



When a hazard ratio is not enough: non-proportional hazards are reshaping the interpretation of phase III trials in hepatocellular carcinoma

Erica Villa

Citation: Villa E. When a hazard ratio is not enough: non-proportional hazards are reshaping the interpretation of phase III trials in hepatocellular carcinoma. *Hepatoma Res.* 2026;12:38. <https://dx.doi.org/10.20517/2394-5079.2026.49>

Received: 22 Apr 2026

First Decision: 29 May 2026

Revised: 29 May 2026

Accepted: 15 Jun 2026

Published: 14 Jul 2026

Academic Editor:

Dar-In Tai

Copy Editor:

Ting-Ting Hu

Production Editor:

Ting-Ting Hu

In hepatocellular carcinoma (HCC), the field has become comfortable with a dangerous simplification: if the hazard ratio is below 1 and the p value crosses the prespecified boundary, the trial is read as positive; if not, it is read as negative. That shortcut was always imperfect. In the immunotherapy era, it is becoming untenable. A hazard ratio is only interpretable as a stable summary of treatment effect when hazards are proportional over time. Once that assumption is violated, the hazard ratio stops being a verdict and becomes, at best, a time-averaged compromise between biologically distinct phases of benefit and risk. Mauro and colleagues make this point with unusual clarity in HCC, and the implications are larger than one paper or one trial^[1].

Non-proportional hazards (NPH) occur when the relative hazard between treatment groups changes during follow-up rather than remaining constant over time^[2]. In practical terms, this means that the treatment effect is dynamic: survival curves may separate early and then converge, remain superimposed initially and diverge only later, or even cross. Under these circumstances, the Cox proportional hazards model is no longer a neutral summary tool. It compresses evolving treatment effects into a single number and, in doing so, risks obscuring the very feature that matters clinically: when benefit emerges, how long it persists, and whether it is offset by early harm or later attenuation. In this setting, a single global hazard ratio may be statistically convenient but clinically incomplete^[1-4]. Although the present discussion focuses on HCC, NPH are not unique to liver cancer. Delayed treatment effects, long-term survival plateaus, and crossing hazards have been extensively documented in immunotherapy trials for melanoma and non-small cell lung cancer, underscoring that this is a broader challenge in contemporary oncology rather than an isolated issue within HCC^[5,6].

This issue is no longer theoretical in HCC. In their analysis of 20 pivotal phase III randomized trials, Mauro *et al.* found evidence of NPH in four studies, corresponding to 20% of the dataset, and all four involved immunotherapy-based strategies^[1]. Three distinct patterns were identified: diminishing effects, delayed



National Institute of Gastroenterology, IRCCS "S. de Bellis" Research Hospital, Castellana Grotte 70013, Italy.

Correspondence to: Prof. Erica Villa, National Institute of Gastroenterology, IRCCS "S. de Bellis" Research Hospital, Castellana Grotte 70013, Italy. E-mail: erica.villa@irccsdebellis.it

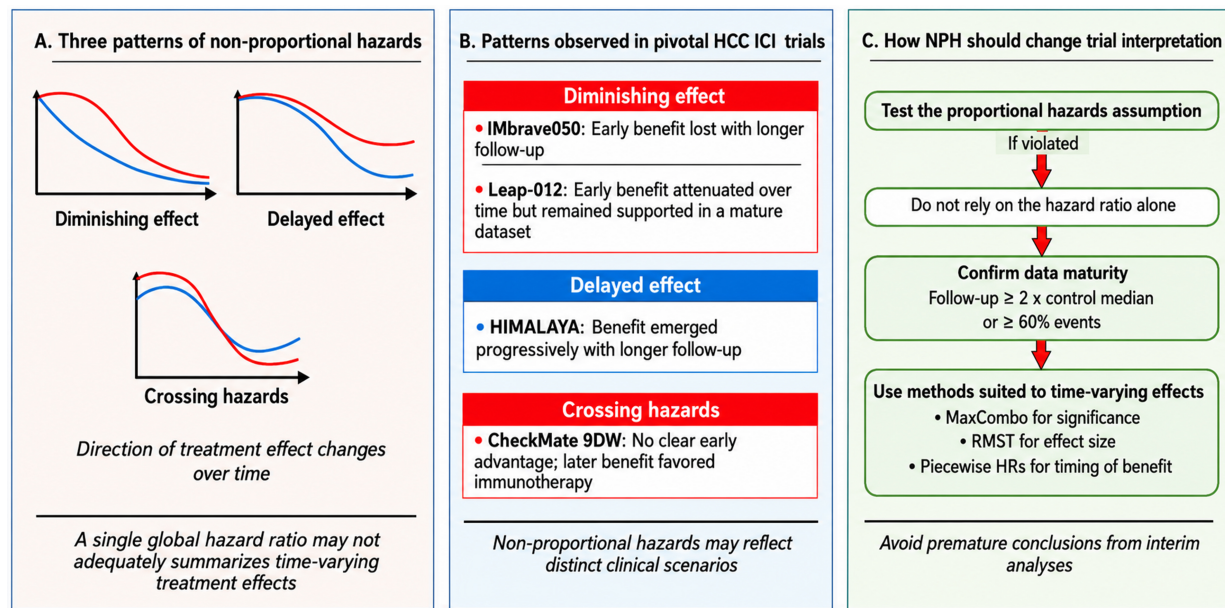


Figure 1. Non-proportional hazards in HCC trials: patterns, examples, and implications for interpretation. (A) Illustrates the three main patterns of non-proportional hazards observed in HCC trials: diminishing effect, delayed effect, and crossing hazards; (B) maps these patterns onto pivotal immunotherapy-based trials, including IMbrave050, LEAP-012, HIMALAYA, and CheckMate 9DW, highlighting the distinct clinical implications of each pattern; (C) summarizes a practical interpretive framework: test the proportional hazards assumption, avoid relying on the hazard ratio alone when NPH are present, confirm data maturity, and apply methods better suited to time-varying treatment effects, including MaxCombo testing, restricted mean survival time, and piecewise hazard ratios. Created in BioRender. Villa, E. (2026) <https://BioRender.com/fclev2z>. HCC: Hepatocellular carcinoma; HR: hazard ratio; NPH: non-proportional hazards; RMST: restricted mean survival time.

effects, and crossing hazards^[1] [Figure 1]. This observation alone should alter how the field reads time-to-event endpoints. NPH are not isolated statistical curiosities; they are recurring features of the very trials that increasingly shape standards of care in HCC. More importantly, these patterns do not carry the same clinical meaning. A diminishing effect raises different interpretive concerns from a delayed effect, and both differ fundamentally from crossing hazards. Treating all of them as though they were adequately summarized by one overall hazard ratio is no longer defensible.

The most cautionary example is IMbrave050, evaluating adjuvant atezolizumab plus bevacizumab^[7]. The interim analysis was hailed as a breakthrough: adjuvant atezolizumab plus bevacizumab improved recurrence-free survival after curative-intent treatment and was rapidly interpreted as practice changing. Yet the later analysis failed to confirm that benefit. Mauro *et al.* show that this inconsistency is best understood through the lens of NPH, specifically a diminishing treatment effect over time^[1]. The early signal was real enough, but it did not endure; with longer follow-up, the curves converged, the piecewise hazard ratio shifted from a strong early advantage to no later benefit, and statistical significance disappeared. Crucially, the trial had not met the maturity thresholds proposed for robust interpretation in the setting of NPH. Thus, the problem was not simply that the effect faded. The deeper issue was that an immature result had been interpreted as though it represented durable truth.

LEAP-012, evaluating lenvatinib plus pembrolizumab, offers an instructive contrast^[8]. It also showed a diminishing-effect pattern, but the interpretation was more robust because the data were more mature and the favorable signal remained coherent across multiple measures, including restricted mean survival time

(RMST), piecewise hazard ratios, and the MaxCombo test^[1]. This distinction is central. NPH does not automatically invalidate a trial result. What weakens interpretation is failing to recognize that the same broad NPH pattern may have very different implications depending on follow-up duration, event accrual, and the analytical tools used. IMbrave050 and LEAP-012 therefore illustrate an important principle for future HCC trial reading: when hazards are non-proportional, data maturity is not a secondary detail but part of the result itself.

The remaining two trials reinforce the same message from a different angle. HIMALAYA, evaluating the STRIDE regimen (tremelimumab plus durvalumab)^[9,10], exemplified a delayed treatment effect: there was little meaningful difference early on, followed by progressively clearer separation later, with increasing benefit over longer follow-up^[1]. This is precisely the pattern one might expect with immunotherapy, where durable survival gain may emerge slowly. An early readout in such a setting may underestimate the true clinical value of treatment. CheckMate 9DW^[11], evaluating nivolumab plus ipilimumab, was even more revealing because it showed crossing hazards: no clear early advantage, followed by later benefit in favor of the immunotherapy arm^[1]. Crossing hazards are perhaps the clearest demonstration that one global hazard ratio is not enough. A single number cannot faithfully describe a situation in which the direction of treatment effect changes over time.

These observations explain why NPH must be addressed before definitive conclusions are drawn, not rationalized after the fact. If the field ignores NPH, it risks making two opposite but equally damaging errors. The first is premature enthusiasm, when early positivity is mistaken for durable benefit, only to erode with longer follow-up. The second is premature dismissal, when a trial is judged underwhelming because the analysis occurs before a delayed treatment effect has had time to emerge. In both cases, the underlying failure is the same: interpretation is driven by the convenience of familiar statistics rather than by the temporal biology of treatment effect. In HCC, where regulatory decisions, guideline recommendations, and therapeutic sequencing increasingly depend on subtle differences in time-to-event outcomes, that is not a harmless oversight. It is a direct threat to sound clinical judgment.

The practical importance of the Mauro paper is that it moves beyond criticism and offers a practical framework^[1]. First, the proportional hazards assumption should be tested rather than taken for granted^[1,2]. Second, when NPH are identified, interim analyses should not be trusted unless the data are sufficiently mature, defined by Mauro *et al.* as follow-up of at least twice the estimated median of the primary endpoint in the control arm or at least 60% events among randomized patients^[1]. Third, when NPH are present, the field should stop relying exclusively on the log-rank test and the Cox-derived hazard ratio. Methods better suited to time-varying effects should be incorporated, notably the MaxCombo test for significance and RMST and piecewise hazard ratios for effect-size description and temporal interpretation^[1,3,5,12]. Importantly, these considerations should be incorporated prospectively, rather than introduced only after NPH have been observed. Future HCC trials, particularly those evaluating immunotherapy-based strategies, should prospectively account for the possibility of time-varying treatment effects in their statistical analysis plans. Pre-specification of proportional hazards testing, alternative significance procedures such as MaxCombo, RMST-based analyses, and clinically meaningful maturity thresholds would help reduce the risk of post hoc interpretation and improve the robustness of trial conclusions. Such an approach would align statistical methodology more closely with the biological mechanisms underlying contemporary cancer therapies.

This is not unnecessary methodological complexity. It is overdue discipline. RMST, in particular, offers an intuitive and clinically interpretable measure because it expresses benefit in units of time rather than as an abstract relative hazard^[3,4]. Piecewise hazard ratios help distinguish early from late treatment effects and can reveal attenuation, reversal, or delayed benefit that a global estimate conceals. As emphasized in broader

methodological guidance, estimands and analytical strategies should reflect the actual clinical question being asked rather than the convenience of default modeling assumptions^[12]. That principle is especially relevant in immunotherapy-based HCC trials, in which biological response patterns are often intrinsically time-dependent. In practical terms, this means that alternative analyses should be prespecified according to the expected pattern of treatment effect. The MaxCombo test is most useful when the direction of non-proportionality is uncertain, because it combines weighted log-rank components sensitive to early, constant, and late effects. RMST should be reported at clinically meaningful and prespecified time horizons, such as 12, 24, and 36 months, or at the last time point at which a sufficient proportion of patients remains at risk. In the framework proposed by Mauro *et al.*, RMST was assessed only when more than 10% of patients remained at risk, while data maturity was defined as follow-up of at least twice the expected median of the primary endpoint in the control arm or at least 60% events among randomized patients^[1]. Piecewise hazard ratios may then be used to describe the timing and direction of treatment effect across predefined intervals, for example before and after 12 months.

The larger lesson is therefore difficult to ignore. Immunotherapy has changed the biology of response in HCC, but trial interpretation has not yet fully caught up. The hazard ratio remains useful, but it can no longer be treated as a self-sufficient verdict. It is one descriptor, conditional on assumptions that are often violated in the very trials now driving therapeutic change. If hazards are not proportional, then a “positive” or “negative” label based on conventional Cox analysis may be incomplete, unstable, or simply wrong.

From a clinical perspective, the presence of NPH also has direct implications for treatment selection in HCC. Therapies characterized by delayed or time-varying effects, particularly immune-based strategies, may offer limited early benefit but more durable long-term outcomes, whereas early signals of efficacy may not be sustained in the setting of diminishing effects. These patterns suggest that treatment decisions should not rely solely on global hazard ratios or early interim results, but should instead consider the temporal profile of benefit, the maturity of the available evidence, and the clinical context of the individual patient. In particular, caution is warranted before adopting new strategies based on immature data, while therapies with delayed benefit should not be prematurely dismissed. The lesson from IMbrave050, LEAP-012, HIMALAYA, and CheckMate 9DW is not merely that NPH exist^[7-11]. It is that HCC can no longer afford to ignore them. Trials should not be stopped early, guidelines should not be rewritten too quickly, and practice should not pivot decisively until the field has first asked three questions: are hazards proportional, are the data mature, and are the methods fit for a time-varying effect? Until those questions become routine, we will remain at risk of mistaking early noise for durable truth or dismissing meaningful benefit simply because it arrives on its own timetable.

DECLARATIONS

Authors' contributions

The author contributed solely to the article.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool ChatGPT (version 5.5) was used solely for language refinement, editing, and improvement of readability. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

None.

Conflicts of interest

Villa E is an Editorial Board Member of the journal *Hepatoma Research*. Villa E was not involved in any stage of the editorial process for this manuscript, including reviewer selection, manuscript handling, or editorial decision-making.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Mauro E, de Castro T, Zeithoeffer M, et al. Strategies to address non-proportional hazards between survival curves - Lessons from phase III trials in hepatocellular carcinoma. *J Hepatol.* 2026;84:567-77. DOI PubMed PMC
2. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81:515-26. DOI
3. Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med.* 2011;30:2409-21. DOI PubMed
4. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol.* 2014;32:2380-5. DOI PubMed PMC
5. Eggermont AMM, Robert C, Ribas A. The new era of adjuvant therapies for melanoma. *Nat Rev Clin Oncol.* 2018;15:535-6. DOI PubMed
6. Pang H, Yang G, Ho JC, et al. Assessing surrogacy using restricted mean survival time ratio for overall survival in non-small cell lung cancer immunotherapy studies. *Chin Clin Oncol.* 2022;11:7. DOI PubMed
7. Yopp A, Kudo M, Chen M, et al. LBA39 updated efficacy and safety data from IMbrave050: phase III study of adjuvant atezolizumab (atezo) + bevacizumab (bev) vs active surveillance in patients (pts) with resected or ablated high-risk hepatocellular carcinoma (HCC). *Ann. Oncol.* 2024;35:S1230. DOI
8. Kudo M, Ren Z, Guo Y, et al. ; LEAP-012 investigators. Transarterial chemoembolisation combined with lenvatinib plus pembrolizumab versus dual placebo for unresectable, non-metastatic hepatocellular carcinoma (LEAP-012): a multicentre, randomised, double-blind, phase 3 study. *Lancet.* 2025;405:203-15. DOI PubMed
9. Abou-Alfa GK, Lau G, Kudo M, et al. Tremelimumab plus durvalumab in unresectable hepatocellular carcinoma. *NEJM Evid.* 2022;1:EVIDoa2100070. DOI PubMed
10. Rimassa L, Chan SL, Sangro B, et al. Five-year overall survival update from the HIMALAYA study of tremelimumab plus durvalumab in unresectable HCC. *J Hepatol.* 2025;83:899-908. DOI PubMed
11. Yau T, Galle PR, Decaens T, et al. ; CheckMate 9DW investigators. Nivolumab plus ipilimumab versus lenvatinib or sorafenib as first-line treatment for unresectable hepatocellular carcinoma (CheckMate 9DW): an open-label, randomised, phase 3 trial. *Lancet.* 2025;405:1851-64. DOI PubMed
12. International Council for Harmonisation. ICH E9(R1): Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. [Last accessed on 3 Jul 2026].

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.