blood

Regular Article

CLINICAL TRIALS AND OBSERVATIONS

A model for predicting effect of treatment on progression-free survival using MRD as a surrogate end point in CLL

Natalie Dimier,[1] Paul Delmar,[2] Carol Ward,[2] Rodica Morariu-Zamfir,[2] Günter Fingerle-Rowson,[2] Jasmin Bahlo,[3] Kirsten Fischer,[3] Barbara Eichhorst,[3] Valentin Goede,[3] Jacques J. M. van Dongen,[4,5] Matthias Ritgen,[6] Sebastian Böttcher,[6,7] Anton W. Langerak,[4] Michael Kneba,[6] and Michael Hallek[3]

[1]Roche Products Ltd, Welwyn, United Kingdom; [2]F. Hoffmann-La Roche Ltd, Basel, Switzerland; [3]Department I of Internal Medicine, Center of Integrated Oncology Cologne-Bonn, University Hospital Cologne, Cologne, Germany; [4]Department of Immunology, Laboratory for Medical Immunology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands; [5]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands; [6]Second Department of Medicine, University of Schleswig-Holstein, Kiel, Germany; and [7]Department of Hematology, Oncology and Palliative Medicine, Center for Internal Medicine, University of Rostock, Rostock, Germany

KEY POINTS

- **Meta-analysis of 3 randomized clinical trials shows a statistically significant relationship between treatment effects on PFS and MRD.**

- **Meta-regression model supports use of MRD as a primary end point in clinical trials of chemoimmunotherapy in CLL.**

Our objective was to evaluate minimal residual disease (MRD) at the end of induction treatment with chemoimmunotherapy as a surrogate end point for progression-free survival (PFS) in chronic lymphocytic leukemia (CLL) based on 3 randomized, phase 3 clinical trials (ClinicalTrials.gov identifiers NCT00281918, NCT00769522, and NCT02053610). MRD was measured in peripheral blood (PB) from treatment-naïve patients in the CLL8, CLL10, and CLL11 clinical trials, and quantified by 4-color flow cytometry or allele-specific oligonucleotide real-time quantitative polymerase chain reaction. A meta-regression model was developed to predict treatment effect on PFS using treatment effect on PB-MRD. PB-MRD levels were measured in 393, 337, and 474 patients from CLL8, CLL10, and CLL11, respectively. The model demonstrated a statistically significant relationship between treatment effect on PB-MRD and treatment effect on PFS. As the difference between treatment arms in PB-MRD response rates increased, a reduction in the risk of progression or death was observed; for each unit increase in the (log) ratio of MRD⁻ rates between arms, the log of the PFS hazard ratio decreased by −0.188 (95% confidence interval, −0.321 to −0.055; P = .008). External model validation on the REACH trial and sensitivity analyses confirm the robustness and applicability of the surrogacy model. Our surrogacy model supports use of PB-MRD as a primary end point in randomized clinical trials of chemoimmunotherapy in CLL. Additional CLL trial data are required to establish a more precise quantitative relationship between MRD and PFS, and to support general applicability of MRD surrogacy for PFS across diverse patient characteristics, treatment regimens, and different treatment mechanisms of action. (*Blood*. 2018;131(9):955-962)

Introduction

In recent years, there has been considerable progress in the treatment of chronic lymphocytic leukemia (CLL), with median progression-free survival (PFS) now approaching 5 years in first-line CLL studies.[1] Because PFS is the standard primary end point used in phase 3 CLL clinical trials, this improvement in outcome requires long-term follow-up in trials of new experimental therapies.

To facilitate the development of novel treatments and ensure timely patient access to more efficacious therapies, shorter-term end points are desired for future CLL clinical trials. A potential surrogate for PFS in this setting is the measurement of minimal residual disease (MRD) response at the end of treatment. Although not formally included in the International Workshop on Chronic Lymphocytic Leukemia (iwCLL) 2008 definition of response,[2] MRD has been shown to be an independent prognostic factor of efficacy in both single-arm/patient series and randomized phase 3 trials of chemotherapy and chemoimmunotherapy agents[3-9] and monoclonal antibodies.[8]

MRD is a sensitive measure of the remaining tumor load after treatment, and therefore is an indicator of the depth of response to treatment. The vast improvement in MRD detection technology over the last 2 decades now allows a robust and reliable quantification of MRD in peripheral blood (PB) and/or bone marrow (BM), and therefore facilitates an objective measurement of response to therapy. Polymerase chain reaction (PCR)–based and 4-color flow

cytometric (MRD flow) techniques have reliably established an MRD detection level of <1 leukemic cell per 10 000 leukocytes (10^{-4}). Both methods are widely used to assess MRD.[10-14]

Results from 3 randomized phase 3 studies of front-line chemoimmunotherapy in CLL conducted by the German CLL Study Group (GCLLSG) provide the rationale for assessing the value of MRD response as a potential surrogate end point for long-term outcome. Data from the CLL8 study support the hypothesis of MRD response as a surrogate end point for both overall survival (OS) and PFS.[3,15] MRD measured in PB at 3 months posttherapy as per iwCLL 2008 guidelines for response assessment in CLL[2] were categorized according to low- (MRD < 10^{-4}, ie, <1 leukemic cell per 10^4 leukocytes), intermediate- ($\geq 10^{-4}$ to <10^{-2}), and high-level ($\geq 10^{-2}$) thresholds, and were associated with median PFS estimates of 68.7, 40.5, and 15.4 months, respectively.[3] Median OS was 48.4 months in patients with high MRD levels, but was not reached for patients with low or medium MRD levels.[3] In CLL10,[4] median PFS was 23.9 months in PB-MRD$^+$ (MRD $\geq 10^{-4}$) patients and 65.2 months in PB-MRD$^-$ patients (MRD < 10^{-4}; Roche data on file). In CLL11,[5,16] median PFS by PB-MRD yielded similar results; median PFS was 19.4 months in PB-MRD$^+$ patients and not reached in PB-MRD$^-$ patients.[5]

Correlation of a short-term end point (MRD) with a long-term end point (PFS) is insufficient to establish surrogacy.[17] Assessment of a potential surrogate end point requires demonstration of the prognostic value of the surrogate for long-term outcome, and evidence that treatment effect on the surrogate reliably predicts treatment effect on the long-term outcome.[18] Here, we evaluate MRD response (negative [MRD < 10^{-4}] vs positive [MRD $\geq 10^{-4}$]) as a potential surrogate end point for PFS in CLL by developing a meta-regression model for predicting the treatment effect on PFS from the treatment effect on MRD. For improved precision, this analysis is based on a combined analysis of the CLL8, CLL10, and CLL11 trials.

Methods

Patients

MRD was prospectively assessed in patients participating in the 3 multicenter, randomized, open-label, phase 3 clinical trials (ClinicalTrials.gov identifiers NCT00281918, NCT00769522, and NCT02053610). In all 3 studies, MRD was assessed in PB in all patients and in BM only in patients with complete response (CR). In CLL8, only patients enrolled in Germany and Austria had MRD assessments conducted. Trial protocols were approved by the relevant institutional review board and ethics committee of each participating center. Patients provided written informed consent to participate in the trials and to undergo MRD testing. The designs of the 3 trials have been previously reported.[4,5,15] Key results are summarized in Table 1. The primary end point of each trial was investigator-assessed PFS. In this analysis, for the noninferiority CLL10 trial, FCR was considered the experimental arm to be consistent with the CLL8 experimental arm. Patients were included in the MRD analysis (MRD-evaluable population) if they have MRD-PB measured at the time of the final response assessment, within 75 to 195 (CLL8 and CLL10) or 56 to 190 (CLL11) days after the last day of treatment. If multiple MRD results within this time window were available, the earliest dated result was used. Patients with no MRD result but death/progressive disease

shortly after last dose (within 90 [CLL8 and CLL10] or 56 [CLL11] days of last dose) were counted as MRD$^+$.

MRD assessment

MRD was quantified using an international standardized approach by flow cytometry analysis in CLL8 and CLL10[12,19] and by allele-specific oligonucleotide real-time quantitative PCR in CLL11 according to the EuroMRD guidelines[13] (supplemental Methods, available on the *Blood* Web site). Concordance between flow-based vs PCR-based MRD assessment has previously been demonstrated and quantitative MRD levels assessed by both techniques were closely correlated, irrespective of therapy. The sensitivity and specificity of MRD flow was not influenced by the presence of rituximab in the PB.[11] PB samples were taken at baseline in each trial, and at predefined postbaseline time points.[3-5] See supplemental Methods for further details on MRD assessment methodology.

BM-MRD samples were taken at final disease staging in patients achieving CR or CR with incomplete BM recovery (CRi) in each of the trials. Due to this small and potentially biased subset of patients with available BM-MRD results, BM-MRD was not included in the present analysis. A summary of BM-MRD results can be found in supplemental Table 1, with cross-tabulation comparing PB-MRD and BM-MRD results in supplemental Table 2.

Prediction model and analysis

To construct a prediction model for PFS, a weighted linear regression model was applied, using the logarithmic PFS HR as the predicted variable. The (log) relative risk of MRD, that is the (log) ratio of the MRD$^-$ rate in the experimental vs the control arm, was used to quantify treatment effect on MRD and was the only predictor in the model. To obtain sufficient data points to fit a regression model, patients were grouped according to region (6 regions in Germany according to the location of the trial site; CLL8, CLL10) or country (7 groups; countries with <45 patients were grouped according to geographic region of the trial site; CLL11).[20] Subgroups were weighted according to the number of PFS events observed (using the inverse of the square of the standard error of the logarithmic HR of PFS).

A relative measure of treatment effect on MRD was used to reflect that different trials may have different proportions of MRD response, dependent on treatment and patient population. The fitted model includes an intercept parameter to represent the expected PFS (log) HR when no difference in MRD$^-$ rates is observed. The "slope" parameter describes how the (log) HR is impacted through changes in the MRD response relative risk. The model was evaluated using the coefficient of determination (R^2), quantifying the proportion of variability in PFS HR that can be explained by MRD, and 95% confidence limits (CLs) and P values for the regression coefficients were calculated. A threshold of 5% was used to conclude statistical significance of model parameters.

As a sensitivity analysis, a regression model based on data from CLL8 and CLL10 only was also constructed. Furthermore, a model with the intercept term fixed at a value of 0 was constructed, such that the predicted HR for PFS is restricted to take a value of 1 (no difference) when there is no observed treatment effect on MRD. A further sensitivity analysis was conducted to create a regression model from CLL8, CLL10 and CLL11 when MRD negativity was defined taking into account the result in BM. In this model, patients

Table 1. PFS and MRD in CLL8, CLL10, and CLL11

	CLL8		CLL10		CLL11	
	FC, n = 184	FCR, n = 209	BR, n = 158	FCR,* n = 179	R-Clb, n = 245	G-Clb, n = 229
Patients	Previously untreated, physically fit		Previously untreated, physically fit; excluding patients with del(17p)		Previously untreated, with comorbidities	
Median observation time, mo	55		61		41	
PFS events, n (%)	119 (65)	107 (51)	104 (66)	87 (49)	220 (90)	163 (71)
PFS HR (95% CI)	0.63 (0.48-0.82)		0.63 (0.47-0.84)		0.44 (0.36-0.54)	
MRD negativity,† n (%)	57 (31)	143 (68)	99 (63)	128 (72)	8 (3)	82 (36)
MRD absolute difference, %	37		9		33	
MRD relative risk‡	2.20		1.14		10.38	

BR, bendamustine and rituximab; CI, confidence interval; FC, fludarabine and cyclophosphamide; FCR, fludarabine, cyclophosphamide, and rituximab; G-Clb, obinutuzumab plus chlorambucil; HR, hazard ratio; R-Clb, rituximab plus chlorambucil.

*For the purpose of the model, FCR was considered the experimental arm (noninferiority trial).

†PB at final response assessment within 75 to 195 (CLL8 and CLL10) or 56 to 190 (CLL11) days after the last day of treatment; if multiple values within this time window were available, the earliest dated result was used; patients with no MRD result but death/progressive disease shortly after last dose (within 90 [CLL8 and CLL10] or 56 [CLL11] days) are included as MRD$^+$ (MRD-evaluable population).

‡Relative risk = MRD$^-$ rate on experimental arm/MRD$^-$ rate on control arm. A value of 0.5 was added to all counts of MRD responders and nonresponders to avoid division by zero.[23] For all trials, PFS results are shown for the MRD-evaluable population. Data as of July 2010 (CLL8), May 2015 (CLL11), September 2016 (CLL10).

were considered as MRD$^-$ if they were negative in both PB and BM, and all other patients were considered as MRD$^+$. As an out-of-sample validation measure, the complete model was used to predict PFS HR in a non-GCLLSG CLL trial (REACH).[21]

Results

Patient population

Baseline demographics of the intention-to-treat (ITT) population (supplemental Table 3) were similar across the 3 trials, acknowledging the increased age expected in patients with comorbidities in CLL11.

Of 2162 patients randomized in the trials, data for PB-MRD or early progressive disease or death were available for 393, 337, and 474 patients in CLL8, CLL10, and CLL11, respectively (MRD-evaluable population) (Table 1). Demographic characteristics between the ITT and MRD-evaluable populations (supplemental Table 4) did not differ substantially across trials, indicating that the MRD-evaluable population is representative of the ITT population. Efficacy end point results were also comparable between MRD-evaluable and ITT populations in all studies.

Prediction of PFS

The proportion of patients with a PFS event and with MRD$^-$ status is shown in Table 1. Across the trials, PFS was longer and a larger proportion of patients achieved MRD negativity in the experimental arm vs the control arm. To assess the association between MRD and PFS within each trial, a Kaplan-Meier plot for PFS was provided for MRD$^-$ vs MRD$^+$ patients (Figure 1) and Cox regression models for PFS, accounting for MRD status with and without treatment, were fit to the data (supplemental Table 5). These models indicate a strong association between MRD and PFS and indicate how much of the effect of treatment on PFS can be captured by MRD. In the CLL8 study, there was no difference in PFS observed between the arms once PB-MRD was accounted for. In the CLL10 and CLL11 studies, PB-MRD captured some,
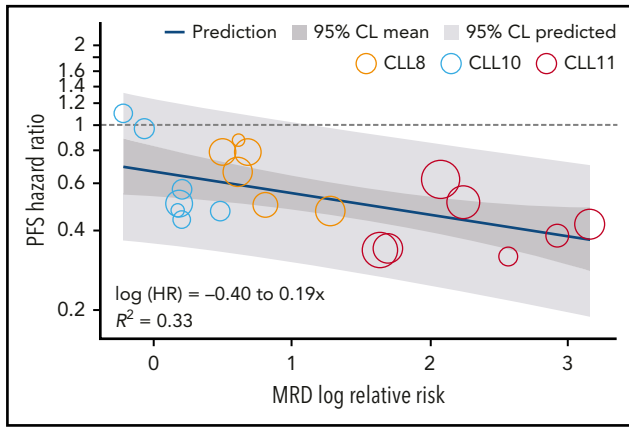
but not all, of the treatment effect in PFS as indicated by the low P values for the MRD-adjusted PFS.

The meta-regression model (Figure 2) showed a significant relationship between treatment effects on MRD and PFS; the log of PFS HR decreased by −0.188 (95% CI, −0.321 to −0.055) for each unit increase in the log relative risk of MRD (P = .008) as depicted by the regression line. This statistically significant slope parameter indicates that an increase in MRD response relative risk between trial arms is associated with improved PFS outcomes. The negative intercept parameter (−0.398; 95% CI, −0.617 to −0.179), representing the difference in PFS between arms when there is no difference in PB-MRD response rates, was also significantly different from zero (P = .001) indicating that some treatment effect remains in PFS when there is no difference in PB-MRD. The coefficient of determination of the model was $R^2 = 0.33$ indicating that approximately one-third of variability in the PFS HR can be explained through the observed MRD results.

Based on this model, predictions of PFS HR using a range of differences in MRD$^-$ rates are summarized in Table 2. These predictions suggest that risk of progression or death decreases as the ratio of MRD response rates increases (ie, a larger relative difference in MRD response rates is associated with a lower PFS HR). Because the model is based on subgroups of the 3 studies, the prediction intervals around future HRs are wide as a result of the low number of events within each subgroup. The prediction intervals were also calculated for a hypothetical phase 3 study with a larger number (170) of observed PFS events, to reflect an HR of 0.65, showing that the prediction is more precise with narrower prediction intervals, as shown in Table 2. When designing a future clinical trial based on MRD as a primary end point, the final column of this table illustrates the prediction interval that would be expected for the unobserved PFS HR based on the observed difference in MRD response rates.
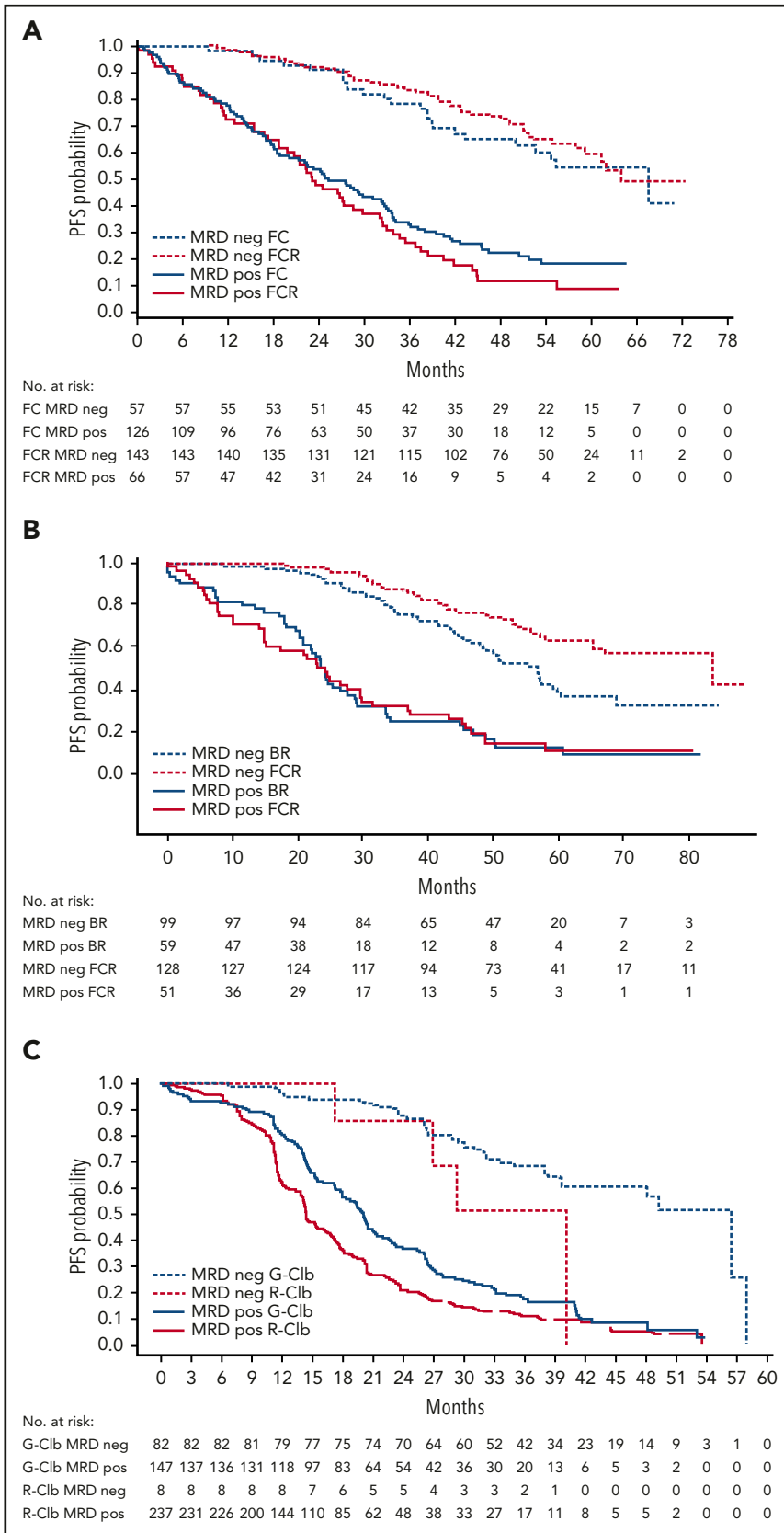
Figure 1. PFS by treatment and MRD response in the CLL8, CLL10, and CLL11 trials. Panels show (A) the CLL8 trial, (B) the CLL10 trial (2016 update), and (C) the CLL11 trial. MRD-evaluable populations in each trial.

A

No. at risk:

	0	6	12	18	24	30	36	42	48	54	60	66	72
FC MRD neg	57	57	55	53	51	45	42	35	29	22	15	7	0	0
FC MRD pos	126	109	96	76	63	50	37	30	18	12	5	0	0	0
FCR MRD neg	143	143	140	135	131	121	115	102	76	50	24	11	2	0
FCR MRD pos	66	57	47	42	31	24	16	9	5	4	2	0	0	0

B

No. at risk:

	0	10	20	30	40	50	60	70	80
MRD neg BR	99	97	94	84	65	47	20	7	3
MRD pos BR	59	47	38	18	12	8	4	2	2
MRD neg FCR	128	127	124	117	94	73	41	17	11
MRD pos FCR	51	36	29	17	13	5	3	1	1

C

No. at risk:

	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
G-Clb MRD neg	82	82	82	81	79	77	75	74	70	64	60	52	42	34	23	19	14	9	3	1	0
G-Clb MRD pos	147	137	136	131	118	97	83	64	54	42	36	30	20	13	6	5	3	2	0	0	0
R-Clb MRD neg	8	8	8	8	8	7	6	5	5	4	3	3	2	1	0	0	0	0	0	0	0
R-Clb MRD pos	237	231	226	200	144	110	85	62	48	38	33	27	17	11	8	5	5	2	0	0	0

Figure 2. Meta-regression based on combined CLL8, CLL10, and CLL11 patient populations (MRD-evaluable populations). Orange circles, CLL8; blue circles, CLL10; red circles, CLL11. Circle size in the figure reflects weighting of each subgroup to the overall model; those with least variability in PFS HR have the largest circle. Clustering of circles by trial reflects overall treatment effect for MRD and PFS in the trials.

Sensitivity analyses

Model including CLL8 and CLL10 only Both CLL8 and CLL10 trials included patients who were considered physically fit (Eastern Cooperative Oncology Group performance status 0-1 in CLL8, Cumulative Illness Rating Scale [CIRS] ≤ 6 and creatinine clearance ≥ 70 mL per minute in both CLL8 and CLL10), whereas CLL11 enrolled only patients with comorbidities (clinically meaningful burden of concomitant illnesses scoring >6 on the CIRS or a creatinine clearance of 30–69 mL per minute). To assess the potential impact of the heterogeneity of the patient population on the predictive value of MRD, the meta-regression model was also developed using data from CLL8 and CLL10 only. Results of this model demonstrate a consistent relationship between treatment effects on MRD and PFS, with an intercept of −0.322 and a slope parameter of −0.296 (P = .025 and .161, respectively, R^2 = 0.17). Although the slope parameter

is no longer statistically significant, the negative value indicates that the difference in PFS increases as the relative difference in MRD⁻ rates increases.

Model without intercept The meta-regression model developed herein enforces no restriction on the intercept term, such that the PFS HR is not constrained to take a value of 1 when there is no difference in MRD response rates. A further sensitivity analysis applied this constraint, to reflect that perfect surrogacy of MRD would mean that a lack of difference in MRD response rates would predict no difference in PFS. This model further demonstrates a strong relationship between treatment effects on MRD response rate and PFS, with a slope parameter of −0.381 (P < .0001 and R^2 = 0.75, Figure 3), further supporting the findings of the primary model.

Model based on MRD-BM To assess the impact of the use of PB in the primary model, a regression model was also constructed incorporating data from BM. In this model, patients were considered MRD⁻ if they had negative MRD status based in both PB and BM. Results demonstrate a consistent relationship between treatment effects on MRD and PFS, with an intercept of −0.252 and a slope parameter of −0.379 (P = .05 and .0015, respectively, R^2 = 0.44). This model is provided in supplemental Figure 1.

Model validation

Validation case study on non-GCLLSG data: REACH trial The REACH trial, which assessed FCR vs fludarabine and cyclophosphamide (FC)[21] in patients with previously treated CLL, was used to independently assess the reliability of the model predictions. MRD was tested in a subset of patients and negativity was observed in 43% and 31% of patients in the FCR and FC arms, respectively, giving a relative risk of 1.39. The model predicted a PFS HR of 0.63, which is consistent with the PFS HR of 0.65 for the REACH trial, thus supporting the reliability of model predictions.

## Table 2. Predictions based on the combined CLL8, CLL10, and CLL11 meta-regression model

Ratio of MRD⁻ rates, relative risk*	Log of relative risk	Predicted PFS HR	Individual prediction,† 95% CL	Mean prediction,‡ 95% CL	Prediction in a phase 3 study,§ 95% CL
2	0.69	0.59	0.32, 1.09	0.50, 0.69	0.43, 0.81
1.75	0.56	0.60	0.33, 1.12	0.51, 0.71	0.44, 0.83
1.5	0.41	0.62	0.33, 1.16	0.52, 0.74	0.45, 0.86
1.37	0.31	0.63	0.34, 1.18	0.52, 0.76	0.45, 0.88
1.25	0.22	0.64	0.34, 1.20	0.53, 0.78	0.46, 0.90
1.2	0.18	0.65	0.35, 1.21	0.53, 0.79	0.46, 0.91
1	0	0.67	0.36, 1.26	0.54, 0.84	0.47, 0.95

*MRD⁻ rate in experimental arm/MRD⁻ rate in control arm.

†Prediction for observation of PFS HR in a single trial.

‡Prediction for PFS HR underlying mean value.

§Prediction for observation of PFS HR in a new study with 170 PFS events (reflects a target HR of 0.65).
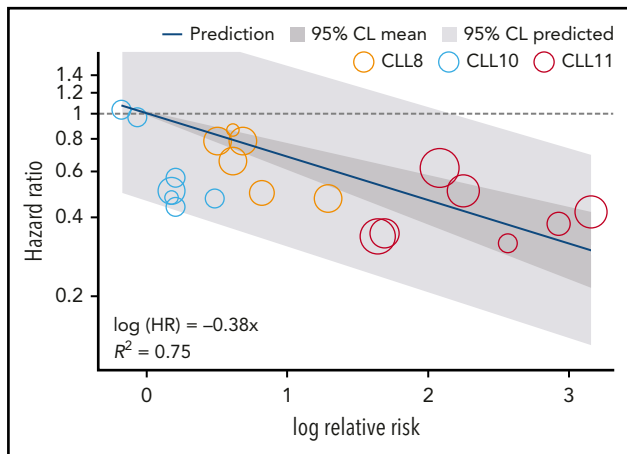
Figure 3. Meta-regression sensitivity analysis restricting PFS HR to be 1 when there is no difference in MRD rates. Based on combined CLL8, CLL10, and CLL11 patient populations (MRD evaluable populations). Orange circles, CLL8; blue circles, CLL10; red circles, CLL11. Circle size in the figure reflects weighting of each subgroup to the overall model; those with least variability in PFS HR have the largest circle. Clustering of circles by trial reflects overall treatment effect for MRD and PFS in the trials.

Discussion

The present analysis was conducted to determine whether the treatment effect on MRD response in PB at the end of induction treatment with chemoimmunotherapy can predict treatment effect on PFS in patients with CLL. To this end, we used PB-MRD data from 3 randomized, phase 3 trials to determine the strength of association between treatment effects using a meta-regression model. A statistically significant relationship between treatment effect on MRD and treatment effect on PFS was observed. The R^2 value measures how close the observed data are to the linear regression model, providing an estimate of how much of the variability in PFS HR can be explained through knowledge of the MRD response rate ratio. The value of 33% indicates that approximately one-third of the variability of the observed PFS HRs can be explained by the model. There are 2 factors to consider in the interpretation of this R^2 value: the variability in the data available for analysis and the significance of model parameters. The model includes data from 3 studies with very different treatment comparisons that are further split into smaller subgroups to enable fitting of the model, an approach discussed by Renfro et al.[22] The variability in observed treatment effects among the small subgroups is therefore apparent and reflected in the wide CIs for future predictions. However, when the model is used to predict treatment effect in a new phase 3 trial, it is expected that there will be a larger number of PFS events observed leading to more precise prediction of the PFS HR.[22] Additionally, the significance of model parameters indicates that even with the observed variability the relationship between the treatment effects on MRD response and PFS is very strong. The significant intercept term of the model indicates that some treatment effect in PFS remains when there is no difference in PB-MRD response rates between treatment arms. As can be seen from Figure 2, such a value lies at the extreme of the observed data and should therefore be interpreted with caution. Sensitivity analysis constraining the intercept term of the model to be zero, such that no difference in MRD response rates predicts no difference in PFS, supports the relationship between treatment effects on MRD and PFS. However, because such a constraint is artificial, further data are required to better quantify the remaining treatment effect in PFS when there is no observed difference in MRD response rates. Successful out-of-sample validation of the model was achieved in the REACH trial with close prediction of the PFS HR.

Data from the CLL8 study also support the hypothesis of MRD response as a surrogate end point for OS.[3] Meta-analysis of OS within the 3 studies included herein was thought to be limited by the shorter follow-up period in studies CLL10 and CLL11, with low numbers of deaths preventing meaningful conclusions. Therefore, OS was not explored.

Although BM is potentially more sensitive to MRD detection compared with PB,[3,5,9,12] BM assessment is limited by the patient burden of obtaining a sample and therefore less practical. Within each of the 3 trials, assessment of BM-MRD was performed at the time of final response staging only in patients achieving suspected CR/CRi, representing a biased subset of patients and preventing clear interpretation. Additionally, low proportions of patients achieving BM-MRD negativity implies that the possibility of meta-regression modeling of such small samples is unlikely. Therefore, BM-MRD data were not considered a more reliable assessment of surrogacy and were not included in the current analysis. Nonetheless, when each of the 3 studies was analyzed using Cox regression analyses, BM-MRD status was also found to be a significant independent prognostic factor for PFS (supplemental Table 5). Furthermore, a sensitivity model using BM-MRD status was consistent with the primary model based on PB-MRD and suggests that use of PB-MRD does not hamper the relationship between treatment effects on MRD and PFS.

In the CLL10 study, the PFS Cox regression and Kaplan-Meier curves indicate a small difference in PFS between BR and FCR in PB-MRD⁻ patients, with those treated with FCR having a slightly better long-term outcome. Although this difference was not observed when assessing BM-MRD, the lack of a statistically significant difference in outcomes based on BM-MRD may be due to the small patient numbers, and/or the bias introduced in this analysis through collection of BM-MRD samples only from responding patients. Measurements based on PB-MRD are taken from an unrestricted patient population, including both responders and nonresponders, making this a more representative sample to compare PFS between treatment groups. Furthermore, based on the baseline characteristics of patients included in the CLL10 study, the difference in outcome for MRD⁻ patients is likely impacted by an imbalance in the proportion of patients with IGHV mutation. In the FCR arm, 41.9% of patients in the MRD-evaluable population had a mutation, compared with 31.6% in the BR arm. Because IGHV mutation is a recognized prognostic factor for CLL, it is possible that this has had a minor impact on the results from this study. Indeed, Cox regression analysis for PFS adjusted for both IGHV status at baseline and MRD in PB indicated that there was no longer a statistically significant treatment difference between FCR and BR at the 5% level (P = .074). This suggests that the IGHV mutation imbalance is contributing to the apparent difference in long-term outcome between treatments. Therefore, the analysis of PB-MRD in CLL10, when adjusting for baseline imbalances, provides results that support the surrogacy relationship between PB-MRD and BM-MRD.

DIMIER et al

The trials selected for this analysis differed with respect to the patient populations and treatments under investigation; CLL8 and CLL10 enrolled patients who were considered physically fit and CLL11 comprised patients with comorbidities. Additionally, 5 different chemoimmunotherapy regimens were evaluated in these trials. However, to obtain a model that is generalizable to a wide range of clinical settings and to avoid excessive extrapolation, it was believed beneficial to have some level of heterogeneity between trials. Sensitivity analyses including only CLL8 and CLL10 data confirmed the relationship between treatment effects on MRD and PFS. The similarity of the results supports the use of MRD as a surrogate end point for PFS in future CLL clinical trials that contain induction treatment, using chemoimmunotherapies with a mechanism of action similar to those investigated in these studies. Inclusion of CLL data from patients with comorbidities did not impact the model conclusions and the added data from the CLL11 trial increased the reliability of the model.

As expected, several limitations may be considered. First, the wide CIs around the PFS prediction show that additional data are required to define a more precise quantitative relationship between treatment effects on PFS and MRD, although these wide CIs would be reduced if there were a higher number of PFS events observed in a future study. Second, although external validation of the model using REACH data suggests general applicability across treatment regimens and patient characteristics, the data used to generate the model were from a single research group (GCLLSG) and 3 clinical trials only. Though data were split into subgroups to generate sufficient data points and facilitate a robust regression analysis, the use of additional trials to serve as individual data points would avoid over-representation of trials with specific baseline and treatment characteristics. Importantly, use of the regression model to predict the PFS HR within key prognostic subgroups in each clinical trial (based on IGHV mutation, age [<65 years vs ≥65 years] and gender), demonstrated good agreement with the observed HRs in those subgroups, further supporting that the model holds in patients with different baseline disease and demographic characteristics. Third, the analysis assessed MRD at the end of induction treatment, in patients who did not receive any post-induction therapy. The effect of maintenance treatment on the ability of MRD to predict PFS and the effect of treatments that are administered continuously until disease progression remain unknown. The effect of treatments that have a different mechanism of action than those studied in this analysis, such as kinase inhibitors, also remains unknown. Finally, it should be noted that the model was not designed to predict the PFS of individual patients, but rather to facilitate design of randomized trials using MRD as a surrogate end point to predict treatment effect on PFS. Further work to investigate the relationship between treatment effects on MRD and PFS for agents that have a different mechanism of action, such as small-molecule inhibitors administered continuously until disease progression, could be considered.

In summary, the present MRD meta-regression model supports the use of MRD as a surrogate primary end point in randomized CLL clinical trials. Future analyses will aim to determine a more precise quantitative relationship between treatment effect on MRD and treatment effect on PFS while also assessing the general applicability of this relationship across CLL treatment regimens and patient populations.

Authorship

Contribution: G.F.-R., V.G., and M.H. designed the research; K.F., B.E., V.G., J.J.M.v.D., M.R., S.B., and M.H. performed the research; J.J.M.v.D., M.R., S.B., A.W.L., and M.K. contributed reagents and analytical tools; J.B. collected data; N.D., P.D., C.W., R.M.-Z., G.F.-R., and J.B. analyzed and interpreted the data; N.D., P.D., and C.W. performed statistical analyses; N.D., P.D., C.W., R.M.-Z., and G.F.-R. wrote the manuscript; and all authors reviewed and approved the manuscript.

Conflict-of-interest disclosure: N.D., P.D., C.W., and G.F.-R. have been employed by and own stock in F. Hoffmann-La Roche. R.M.-Z. has been employed by F. Hoffmann-La Roche. J.B. received honoraria and travel support from Roche. K.F. received travel grants from Roche. B.E. received research funding from Roche, AbbVie, Gilead Sciences, and Janssen Pharmaceuticals; has had a consulting/advisory role for AbbVie, Roche, Gilead, Janssen, and Novartis; and been on speakers' bureaus for Roche, Janssen, Gilead, and Celgene. V.G. received a research grant from Roche; was an advisory board member or had an advisory role for Roche, Gilead, and Janssen; received speaker honoraria from Roche, GlaxoSmithKline, Mundipharma, and Bristol-Myers Squibb; and received travel grants from Roche and Janssen. J.J.M.v.D. received: consultancy fees from Roche; patents and royalties from BD Biosciences, Cytognos, DAKO, InVivoScribe, and Immunostep; and laboratory services from Roche and BD Biosciences. M.R. received research funding from Roche and was a member of the Roche Board Of Directors and Advisory Board. S.B. received research funding from Roche, AbbVie, and Celgene, and received honoraria from Roche and AbbVie. A.W.L. received research funding from Roche and patents and royalties from InVivoScribe Technologies. M.K. received research funding from Gilead Sciences, Roche, and Mundipharma; received honoraria from AbbVie, Roche, and Mundipharma; and had a consulting/advisory role for AbbVie and Roche. M.H. was an advisory board member or had an advisory role for, and received honoraria and research support from, AbbVie, Amgen, Celgene, Roche, Gilead, Janssen, and Mundipharma.

ORCID profile: N.D., 0000-0002-8537-4962.

Correspondence: Natalie Dimier, Roche Products Limited, Hexagon Pl, 6 Falcon Way, Shire Park, Welwyn Garden City, Hertfordshire AL7 1TW, United Kingdom; e-mail: natalie.dimier@roche.com.

REFERENCES

1. Fischer K, Bahlo J, Fink AM, et al. Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood*. 2016; 127(2):208-215.

2. Hallek M, Cheson BD, Catovsky D, 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): 5446-5456.

3. Böttcher S, Ritgen M, Fischer K, et al. Minimal residual disease quantification is an independent predictor of progression-free and overall survival in chronic lymphocytic leukemia: a multivariate analysis from the randomized GCLLSG CLL8 trial. *J Clin Oncol*. 2012;30(9):980-988.

4. Eichhorst B, Fink AM, Bahlo J, et al; German CLL Study Group (GCLLSG). First-line chemoimmunotherapy with bendamustine and rituximab versus fludarabine, cyclophosphamide, and rituximab in patients with advanced chronic lymphocytic leukaemia (CLL10): an international, open-label, randomised, phase 3, non-inferiority trial. *Lancet Oncol*. 2016;17(7):928-942.

5. Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. *N Engl J Med*. 2014;370(12):1101-1110.

6. Kwok M, Rawstron AC, Varghese A, et al. Minimal residual disease is an independent predictor for 10-year survival in CLL. *Blood*. 2016;128(24):2770-2773.

7. Santacruz R, Villamor N, Aymerich M, et al. The prognostic impact of minimal residual disease in patients with chronic lymphocytic leukemia requiring first-line therapy. *Haematologica*. 2014;99(5):873-880.

8. Moreton P, Kennedy B, Lucas G, et al. Eradication of minimal residual disease in B-cell chronic lymphocytic leukemia after alemtuzumab therapy is associated with prolonged survival. *J Clin Oncol*. 2005;23(13): 2971-2979.

9. Kovacs G, Robrecht S, Fink AM, et al. Minimal residual disease assessment improves prediction of outcome in patients with chronic lymphocytic leukemia (CLL) who achieve partial response: comprehensive analysis of two phase III studies of the German CLL Study Group. *J Clin Oncol*. 2016;34(31):3758-3765.

10. Ghia P. A look into the future: can minimal residual disease guide therapy and predict prognosis in chronic lymphocytic leukemia? *Hematology Am Soc Hematol Educ Program*. 2012;2012:97-104.

11. Böttcher S, Stilgenbauer S, Busch R, et al. Standardized MRD flow and ASO IGH RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis. *Leukemia*. 2009; 23(11):2007-2017.

12. Rawstron AC, Villamor N, Ritgen M, et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. *Leukemia*. 2007;21(5):956-964.

13. van der Velden VH, Cazzaniga G, Schrauder A, et al; European Study Group on MRD detection in ALL (ESG-MRD-ALL). Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia*. 2007;21(4):604-611.

14. van der Velden VH, van Dongen JJ. MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR. *Methods Mol Biol*. 2009;538:115-150.

15. Hallek M, Fischer K, Fingerle-Rowson G, et al; German Chronic Lymphocytic Leukaemia Study Group. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet*. 2010;376(9747):1164-1174.

16. Goede V, Fischer K, Bosch F, et al. Updated survival analysis from the CLL11 study: obinutuzumab versus rituximab in chemoimmunotherapy-treated patients with chronic lymphocytic leukemia [abstract]. *Blood*. 2015;126(23). Abstract 1733.

17. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125(7):605-613.

18. Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biom J*. 2016;58(1):104-132.

19. Rawstron AC, Böttcher S, Letestu R, et al; European Research Initiative in CLL. Improving efficiency and sensitivity: European Research Initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. *Leukemia*. 2013;27(1):142-149.

20. Buyse M, Michiels S, Squifflet P, et al. Leukemia-free survival as a surrogate end point for overall survival in the evaluation of maintenance therapy for patients with acute myeloid leukemia in complete remission. *Haematologica*. 2011;96(8):1106-1112.

21. Robak T, Dmoszynska A, Solal-Céligny P, et al. Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia. *J Clin Oncol*. 2010; 28(10):1756-1765.

22. Renfro LA, Shi Q, Xue Y, Li J, Shang H, Sargent DJ. Center-within-trial versus trial-level evaluation of surrogate endpoints. *Comput Stat Data Anal*. 2014;78:1-20.

23. Agresti A. Categorical Data Analysis, 2nd ed. New York, NY: Wiley-Interscience; 2002.

DIMIER et al