaretti 2014	Prognostic factor study
Christiansen 1994	Prognostic factor study
Chuang 2012	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Ciccone 2012	Prognostic factor study

Study	Reason for exclusion
Ciocoiu 1988	Study focus on staging system(s)
Claus 2012	Prognostic factor study
Claus 2014	Prognostic factor study
Cmunt 2002	Prognostic factor study
Cocco 2005	Prognostic factor study
Corcoran 2005	Prognostic factor study
Cordone 1998	Prognostic factor study
Cortese 2014	Prognostic factor study
Coscia 2012	Prognostic factor study
Crespo 2003	Prognostic factor study
Criel 1997	Characterisation of CLL and prognostic factors
Cro 2009	Prognostic factor study
Cro 2010	Characterisation of CLL and prognostic factors
Cuneo 2004	Characterisation of CLL and prognostic factors
Cuneo 2018	Observational efficacy study
D'Arena 2001	Prognostic factor study
D'Arena 2007	Prognostic factor study
D'Arena 2012	Characterisation of CLL and prognostic factors
Damle 1999	Prognostic factor study
Dasgupta 2015	Study focus on identification of the threshold for a prognostic factor
Davis 2016	Study focus on identification of the threshold for a prognostic factor
De Faria 2000	Study focus on staging system(s)
De Rossi 1989	Study focus on staging system(s)
DeAndres-Galiana 2016	Study focus on identification of a prognostic factor
Degan 2004	Prognostic factor study
Del Guidice 2011	Prognostic factor study
Del Poeta 2010	Prognostic factor study
Del Principe 2004	Prognostic factor study

Study	Reason for exclusion
Del Principe 2006	Prognostic factor study
Delgado 2009	Prognostic factor study
Delgado 2014	Prognostic factor study
Deslandes 2007	Methodological study
Di Raimondo 2001	Prognostic factor study
Dimier 2018	Prediction model to evaluate the relationship between surrogacy outcome(s) and standard outcomes
Dong 2011	Prognostic factor study
Dong 2014	Prognostic factor study
Durak 2009	Prognostic factor study
El-Kinawy 2012	Prognostic factor study
Fang 2019	Study applies the CLL-IPI, but without the aim of validating it
Ferrara 1981	Study focus on staging system(s)
Ferreira 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
French Cooperative Group on CLL 1988	Outdated staging system, which is no longer used in today's clinical practice
Friedman 2009	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Gattei 2008	Prognostic factor study
Gdynia 2018	Prognostic factor study
Geisler 1997	Characterisation of CLL and prognostic factors
Gentile 2016	Prognostic factor study
Gentile 2018	Score or model included predictors available only at treatment initiation
Giles 2003	Patient population did not match the review question; study included previously treated CLL patients
Giudice 2018	Prognostic factor study
Gogia 2014	Prognostic factor study
Gonzalez 2013	Characterisation of CLL and prognostic factors
Gonzalez-Gascon 2015	Characterisation of CLL and prognostic factors
Gonzalez-Rodriguez 2010	Characterisation of CLL and prognostic factors

Study	Reason for exclusion
Grabowski 2005	Prognostic factor study
Grever 2006	Abstract only
Hallek 1999	Characterisation of CLL and prognostic factors
Han 1984	Prognostic factor study
Herold 2011	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Hock 2010	Prognostic factor study
Houldsworth 2014	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Hus 2006	Prognostic factor study
Jaksic 1981	Prognostic factor study
Jaksic 1992	Study focus on staging system(s)
Jaksic 2014	Comment or addition referring to an excluded study
Josefsson 2007	Prognostic factor study
Juliusson 1986	Prognostic factor study
Juliusson 1990	Prognostic factor study
Kahraman 2014	Prognostic factor study
Kardum-Skelin 2008	Prognostic factor study
Kardum-Skelin 2009	Patient population did not match the review question; study included previously treated CLL patients
Karmiris 1994	Prognostic factor study
Kay 2018	Study applied the CLL-IPI, but without the aim of validating it
Keating 2000	Patient population did not match the review question; study included previously treated CLL patients
Khalifa 2002	Prognostic factor study
Kim 2004	Prognostic factor study
Kimby 1988	Prognostic factor study
Kleinstern 2018	Risk factor study
Knospe 1977	Prognostic factor study
Koberda 1989	Prognostic factor study
Korycka-Wolowiec 2011	Prognostic factor study

Study	Reason for exclusion
Krober 2002	Prognostic factor study
Krober 2006	Patient population did not match the review question; study included previously treated CLL patients
Kryachok 2011	Prognostic factor study
Kurec 1992	Prognostic factor study
Lai 2002	Prognostic factor study
Lech-Maranda 2012	Prognostic factor study
Lech-Maranda 2013	Prognostic factor study
Lecouvet 1997	Prognostic factor study
Li 2008	Prognostic factor study
Li 2017a	Prognostic factor study
Lin 2002	Prognostic factor study
Lin 2014	Prognostic factor study
Lozano-Santos 2014	Prognostic factor study
Lucas 2015	Prognostic factor study
Maffei 2007	Prognostic factor study
Maffei 2010	Prognostic factor study
Mandelli 1987	Development study of an outdated staging system, which is no longer used in today's clinical practice
Mansouri 2013	Prognostic factor study
Marasca 2005	Prognostic factor study
Marasca 2013	Prognostic factor study
Martinelli 2008	Prognostic factor study
Masic 1998	Prognostic factor identification study
Mateva 2001	Prognostic factor model, no full-text for further investigation
Matthews 2006	Prognostic factor study
Matthews 2007	Prognostic factor study
Matutes 2013	Prognostic factor study
Matutes 2017	Review

Study	Reason for exclusion
Melo 1987	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Miao 2018	Prognostic factor study
Molica 1984	Study focus on staging system(s)
Molica 1986	Prognostic factor study
Molica 1988	Prognostic factor study
Molica 1991	Prognostic factor study
Molica 1994	Prognostic factor study
Molica 1998	Prognostic factor study
Molica 1999a	Prognostic factor study
Molica 1999b	Prognostic factor study
Molica 2008	Prognostic factor study
Montserrat 1991	Prognostic factor study
Morabito 2001	Prognostic factor study
Morabito 2015a	Prognostic factor study
Morabito 2015b	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Morabito 2018c	Study aimed to identify cut-off points for a specific prognostic factor
Moreno 2019	Study focus on staging system(s)
Morilla 2008	Prognostic factor study
Nabhan 2017	Score or model included predictors available only at treatment initiation
Nedeva 2018	Evaluation of an excluded model
Nenova 2000	Prognostic factor study
Nipp 2014	Prognostic factor study
Nola 2004	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Nowakowski 2009	Prognostic factor study
Nuckel 2006	Prognostic factor study
Nuckel 2009	Prognostic factor study

Study	Reason for exclusion
O'Brien 1993	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Ocana 2007	Prognostic factor study
Oliveira 2011	Prognostic factor study
Orgueira 2019	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Oscier 1990	Prognostic factor study
Paolino 1984	Prognostic factor study
Plesingerova 2017	Study focus on surrogacy outcome(s)
Prokocimer 1985	Prognostic factor study
Qin 2017	Prognostic factor study
Queiros 2015	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Rai 1975	Study focus on staging system(s)
Rai 1990	Study focus on staging system(s)
Raponi 2018	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Resegotti 1989	Prognostic factor study
Rissiek 2014	Prognostic factor study
Rodriguez 2007	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Ronchetti 2016	Prognostic factor study
Rossi 1986	Study focus on staging system(s)
Rossi 2010a	Study focus on identification of the threshold for a prognostic factor
Rossi 2010b	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Rossi 2011	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Rozman 1979	Study focus on staging system(s)
Salomon-Nguyen 1995	Diagnostic score
Santoro 1979	Study focus on staging system(s)
Sarmiento 2002	Prognostic factor study
Savvopoulos 2016	Methodological study

Study	Reason for exclusion
Scolozzi 1981	Study focus on staging system(s)
Shanafelt 2010	Prognostic factor study
Shanafelt 2017	Outcome not relevant for the review
Spacek 2009	Prognostic factor study
Stamatopoulos 2009	Prognostic score, however, this score was built with a proxy as outcome ('distinguished mutated and unmutated cases')
Stamatopoulos 2017	Prognostic factor study
Strefford 2015	Prognostic factor study
Szymczyk 2018	Prognostic factor study
Tallarico 2018	Outcome not relevant for the review
Tobin 2005a	Study focus on identification of the threshold for a prognostic factor
Tobin 2005b	Study focus on identification of the threshold for a prognostic factor
Vallat 2013	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Van Damme 2012	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering
Velardi 1980	Study focus on staging system(s)
Vojdeman 2017	Prognostic factor study
Vural 2014	Prognostic factor study
Weinberg 2007	Patient population did not match the review question; study was not restricted to previously untreated CLL patients
Weiss 2011	Prognostic factor study
Wierda 2003	Prognostic factor study
Winkler 2010	Prognostic factor study
Wu 2010	Study focus on staging system(s)
Zengin 1997	Study focus on staging system(s)
Zenz 2009	Prognostic factor study
Zucchetto 2006	Genetic analysis, e.g. genetic subgrouping, genetic signature(s) or genetic clustering

CLL: chronic lymphocytic leukaemia
CLL-IPI: chronic lymphocytic leukaemia International Prognostic Index

Characteristics of studies awaiting classification *[ordered by study ID]*

Antic 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Baumann 2014

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Bulian 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Bulian 2012

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Cavallini 2017

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Del Guidice 2005

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Friedrichs 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Furundarena 1994

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Gentile 2009

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Haferlach 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Jarque 1991

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Lee 1987

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Leotard 2000

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Letestu 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Li 2017b

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Liang 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Metze 2000

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Miao 2019

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Molica 1990

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Molica 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Molica 2015

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Molica 2019

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Morabito 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Pepper 2012

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Qin 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Rozman 1982

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Rozman 1984

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Schweighofer 2011

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Stamatopoulos 2010

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Tsimberidou 2007

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Vetro 2018

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Visentin 2015

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Wierda 2009

Notes	Details on study design, recruitment period, prognostic factors and more can be found in Appendix 7.

Characteristics of ongoing studies *[ordered by study ID]*

NCT00275054

Study name	Rituximab, fludarabine, and cyclophosphamide or observation alone in treating patients with stage 0, stage I, or stage II chronic lymphocytic leukemia
Starting date	2005
Contact information	Michael Hallek, MD, Medizinische Universitaetsklinik I at the University of Cologne
Notes	• inclusion reason: "Development of a new prognostic staging system" • no staging system or prognostic model published yet

NCT03436524

Study name	A prognostic tool for early stage CLL
Starting date	2018
Contact information	Davide Rossi, MD, PhD, Principal investigator, Oncology Institute of Southern Switzerland
Notes	• inclusion reason: "The study aims at developing a model for the prediction of time to first treatment in chronic lymphocytic leukemia patients presenting with asymptomatic early stage disease". • study is currently recruiting

APPENDICES

Appendix 1. MEDLINE (via Ovid) search strategy up to 19 September 2018

1	exp LEUKEMIA, LYMPHOCYTIC, CHRONIC, B-CELL/
2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em$ adj3 (chronic$ or cronic$ or chroniq$ or well-differentia$)).tw,kf,ot.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic$ or cronic$ or chroniq$ or well-differentia$)).tw,kf,ot.
4	((small cell$ or small-cell$) adj3 lymphom$).tw,kf,ot.
5	(lymphom$ adj2 lymphocyt$).tw,kf,ot.
6	lymphoplasma?ytoid.tw,kf,ot.

(Continued)

7	(cll or b-cll or bcll).tw.
8	sll.tw.
9	or/1-8
10	(predict$ or clinical$ or outcome$ or risk$).mp.
11	validat$.mp. or predict$.ti. or rule$.mp.
12	(predict$ and (outcome$ or risk$ or models$)).mp.
13	((history or variable$ or criteria or scor$ or characteristic$ or finding$ or factor$) and (predict$ or model$ or decision$ or identif$ or prognos$)).mp.
14	decision$.mp. and ((model$ or clinical$).mp. or LOGISTIC MODELS/)
15	(prognostic and (history or variable$ or criteria$ or scor$ or characteristic$ or finding$ or factor$ or model$)).mp.
16	or/10-15
17	validat$.mp. or predict$.ti. or DECISION SUPPORT TECHNIQUES/ or rule$.mp. or PREDICTIVE VALUE OF TESTS/
18	(predict$ and (clinical$ or identif$)).mp.
19	17 or 18
20	RISK ASSESSMENT/
21	(risk$ adj scores$).tw,kf,ot.
22	exp RISK FACTORS/
23	(risk$ adj (score$ and factor$)).tw,kf,ot.
24	DECISION SUPPORT TECHNIQUES/
25	(decision$ adj2 (techniqu$ or model$)).tw,kf,ot.
26	(decision$ and support$ and technique$).tw,kf,ot.
27	(prediction$ and rule$ and clinical$).tw,kf,ot.
28	(decision$ adj2 (modeling$ or aid$ or analys$ or technique$)).tw,kf,ot.
29	or/20-28
30	ANIMALS/ not HUMANS/
31	9 and (16 or 19 or 29)
32	31 not 30

Appendix 2. MEDLINE (via Ovid) search strategy, 19 September 2018 to 24 June 2019, including study filter

1	exp LEUKEMIA, LYMPHOCYTIC, CHRONIC, B-CELL/
2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em$ adj3 (chronic$ or cronic$ or chroniq$ or well-differentia$)).tw,kf,ot.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic$ or cronic$ or chroniq$ or well-differentia$)).tw,kf,ot.
4	((small cell$ or small-cell$) adj3 lymphom$).tw,kf,ot.
5	(lymphom$ adj2 lymphocyt$).tw,kf,ot.
6	lymphoplasma?ytoid.tw,kf,ot.
7	(cll or b-cll or bcll).tw.
8	sll.tw.
9	or/1-8
10	Validat$.tw. or Predict$.ti. or Rule$.tw. or (Predict$ and (Outcome$ or Risk$ or Model$)).tw. or ((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).tw. or (Decision$.tw. and ((Model$ or Clinical$).tw. or LOGISTIC MODELS/)) or (Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or c-statistic or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable").tw.
11	9 and 10
12	exp ANIMALS/ not HUMANS/
13	11 not 12
14	limit 13 to ed=20180919-20190624

key: exp # /: explode # MeSH subject heading, tw: text word, kf: keyword heading word, ot: original title, * or $: truncation, ?: wildcard, adj#: adjacent within # number of words
search line #10: Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG, et al. (2012) Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLOS One 2012; 7(2): e32844, doi:10.1371/journal.pone.0032844

Appendix 3. Embase search strategy up to 24 June 2019

1	exp CHRONIC LYMPHATIC LEUKEMIA/

(Continued)

2	((lymphocytic* or b-cell or b-lymphocytic or lymphoblastic or lymphatic) adj1 leuk?em* adj3 (chronic* or cronic* or chroniq* or well-differentia*)).tw,kw.
3	((lymphoplasmacytoid or lymphocytic) adj1 lymphom* adj3 (chronic* or cronic* or chroniq* or well-differentia*)).tw,kw.
4	((small cell* or small-cell*) adj3 lymphom*).tw,kw.
5	(lymphoma* adj2 lymphocyt*).tw,kw.
6	lymphoplasma?ytoid.tw,kw.
7	(cll or b-cll or bcll).tw.
8	sll.tw.
9	or/1-8
10	Validat*.tw. or Predict*.ti. or Rule*.tw. or (Predict* and (Outcome* or Risk* or Model*)).tw. or ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*)).tw. or (Decision*.tw. and ((Model* or Clinical*).tw. or STATISTICAL MODEL/)) or (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model*)).tw. or ("Stratification" or "Discrimination" or "Discriminate" or c-statistic or "Area under the curve" or "AUC" or "Calibration" or "Indices" or "Algorithm" or "Multivariable").tw.
11	9 and 10
12	exp ANIMAL/ not HUMAN/
13	11 not 12
14	limit 13 to embase
15	limit 14 to em=197401-201926

key: exp # /: explode # MeSH subject heading, tw: text word, kw: keyword, *: truncation, ?: wildcard, adj#: adjacent within # number of words searchline #10: Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLOS One 2012; 7(2): e32844, doi:10.1371/journal.pone.0032844 (translated in Embase)

Appendix 4. ClinicalTrials.gov search strategy up to 05 March 2020

Basic search

prognostic OR predictive OR model OR score | chronic lymphocytic leukemia OR CLL

Advanced search

Intervention/treatment: prognostic OR predictive OR Model OR score

Recruitment: all studies

Study type: all studies

Appendix 5. WHO ICTRP search strategy

Basic search

Chronic lymphocytic leukemia AND prognostic

Chronic lymphocytic leukemia AND predictive

Chronic lymphocytic leukemia AND model

Chronic lymphocytic leukemia AND score

CLL AND prognostic

CLL AND predictive

CLL AND model

CLL AND score

Advanced search

Condition: chronic lymphocytic leukemia OR CLL

Intervention: prognostic OR predictive Or model OR score

Recruitment status: ALL

Appendix 6. Prognostic models with external validation studies

Author, year	Model (= 1) or Score (= 2)	Outcome	Recruitment period	Number of patients	Study design	Number of predictors included in final model	Discrimination	Calibration
Delgado 2017	2	OS	NR	524	Retrospective cohort	IgHV; FISH abnormalities (2)	0.682	Calibration plot for survival at 5 years
International CLL IPI working group	2	OS	1997 to 2007	3472	RCTs	IgHV; B2M; age; stage; TP53 (5)	0.723	NR
Morabito 2009	2	OS	NR	262	Unclear	IgHV; ZAP-70; CD38 (3)	NR	NR
Pflug 2014	2	OS	1997 to 2006	1223	RCTs	IgHV; B2M; age; gender; del17p; ECOG; sTK; del11q (8)	0.75	NR
Wierda 2007	1	OS	1981 to 2004	1674	Retrospective cohort	IgHV; B2M; age; stage; gender; ALC (6)	0.84	Calibration plot for survival at 5 years
Stephens 2015	2	OS	2002 to 2012	114	Retrospective cohort	age; LDH; ECOG (3)	0.73	NR
Molica 2005	2	PFS	1991 to 2000	1138	Retrospective cohort	stage; gender; lymphocytosis; LDT (4)	NR	NR
Baliakas 2019	2	TFS	NR	1900	Retrospective cohort	IgHV; TP53; FISH abnormalities; gender (4)	0.745	NR
Gentile 2016	2	TFS	2007 to NR	480	Prospective cohort	IgHV; B2M; stage; ALC (4)	0.75	Hosmer-May test

(Continued)

Rossi 2013	2	TFS	NR	673	Retrospective cohort	TP53; FISH abnormalities; SF3B1 (3)	0.642	Bias-corrected calibration slope, 0.965
Wierda 2011	1	TFS	2004 to 2009	930	Retrospective cohort	IgHV; LDH; number of involved lymph nodes; FISH abnormalities; lymph node size in neck; interaction term (6)	NR	NR
Stephens 2015	2	TFS	2002 to 2012	114	Retrospective cohort	age; stage; ECOG; del11q; WBC count (5)	0.84	NR

Abbreviations: OS: overall survival, PFS: progression-free survival; TFS: treatment-free survival; NR: not reported; RCT: randomised controlled trial; IgHV: immunoglobulin heavy chain variable region gene mutational status; B2M: beta 2 microglobulin; FISH: fluorescence in situ hybridisation genomic aberrations; Zap-70: zeta-chain-associated protein kinase 70; CD38: cluster of differentiation 38; LDH: lactate dehydrogenase; LPL: lipoprotein lipase; ECOG: Eastern Cooperative Oncology Group performance status

Appendix 7. Prognostic models without external validation studies

Author, year	Model (=1) or Score (=2)	Outcome	Recruitment period	Number of patients	Study design	Predictors included in final model (number)	Discrimination	Calibration
Baumann 2014	2	OS	1990 to 2012	949	Retrospective cohort	B2M; stage; Zap-70; Comorbirities (4)	NR	NR
Bulian 2012	1	OS	1996 to 2008	620	Retrospective cohort	IgHV; B2M; age stage; FISH; gender (6)	0.79	NR
Bulian 2011	1	OS	1996 to NR	1480	Retrospective cohort	B2M; age; stage; gender (4)	0.78	Calibration plot for survival at 5 years
Furundarena 1994	1	OS	1973 to 1992	150	Retrospective cohort	age; stage; gender; splenomegaly (4)	NR	NR
Haferlach 2010	2	OS	2005 to 2008	399	NR	IgHV; age; TP53; translocation IGH@ on 14q32; number of chromosome aberrations based on CBA; WBC count (6)	NR	NR
Jarque 1991, model A	1	OS	1969 to 1988	187	Retrospective cohort	Age; Spinal infiltration; BUN (3)	NR	NR
Jarque 1991, model B	1	OS	1969 to 1988	187	Retrospective cohort	Albumin; Spinal infiltration; BUN; cervical adenopathies (4)	NR	NR

(Continued)

Lee 1987	1, 2	OS	1970 to 1983	325	Retrospective cohort	Age; LDH; uric acid; alkaline phosphatase; external lymphadenopathy (5)	NR	NR
Liang 2018, model 3a	2	OS	2000 to 2014	501	Retrospective cohort	IgHV; B2M; stage; TP53; albumin; HBV infection (6)	time-dependent ROC curve	NR
Molica 1990	1	OS	NR	221	Unclear	age; stage; LAR (3)	NR	O/E ratios
Molica 2019	1, 2	OS	2000 to NR	108	Retrospective cohort	age; ADL; CIRS (3)	c = 0.70 (95% CI 0.53 to 0.87)	NR
Pepper 2012	1	OS	NR	1154	Unclear	IgHV; age; CD38; LDT (4)	NR	NR
Rozman 1982	1	OS	NR	150	Unclear	Splenomegaly; lymphocytosis; anaemia; thrombocytopenia (4)	NR	O/E probabilities
Rozman 1984	1	OS	NR	329	Unclear	lymphadenopathy; haemoglobin; bone marrow pattern; hepatomegaly (4)	NR	O/E ratios
Stamatopoulos 2010	2	OS	NR	170	Unclear	ZAP-70; LPL; miR-29 (3)	NR	NR
Tsimberidou 2007	2	OS	1985 to 2005	1893	Retrospective cohort	Age; B2M; Del17p; albumin; creatinine (5)	NR	NR
Visentin 2015	2	OS	1983 to 2013	608	Retrospective cohort	IgHV; CD38; FISH abnormalities (3)	0.88	NR

(Continued)

Wierda 2009	1, 2	OS	1985 to 2004	595	Unclear	Age; B2M; treatment (3)	NR	Concordance index for calibration curve (nomogram): 0.81
Friedrichs 2011	2	PFS	NR	134	NR	IgHV; CD38; iLR (3)	NR	NR
Leotard 2000	1	PFS	1985 to 1997	88	Prospective cohort	B2M; LDH; albumin; sCD23 (4)	NR	NR
Letestu 2010	1, 2	PFS	NR	339	NR	B2M; CD38; lymphocytosis; sTK (4)	NR	NR
Antic 2011	2	TFS	NR	33	Unclear	B2M; LPL; sVEGF (3)	NR	NR
Cavallini 2017	2	TFS	NR	125	NR	SF3B1; ERK1/2 phosphorylation (2)	NR	NR
Del Guidice 2005	2	TFS	2003 to 2004	201	Prospective cohort	ZAP-70; CD38 (2)	NR	NR
Gentile 2009	2	TFS	NR	222	NR	IgHV; B2M; CD38 (3)	NR	NR
Haferlach 2010	2	TTT	2005 to 2008	399	NR	IgHV; ATM deletion; translocation involving IGH@ locus on 14q32; number of chromosome abnormalities (4)	NR	NR
Li 2017b	2	TFS	2007 to 2015	406	NR	IgHV; stage; Del17p (3)	NR	NR

(Continued)

Liang 2018 model 3b	2	TFS	2000 to 2014	501	Retrospective cohort	IgHV; LD; lymphocytosis; platelets; HBV (5)	time-dependent ROC curve	NR
Metze 2000	2	TFS	1995 to 1998	57	Prospective cohort	total tumor mass score; percentage of cells with 1 AgNOR cluster (2)	NR	NR
Miao 2019, model A (CLL-IPI-S)	1, 2	TFS	2000 to 2017	399	Retrospective cohort	IgHV; B2M; Age; stage; TP53; SF3B1 (6)	AUC = 0.762	calibration plot
Miao 2019, model B (CLL-PI)	1, 2	TFS	2000 to 2017	399	Retrospective cohort	IgHV; B2M; stage; TP53; SF3B1 (5)	AUC = 0.773	calibration plot
Molica 2010	2	TFS	1998 to 2008	150	Retrospective cohort	IgHV; BAFF (2)	0.86	NR
Molica 2015	1	TFS	NR	322	Unclear	IgHV; B-cell count; interaction term (3)	NR	NR
Morabito 2011	2	TFS	NR	449	Retrospective cohort	stage; ZAP-70; FISH abnormalities; SFLC (kappa + delta) (4)	NR	NR
Pepper 2012	1	TFS	NR	1154	Unclear	IgHV; age; LDT; CD38 (4)	NR	NR
Qin 2018	2	TTT	2008 to 2016	334	Retrospective cohort	IgHV; EBV DNA positive; stage; TP53; ALC; HBsAg+; (6)	AUC = 0.768	NR
Schweighofer 2011	1	TFS	NR	131	Retrospective cohort	SKI (gene); SLAMF1 (gene) (2)	NR	NR

(Continued)

Stamatopoulos 2010	2	TFS	NR	170	Unclear	ZAP-70; LPL; miR-29c (3)	NR	NR
Vetro 2018	1	TTT	NR	171	Unclear	IgHV; SF3B1; percentage of B cells; genomic aberrations (4)	NR	NR
Wierda 2009	1	TFS	1985 to 2004	595	Unclear	Age; B2M; treatment; bone marrow lymphocytes (4)	NR	NR

Abbreviations: OS: overall survival, PFS: progression-free survival; TFS: treatment-free survival; TTT: time to treatment; NR: not reported; RCT: randomised controlled trial; IgHV: immunoglobulin heavy chain variable region gene mutational status; B2M: beta 2 microglobulin; FISH: fluorescence in situ hybridisation genomic aberrations; Zap-70: zeta-chain-associated protein kinase 70; CD38: cluster of differentiation 38; LDH: lactate dehydrogenase; LPL: lipoprotein lipase; ECOG: Eastern Cooperative Oncology Group performance status; LAR: lymphocyte accumulation time; sVEGF: soluble vascular endothelial growth factor; iLR: immature laminin receptor; CIRS: Cumulative Illness Rating Score; ADL: Katz Activity of Daily Living; EBV: Epstein-Barr virus; HBsAg: surface antigen of hepatitis B virus

Appendix 8. Reporting deficiencies in all included studies with at least one external validation study (development + validation studies, listed per model)

Prognostic model	Number of cohorts	Recruitment period reported in:	Study design reported in:	Observation time (range) reported in:	Total sample size reported in:	Number of events reported in:	Calibration reported in:	Discrimination (95% CI or SE) reported in:
Baliakas (Baliakas 2019)	2	0	2	1 (1)	2	0	0	1 (1)
Barcelona-Brno (Delgado 2017)	6	3	4	6 (6)	6	5	1	5 (3)
CLL-IPI (Bahlo 2016)	11	9	9	10 (6)	10	9	5	8 (6)
GCLLSG (Pflug 2014)	2	1	2	2 (0)	2	2	2	2 (2)
GIMEMA (Molica 2005)	2	2	2	1 (1)	2	1	0	1 (1)
MDACC 2007 (Wierda 2007)	10	7	8	8 (9)	9	9	2	8 (7)
MDACC 2011 (Wierda 2011)	2	1	2	2 (2)	2	2	0	1 (1)
Morabito model (Morabito, 2009)	2	2	1	2 (2)	2	2	0	0 (0)
O-CLL1 model (Gentile 2016)	3	1	1	3 (2)	3	3	0	3 (2)
Rossi (Rossi 2013)	3	3	3	3 (0)	3	2	0	2 (0)
Stephens (Stephens 2015)	2	1	1	0 (0)	2	0	0	2 (0)
Stephens (Stephens 2015)	2	1	1	0 (0)	2	0	0	2 (0)

HISTORY

Protocol first published: Issue 1, 2016
Review first published: Issue 7, 2020

CONTRIBUTIONS OF AUTHORS

Nina Kreuzberger: screening and selection of studies, development of data extraction form and data extraction, characteristics of studies, 'Risk of bias' assessment, statistical analysis, writing and drafting of the review, communication with and between authors

Johanna AAG Damen: statistical analysis

Marialena Trivella: statistical input, screening and selection of studies, data extraction, 'Risk of bias' assessment

Lise J Estcourt: medical and content input, screening and selection of studies, data extraction, 'Risk of bias' assessment

Angela Aldin: screening and selection of studies, data extraction, 'Risk of bias' assessment

Lisa Umlauff: data extraction, 'Risk of bias' assessment, characteristics of included studies

Maria Vazquez Montes: prototype of data extraction form, support in data extraction, 'Risk of bias' assessment

Robert Wolff: 'Risk of bias' input

Karel GM Moons: methodological input on reviews of prognosis studies

Ina Monsef: search strategy development

Farid Foroutan: 'Risk of bias' input

Karl-Anton Kreuzer: medical and content input

Nicole Skoetz: protocol development, screening and selection of studies, data extraction, 'Risk of bias' assessment, extensive proofread and comments on the review draft

DECLARATIONS OF INTEREST

Nina Kreuzberger: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Johanna AAG Damen: none known

Marialena Trivella: I am working as a statistical editor in a number Cochrane groups, and I declare that my work as a statistical editor is independent to this published work where I participate. The University of Oxford received a small part of the grant from the Federal Ministry of Education and Research, Germany, as reimbursement for my time spent on the project.

Lise J Estcourt: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Angela Aldin: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Lisa Umlauff: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

Maria Vazquez Montes: I am sponsored by the BHF to contribute on a similar prognostic models review for heart failure which allowed me to effectively participate in the work under consideration.

Robert Wolff: As employee of Kleijnen Systematic Reviews, I was the lead author of a systematic review on prostate cancer, commissioned by Elekta, Nucletron.

Karel Moons: none known

Ina Monsef: none known

Farid Foroutan: none known

Karl-Anton Kreuzer: board member and consultant for AbbVie, Alexion, Amgen, Ariad, Baxter, Bayer Health Care, Biotest, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Chugai, Gilead, Glaxo-SmithKline, Grifols, Hexal, Janssen, Jazz Pharmaceuticals, Leo, Mundipharma, MSD, Novartis, Pfizer, Roche, Shire, Teva. Grants, fees, honoraria and travel grants from AbbVie, Alexion, Amgen, Ariad,

Baxter, Bayer Health Care, Biotest, Boehringer Ingelheim, Bristol Myers Squibb, Celgene, Chugai, Gilead, Glaxo-SmithKline, Grifols, Hexal, Janssen, Jazz Pharmaceuticals, Leo, Mundipharma, MSD, Novartis, Pfizer, Roche, Shire, Teva.

Nicole Skoetz: My institution received a grant from the Federal Ministry of Education and Research, Germany to conduct this review.

SOURCES OF SUPPORT

Internal sources

- University Hospital of Cologne, Department I of Internal Medicine, Germany
- NHS Blood and Transplant, UK

External sources

- Grant by the Federal Ministry of Education and Research (Grant no. 01KG1711), Germany

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We changed the title of the review to clarify the included population (newly-diagnosed adults with CLL). In the protocol, the title was 'Prognostic models for chronic lymphocytic leukaemia: an exemplar systematic review and meta-analysis'.

Structure of the review

As prognosis reviews are a new review type within Cochrane, little guidance had been published at protocol stage. Recently, a review template has been developed and published by the Cochrane Prognosis Methods group, which we have adopted for this review (methods.cochrane.org/prognosis/our-publications). In addition, we have added two sections ('Selection of studies' and ''Risk of bias' assessment') to report our group decisions transparently.

Search methods

We did not search PubMed, because content from PubMed can be identified via MEDLINE.

We added a search for Embase based on editorial comments.

We added a search of the following databases for ongoing trials instead of the metaRegister of Controlled Trials:

- ClinicalTrials.gov
- World Health Organization International Clinical Trials Registry Platform (WHO ICTRP)

We did not search conference proceedings for abstracts, because the paucity of information would not allow us to apply the 'Risk of bias' tool. To ensure that we identified all validation studies corresponding to a developed model, we screened the citations of all included studies (Web of Science, October 2018).

Inclusion and exclusion of studies

Originally, we planned to include only models published after 1990. However, prognostic model studies are often conducted based on retrospective data, which sometimes include the analysis of blood samples frozen at diagnosis and analysed many years after. To give the same chance to studies with samples that were analysed immediately and after several years for inclusion in this review, we decided not to limit our search strategy based on date of publication. As a clarification, any studies that explicitly aimed at defining a new staging system was excluded during title and abstract screening.

We decided to include prognostic models with the outcomes 'time-to-first treatment' and 'treatment-free survival' as additional outcomes, because individuals with low-risk CLL can often live a long time without disease progression and the need for treatment. Usually, at disease progression, treatment is indicated, which means that these two outcomes are quite similar and both are meaningful for the patient.

After several rounds of discussion, we decided that genetic signatures are beyond the scope of this review because: a) we prespecified that we would exclude factor identification studies, which most of the genetic signature studies are to start with; b) they are not yet ready to be applied in clinical practice due to the unavailability of detailed genetic information that these signatures use and the complexity of the algorithms; and c) more often than not, it is not feasible to assess a genetic signature study in the same context as a prognostic model.

Analysis and 'Risk of bias' assessment

Instead of the CHARMS checklist, we used the recently published PROBAST-tool to assess the risk of bias of individual studies (Wolff 2019). The checklist does not originally aim at the assessment of risk of bias, but as a guide for critical appraisal and data extraction. However, at the protocol stage, no tool specific to 'Risk of bias' assessment for prognostic models had been published.

The analysis was not specified precisely in the protocol, as methods for systematic reviews of prognosis are just developing. As described in the protocol, we decided to follow the current recommendations of the Cochrane Prognosis Methods Group. We added a more detailed description in the section Data collection and analysis. In short, we meta-analysed the performance measures of the various external validation studies per model where data were available. We planned to summarise the measures of calibration. Unfortunately, calibration in the form of O:E ratios or calibration plots was rarely reported. Instead, many studies only presented the observed outcome frequency per subgroup at a specific time point. To gain an overview of this information among several external validations, during the review process and in collaboration with the Cochrane Prognsosis Methods Group, we decided to represent the outcome frequency graphically together with the pooled outcome frequency in a table.

As we only meta-analysed the external validation studies of a prognostic model, we analysed only models that were externally validated several times. Although we have included all prognostic models and scores that we identified, we did not describe models without any external validation studies in detail, because it is not recommended to use prognostic scores or models without any testing in independent cohorts, especially when the development sample was small (Moons 2015; Steyerberg 2013). References to these studies can be found in the list of studies awaiting classification; further information can be found in Appendix 7.

In the protocol, we did not specify any sensitivity analyses. We decided post hoc to explore the effect of including area under the curve (AUC) as a performance measure for discrimination instead of the c-statistic (Figure 23), the effect of the Newcombe estimation method for 95% CIs of the c-statistic (Figure 24; Figure 25), and the difference between studies using the original predictor or a proxy (del(17p) instead of TP53 mutation (Figure 26).

GRADE

We did not apply GRADE as no GRADE guidance has yet been developed to assess certainty of evidence from meta-analysis of prognostic models.

NOTES

Parts of this review, especially the methods, are from the Cochrane Haematology standard template.

INDEX TERMS

Medical Subject Headings (MeSH)

Age Factors; Bias; Biomarkers, Tumor; Calibration; Confidence Intervals; Discriminant Analysis; Disease-Free Survival; Genes, p53 [genetics]; Immunoglobulin Heavy Chains [genetics]; Immunoglobulin Variable Region [genetics]; Leukemia, Lymphocytic, Chronic, B-Cell [*mortality] [pathology]; *Models, Theoretical; Neoplasm Staging; Prognosis; Progression-Free Survival; Receptors, Antigen, B-Cell [genetics]; Reproducibility of Results; Tumor Suppressor Protein p53 [genetics]

MeSH check words

Adult; Female; Humans; Male