

## TO THE EDITOR:

## The fragility index in 2010-2021 chronic lymphocytic leukemia randomized controlled trials

Cristina Bagacean,<sup>1,2</sup> Jean-Christophe Ianotto,<sup>1</sup> Nanthara Sritharan,<sup>3</sup> Florence Cymbalista,<sup>4</sup> Christian Berthou,<sup>1,2</sup> and Vincent Lévy<sup>3,5</sup>

<sup>1</sup>Department of Hematology, CHU Brest, Brest, France; <sup>2</sup>Department of Hematology, Université de Brest, INSERM, UMR1227, Lymphocytes B et Auto-immunité, Brest, France; <sup>3</sup>Département de Recherche Clinique, Assistance Publique-Hôpitaux de Paris, Hôpital Avicenne, Université Sorbonne Paris Nord, Bobigny, France; <sup>4</sup>Laboratoire d'Hématologie, Assistance Publique-Hôpitaux de Paris, Hôpital Avicenne, Bobigny, France; and <sup>5</sup>CRESS UMR 1153 ECSTRRA team, INSERM, Paris, France

Randomized controlled trials (RCTs) are designed to objectively assess the safety and efficacy of a specific intervention and represent a critical component of evidence-based medicine. Traditionally, frequentist analysis and threshold  $P$  values have been viewed as the arbiters of whether an intervention is effective.<sup>1</sup> This approach has been often criticized as being simplistic and insufficient, as the statistical methods used to analyze a randomized trial can easily modify the  $P$  value.<sup>2</sup> One approach to better communicate the limitations of  $P$  value thresholds is to report an additional metric that demonstrates how easily significance based on a threshold  $P$  value may be exceeded. The fragility index (FI) has been proposed as a mean to complement  $P$  value and inform its interpretation.<sup>3</sup> For a trial that demonstrates a statistically significant result ( $P < .05$ ), the FI is defined as the number of “nonevents” in the trial treatment group with the lowest event rate that must be changed to “events” in order for the  $P$  value calculated by the Fisher exact test to equal or exceed 0.05.<sup>1</sup> A lower FI will therefore indicate less statistically robust results. Since its first application in 2014,<sup>3</sup> the FI has been applied to several medical fields, including oncology.<sup>4,5</sup>

The past decade brought enormous advancements in the understanding of chronic lymphocytic leukemia (CLL) biology that incited numerous drug development programs.<sup>6</sup> Hence, new drug approvals were obtained that fundamentally changed the management of CLL and improved patient outcomes. The main objective of our study was to evaluate the robustness of CLL trials published during this dynamic and prolific period by calculating the FI.

We searched PubMed to identify RCTs for CLL between January 2010 and June 2021. Two reviewers independently screened all identified abstracts and performed data extraction (V.L. and C. Bagacean). Discrepancies were resolved with the involvement of a third author (F.C.). We included prospective phase 2 and 3 RCTs that (1) were 2 parallel arm or had a primary endpoint based on 2 arms, (2) had a primary endpoint based on response (complete response rate [CRR], overall response rate [ORR], or survival [progression-free survival [PFS] and event-free survival [EFS]) (Figure 1). Endpoints were defined according to the International Workshop on Chronic Lymphocytic Leukemia 2008 criteria.<sup>7</sup> We excluded secondary/cost-effectiveness studies, methodology studies, noninferiority trials, RCTs that reported statistically nonsignificant primary outcomes ( $P \geq .05$ ), and RCTs with incomplete information on the number of events, not permitting the FI calculation.

The FI was calculated from a 2-by-2 contingency table by the iterative addition of an event to the experimental group and concomitant subtraction of a nonevent to the same group, thereby maintaining a total constant number of events plus nonevents, until positive significance (defined as  $P < .05$ ) was lost.  $P$  values were calculated with Fisher exact test.<sup>3</sup> Other methods used for statistical analysis are briefly described in the supplemental Methods.

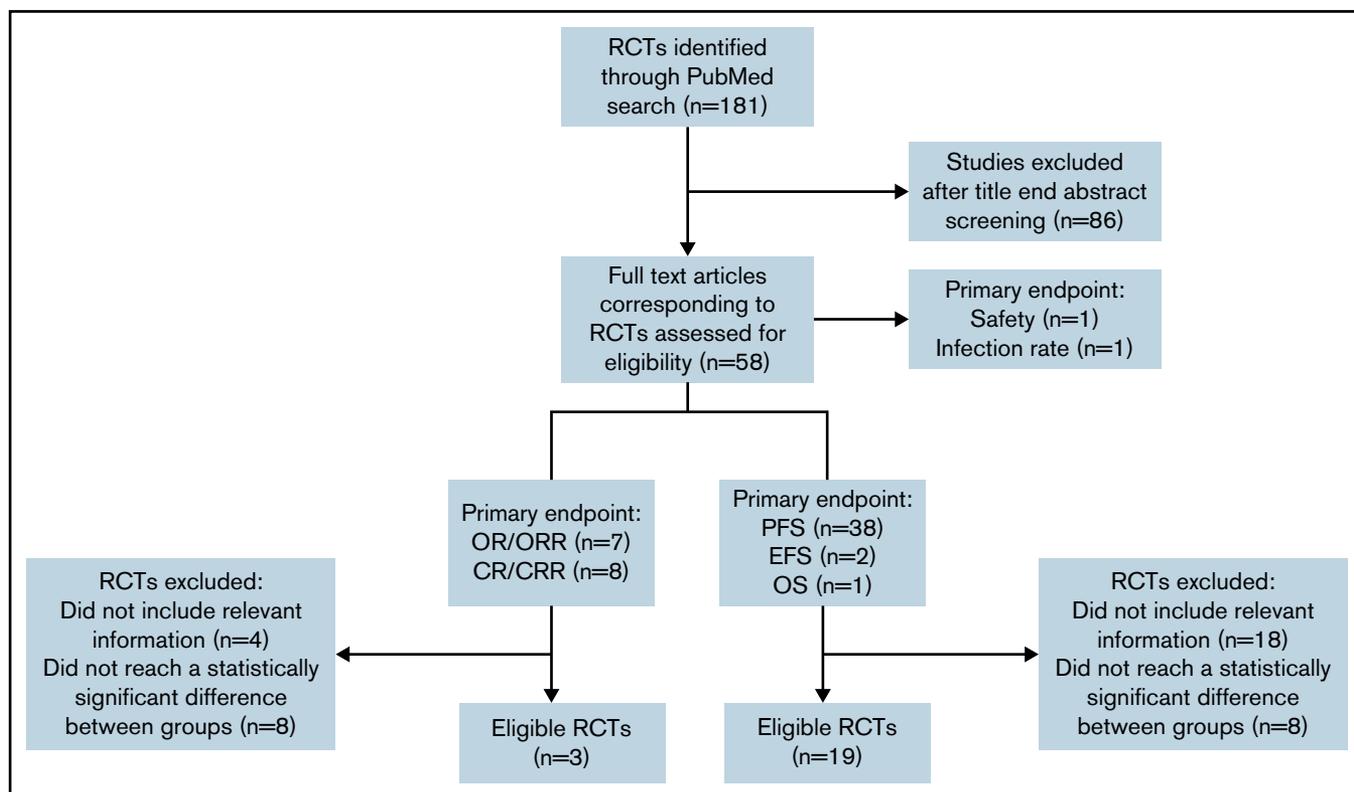
Our search for CLL RCTs for the 2010-2021 period identified 181 results, whereas only 57 results were identified when using the same search term and filtered for the 2000-2010 period, indicating an extremely prolific CLL drug development program in the last 10 years. From the 181 results, 58 RCTs were selected for further analysis in our study.

Submitted 28 October 2021; accepted 4 November 2021; prepublished online on *Blood Advances* First Edition 9 November 2021; final version published online 10 January 2022. DOI 10.1182/bloodadvances.2021006473.

Requests for data sharing may be submitted to Cristina Bagacean (cristina.bagacean@chu-brest.fr).

The full-text version of this article contains a data supplement.

© 2022 by The American Society of Hematology. Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.



**Figure 1.** Flowchart showing the selection process used to retain CLL RCTs eligible for FI calculation. CR, complete response; OR, overall response; OS, overall survival.

The 4 journals with the most CLL RCTs published were *Blood* (8 RCTs, 13.79%), *New England Journal of Medicine* (7 RCTs, 12.07%), *Lancet Hematology* (6 RCTs, 10.34%), and *Lancet Oncology* (6 RCTs, 10.34%). The median impact factor of the journals at the time of the trial publication was 10.30 (range: 2.38-74.69). A total of 17 057 patients were included, with a median

sample size of 253 patients (range: 44-817). The median age of the patients was 63 years (median age range: 54-73, age range: 22-94), and a male predominance was reported in all CLL RCTs, with an overall male-to-female ratio of 2.13 (range: 1.26-5.85). The primary endpoint evaluated in most of the trials was PFS/EFS/OS (41 RCTs, 70.69%), followed by CR/CRR (8 RCTs, 13.79%), OR/ORR (7 RCTs, 12.06%), and safety and infection rate (1 RCT, 1.72%, each).

From the 41 RCTs with PFS, EFS, and OS as primary endpoints, 19 (46.34%) met our eligibility criteria, and all of them were phase 3 trials. From the 15 RCTs with CR/CRR and OR/ORR as primary endpoints, only 3 (20%) met our eligibility criteria.

Supplemental Table 1 summarizes the characteristics of the 22 CLL RCTs included. The median FI for included RCTs was 22.50 (range: 1.00-103.00; interquartile range, 6.00-35.25), for instance, a median of 22 events was required to change the results of the endpoint analysis from significant to nonsignificant. The oncology study of Del Paggio and Tannock that used the same method for FI calculation for the RCTs that led to Food and Drug Administration approval of cancer drugs between 2014 and 2018 reported a median FI of 2.<sup>4</sup> We can also compare our results with the calculated FI for RCTs published in high-impact general medical journals, as all eligible CLL RCTs were published in such journals. The median FI calculated for RCTs published in high-impact general medical journals was 8, as reported by Walsh et al.<sup>3</sup> Therefore, compared with other RCTs, the FI and robustness of the positive CLL RCTs results seem satisfactory.

**Table 1.** Associations between trial characteristics and FI

Continuous characteristics	Correlation ( $R_s$ )	$R_s$ P
Number of patients included	0.47	.03
Median age of patients included	0.1	.65
Number of events	0.47	.02
Median follow-up	-0.41	.08
Journal impact factor	0.56	.006
Hazard ratio	-0.60	.009
Categorical characteristics	FI (IQR)	P
Publication year		.81
2010 ≤ publication year < 2015	23.5 (10.5-32.0)	
2015 ≤ publication year < 2018	11.0 (6-38.5)	
2018 ≤ publication year ≤ 2021	29.0 (20.3-34.8)	
Treatment line		.14
First line	17.5 (6-34)	
Relapsed/refractory	31 (25-51.5)	

IQR, interquartile range.

The evaluation of associations between the FI and trial characteristics revealed differences in the FI on the basis of the number of patients included (Spearman correlation [ $R_S$ ] = 0.47,  $P$  = .03), number of reported events ( $R_S$  = 0.47,  $P$  = .02), journal impact factor ( $R_S$  = 0.57,  $P$  = .006), and hazard ratio ( $R_S$  = -0.60,  $P$  = .009) (Table 1). From the 22 eligible trials, only 5 RCTs (22.73%) were academic. No statistically significant difference was revealed between the FI of the academic and pharmaceutical industry sponsored RCTs.

Our study is limited by the small sample size, which was mainly reduced by the exclusion of nonsignificant trials and of trials with missing relevant information. The operating characteristics of the FI also limit its use in time to event data: in situations where the number of events is similar between 2 groups, but a difference in timing exists, the FI might be overly sensitive in concluding fragility.<sup>4</sup>

Our results lead us to conclude that the majority of positive CLL trials from the last decade are statistically robust compared with RCTs performed in other medical fields. This evaluation is supported by the substantial changes with regard to standard-of-care therapy and the continuous increase of survival in CLL patients during this time period.<sup>8</sup> However, clinicians should remain wary of basing their decisions exclusively on a  $P$  value, as the significant results may hinge on very few events, as suggested by some of the RCTs included in our study.

**Acknowledgment:** The authors thank Thomas Marshall for editing the manuscript.

**Contribution:** V.L. and C. Bagacean designed the study, performed data collection, and wrote the manuscript; F.C. validated the accuracy of data collection when discrepancies between V.L. and C. Bagacean occurred; J.-C.I. and C. Berthou helped design the study; C. Bagacean and N.S. performed statistical analysis; all authors provided final approval of the manuscript.

**Conflict-of-interest disclosure:** The authors declare no competing financial interests.

**Correspondence:** Cristina Bagacean, CHRU de Brest, Hôpital Morvan, 2 Av Foch, 29609 Brest Cedex, France; e-mail: cristina.bagacean@chu-brest.fr.

## References

1. Ridgeon EE, Young PJ, Bellomo R, Mucchetti M, Lembo R, Landoni G. The fragility index in multicenter randomized controlled critical care trials. *Crit Care Med*. 2016;44(7):1278-1284.
2. Kahan BC, Ahmad T, Forbes G, Cro S. Public availability and adherence to prespecified statistical analysis approaches was low in published randomized trials. *J Clin Epidemiol*. 2020;128:29-34.
3. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. *J Clin Epidemiol*. 2014;67(6):622-628.
4. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis. *Lancet Oncol*. 2019;20(8):1065-1069.
5. Desnoyers A, Wilson BE, Nadler MB, Amir E. Fragility index of trials supporting approval of anti-cancer drugs in common solid tumours. *Cancer Treat Rev*. 2021;94:102167.
6. Yosifov DY, Wolf C, Stilgenbauer S, Mertens D. From biology to therapy: the CLL success story. *HemaSphere*. 2019;3(2):e175.
7. 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.
8. van der Straten L, Levin MD, Visser O, et al. Survival continues to increase in chronic lymphocytic leukaemia: a population-based analysis among 20 468 patients diagnosed in the Netherlands between 1989 and 2016. *Br J Haematol*. 2020;189(3):574-577.