



PARTNER 3 at 7 years: temporal heterogeneity, endpoint architecture, and the limits of long-term equivalence

PARTNER 3 a 7 años: heterogeneidad temporal, arquitectura de los puntos finales y los límites de la equivalencia a largo plazo

Ovidio A. García-Villarreal

Colegio Mexicano de Cirugía Cardiovascular y Torácica, A.C. Ciudad de México, México.

ABSTRACT

The 7-year follow-up of the PARTNER 3 trial provides the most robust randomized evidence comparing balloon-expandable transcatheter aortic valve implantation (TAVI) with surgical aortic valve replacement (SAVR) in patients with severe aortic stenosis at low surgical risk. Although early analyses demonstrated noninferiority of TAVI, longer-term data reveal temporal heterogeneity of treatment effects, nonproportional hazards, and interpretative dependence on composite endpoint structure. While the primary non-hierarchical composite endpoint shows no statistically significant difference at seven years, all-cause mortality is numerically higher in the TAVI group (19.5 vs. 16.8%). Hazard ratios converge and appear to invert beyond the first year, and restricted mean survival time analyses integrate early benefit with late-phase attenuation. This Point of View examines the methodological and temporal architecture of the 7-year analysis and discusses its implications for durability assessment and long-term decision-making in low-risk patients with extended life expectancy.

Keywords: aortic valve stenosis, clinical trials, TAVI, SAVR, evidence-based medicine.

RESUMEN

El seguimiento a siete años del ensayo PARTNER 3 proporciona la evidencia aleatorizada más robusta que compara la implantación de válvula aórtica transcatóter (TAVI, por sus siglas en inglés) con expansión de balón con la sustitución quirúrgica de la válvula aórtica (SAVR, por sus siglas en inglés) en pacientes con estenosis aórtica grave con bajo riesgo quirúrgico. Aunque los análisis iniciales demostraron no inferioridad de TAVI, los datos a más largo plazo revelan heterogeneidad temporal de los efectos del tratamiento, riesgos no proporcionales y dependencia interpretativa de la estructura del criterio de valoración compuesto. Si bien el criterio de valoración compuesto primario no jerárquico no muestra una diferencia estadísticamente significativa a los siete años, la mortalidad por todas las causas es numéricamente mayor en el grupo TAVI (19.5 frente a 16.8%). Las razones de riesgo convergen y parecen invertirse más allá del primer año, y los análisis de tiempo de supervivencia medio restringido integran el beneficio temprano con la atenuación en la fase tardía. Este punto de vista examina la arquitectura metodológica y temporal del análisis a siete años y discute sus implicaciones para la evaluación de la durabilidad y la toma de decisiones a largo plazo en pacientes de bajo riesgo con esperanza de vida prolongada.

Palabras clave: estenosis de la válvula aórtica, ensayos clínicos, implante de válvula aórtica transcatóter, reemplazo quirúrgico de válvula aórtica, medicina basada en evidencia.

How to cite: García-Villarreal OA. PARTNER 3 at 7 years: temporal heterogeneity, endpoint architecture, and the limits of long-term equivalence. *Cir Card Mex.* 2026; 11 (2): 54-58. <https://dx.doi.org/10.35366/122952>

©2026 by the Sociedad Mexicana de Cirugía Cardíaca, A.C.

Received: 13-02-2026. Accepted: 21-02-2026.

Correspondence: Dr. Ovidio A. García-Villarreal. **E-mail:** ovidiocardiotor@gmail.com



The PARTNER 3 trial (safety and effectiveness of the SAPIEN 3 transcatheter heart valve in low-risk patients with aortic stenosis; NCT02675114) constitutes the most robust randomized study available to date comparing transcatheter aortic valve implantation (TAVI) with conventional surgical aortic valve replacement (SAVR) in patients with severe symptomatic aortic stenosis and low surgical risk.¹

This was a prospective, multicenter, randomized clinical trial that enrolled 1,000 patients assigned in a 1:1 ratio to TAVI with a balloon-expandable valve (n = 503) or to surgery with a conventional biological prosthetic valve (n = 497). Low risk was primarily defined by an STS-PROM score < 4%, accompanied by anatomical and clinical criteria that excluded patients with significant surgical complexity. The mean age was approximately 73 years, representing a cohort with extended life expectancy within the context of degenerative valvular disease.

The primary endpoint was a non-hierarchical composite of all-cause mortality, disabling stroke, and valve-related or heart failure-related rehospitalization. In addition, a hierarchical analysis was performed using the win ratio method. Secondary outcomes included individual mortality, stroke, valve reintervention, echocardiographic parameters of function and durability, as well as complications such as atrial fibrillation, valve thrombosis, and paravalvular leak.

The results demonstrated noninferiority of TAVI compared with SAVR. The 7-year follow-up represents the longest available time horizon in this low-risk population and has been widely interpreted as evidence of sustained equivalence.² However, a careful examination of endpoint architecture, the temporal evolution of risk, and late mortality necessitates a more nuanced reading of this trial.

ENDPOINT ARCHITECTURE, ANALYTICAL FRAMING, AND THE LIMITS OF STATISTICAL NEUTRALITY

The 7-year evaluation rests on two complementary analytical strategies constructed from overlapping clinical events. The first employs a conventional non-hierarchical composite—death, any stroke, or valve- or heart failure-related rehospitalization—analyzed through time-to-first-event methods (Kaplan-Meier estimates and Cox proportional hazards models). The second applies a hierarchical composite assessed by the win ratio, which orders outcomes according to clinical priority: death, disabling stroke, nondisabling stroke, and cumulative days of rehospitalization.

While the clinical components are largely shared, the inferential architecture differs. Time-to-first-event analysis treats all first occurrences equivalently from a statistical standpoint, regardless of severity. The win ratio, in contrast,

ranks patient pairs according to predefined clinical gravity, thereby incorporating an explicit value hierarchy into the comparison. Both approaches are methodologically sound and increasingly utilized in contemporary cardiovascular trials. However, they shed light on distinct dimensions of treatment effect—event incidence versus comparative clinical ranking—and neither, by itself, resolves questions extending beyond the observed follow-up.

At seven years, the non-hierarchical composite occurred in 36.3% of patients undergoing TAVI and 34.5% of those undergoing SAVR (HR \approx 1.05; 95%CI 0.90-1.22). The range of this interval encompasses both a modest relative reduction in risk and a potentially meaningful excess hazard. Statistically, the null hypothesis cannot be rejected. Yet statistical non-significance is not synonymous with equivalence, particularly in a study not designed or powered to establish noninferiority at this temporal horizon. As follow-up continues and the accumulation of events stabilizes, diminished power heightens susceptibility to type II error;^{3,4} neutrality under such conditions means indeterminacy rather than interchangeability.

Interpretation is further complicated by the structure of composite endpoints.⁵ Even when hierarchically ordered, the aggregation of outcomes with different prognostic and ethical weight generates conceptual conflicts. Death, disabling stroke, nondisabling stroke, and rehospitalization differ fundamentally in their implications for survival, autonomy, and long-term quality of life. Statistical frameworks can reorder or group these events, but they cannot erase their qualitative asymmetry.

For younger patients with low operative risk and substantial life expectancy, the pivotal issue is not merely the absence of demonstrable difference, but whether the available evidence justifies declaring two distinct interventions therapeutically interchangeable over the long term. The 7-year findings do not provide that level of certainty. Rather, they delineate a rigorously analyzed yet intrinsically uncertain space in which methodological sophistication coexists with unresolved clinical doubt.

TEMPORAL HETEROGENEITY, NONPROPORTIONAL HAZARDS, AND THE LIMITS OF INTEGRATED METRICS

The 7-year analysis explicitly acknowledges evidence of nonproportional hazards for key mortality-related outcomes, indicating that the assumption of a constant hazard ratio over time is not fully sustained. This is a critical interpretative issue, not a minor statistical detail. When proportional hazards are violated, a single Cox-derived hazard ratio no longer represents a stable treatment effect; instead, it becomes a time-averaged summary across different risk phases.

In the context of PARTNER 3, the risk pattern likely includes an early procedural phase favoring TAVI, an intermediate period where risks converge, and a later phase where the relative hazard may decrease or even reverse. The survival curves support this pattern: they separate early, then converge, with hazard ratios beyond the first year appearing less favorable to TAVI and diverging over time. This time-dependent pattern is plausible in structural valve interventions, possibly due to differences in valve durability, leaflet integrity, hemodynamic performance, patient-prosthesis interaction, and device-related issues that accumulate over time.

Methodologically, the investigators respond appropriately to this temporal heterogeneity. It is well established that the two principal approaches developed to address departures from the proportional hazards assumption are weighted log-rank procedures and analyses based on restricted mean survival time (RMST).^{6,7} By incorporating RMST for both the primary composite end point and all-cause mortality, the authors adopt a framework that does not rely on proportional hazards and instead estimates the average event-free survival up to a prespecified time horizon (2,555 days) by integrating the area under the survival curve. There is therefore no statistical inconsistency in presenting hazard ratios alongside RMST estimates; the latter serves as a robustness measure when treatment effects vary over time.

However, RMST, by design, integrates early benefit and later attenuation into a single cumulative metric. In this analysis, the primary composite endpoint modestly favors TAVI, with approximately 134 additional event-free days, whereas the difference for all-cause mortality is small (approximately -15 days) and not statistically significant. While RMST provides a stable and assumption-free average measure, it does not describe the directional evolution of hazard beyond the restriction horizon, nor does it distinguish between front-loaded procedural advantage and late-phase risk convergence. In interventions where durability is a concern, this integration may hide emerging late differences. As a result, when hazards are not proportional, long-term equivalence cannot be assumed based on overall, time-averaged data alone. The clinical question shifts from whether average survival differs to how risk unfolds across time—a distinction that is not statistical nuance, but a determinant of durable therapeutic inference.

LATE MORTALITY: A SIGNAL THAT WARRANTS ATTENTION

At seven years, all-cause mortality was 19.5% in the TAVI group compared with 16.8% in the SAVR group. The difference was not statistically significant, yet it remains numerically higher. As in PARTNER 3 trial, in a cohort

with a mean age of 73 years and a potential life expectancy exceeding a decade, modest late divergences may carry greater strategic relevance than early procedural advantages. Over extended horizons, small absolute differences can translate into clinically meaningful trajectories.

The manuscript narrative suggests that early benefits may be counterbalanced by later events. However, the underlying mechanisms that could account for such attenuation—structural durability, persistent paravalvular leak, subclinical valve thrombosis, or prolonged hemodynamic interaction—are not causally explored.

Statistical non-significance does not exempt from the obligation of continuous clinical monitoring.

GUIDELINE-DEFINED LOW RISK VERSUS TRIAL-SELECTED LOW RISK: ALIGNMENT AND TENSION

Although PARTNER 3 formally aligns with the 2020 ACC/AHA definition of low surgical risk—requiring a low STS-PROM score, absence of frailty, no major end-organ dysfunction, and no procedure-specific impediments—the trial applies this definition in an exceptionally stringent manner.⁸ The enrolled population represents not merely “low risk” in the pragmatic clinical sense, but a carefully selected group optimized for dual eligibility: anatomically suitable for transfemoral TAVI, free of bicuspid morphology, devoid of complex coronary disease, and largely spared significant renal, pulmonary, neurologic, or hepatic dysfunction. In doing so, the trial achieves high internal validity and methodological clarity.

However, this rigor simultaneously narrows its external applicability. In routine practice, many patients considered low surgical risk under guideline frameworks present with gradations of comorbidity, coronary complexity, or anatomical variation that were systematically excluded from PARTNER 3.

Thus, the study does not so much misclassify risk as idealize it. Importantly, even within this optimized and physiologically favorable cohort, long-term equivalence remains statistically neutral and clinically nuanced, with late mortality numerically favoring surgery and evidence of non-proportional hazards over time. If uncertainty persists under idealized conditions, extrapolation to broader, more heterogeneous low-risk populations warrants measured restraint rather than conceptual substitution of one therapy for the other.

Therefore, the critical issue is not patient selection per se, but extrapolation. The concern emerges when an ultra-selected experimental phenotype is implicitly transformed into a mandate for broad clinical expansion. Guidelines may define “low surgical risk” using structured and stringent criteria; however, in real-world practice, that designation often

shifts in meaning over time. The term expands to encompass patients with increasing anatomical complexity, incremental comorbidity, or longer life expectancy than those represented in the trial population.

At that point, the basis of the evidence as evidentiary foundation becomes less directly transferable. What was demonstrated under tightly controlled, idealized conditions risks being generalized to a wider and more heterogeneous population in whom the balance between early procedural advantage and long-term durability may differ substantially. The distinction between internal validity and clinical universality is subtle but decisive – and it is exactly here that caution becomes not about being overly careful, but about being methodologically responsible.

If uncertainty persists under idealized experimental conditions, its extrapolation to broader and more heterogeneous low-risk populations demands prudence.

DURABILITY AND CONCEPTUAL REDEFINITION

Durability assessment is grounded in VARC-3 criteria.⁹ Under these standardized definitions, TAVI and SAVR appear broadly comparable. However, several considerations complicate this apparent symmetry:

1. Clinically significant valve thrombosis is more frequent following TAVI (2.8 vs. 0.5%; HR 5.7; 95%CI 1.29-25.25).
2. Paravalvular leak persists in a non-negligible proportion of TAVI recipients (17.7 vs. 2.2%).
3. A 7-year horizon does not approximate the biological lifespan of a patient aged 65-70 years.

Durability cannot be reduced to the absence of reintervention:¹⁰ it entails sustained physiological integration over time.

DURATION OF FOLLOW-UP AND CONDITIONAL LIFE EXPECTANCY

According to the National Center for Health Statistics, life expectancy at birth in the United States in 2025 is 78.4 years overall (75.8 years for men and 81.1 years for women).^{11,12} However, this metric is not the relevant parameter when analyzing a cohort with a mean age of 73 years, as in PARTNER 3. At that age, the appropriate reference is conditional life expectancy: a 73-year-old man has an average additional life expectancy of 10.95 years, and a woman 13.28 years; even at age 65, the additional expectancy is 16.06 and 19.06 years, respectively.¹³ These figures imply

that a substantial proportion of patients enrolled in the trial –carefully selected, low surgical risk, and without significant frailty– are likely to survive well beyond the first decade following valve implantation.

Given that structural valve degeneration in biological prosthetic aortic valves typically becomes clinically meaningful after 8-10 years and may accelerate thereafter, follow-up at one, five, or even seven years primarily captures procedural performance and intermediate outcomes rather than true long-term durability. In this context, 7-year data cannot reasonably be regarded as definitive for informing therapeutic decisions in low-risk patients whose conditional life expectancy substantially exceeds that time horizon.

WHAT THE STUDY DOES WELL

Critical appraisal does not entail dismissing methodological rigor. PARTNER 3 stands as one of the most robustly designed trials in TAVI: randomized, multicenter, supported by independent event adjudication, and grounded in the rigorous application of VARC-3 definitions. It has demonstrated that expansion into low-risk populations has not resulted in early clinical failure.

However, the absence of early collapse is not synonymous with superiority, nor does it establish full therapeutic interchangeability with SAVR.

FINAL COMMENTARY

The 7-year follow-up of PARTNER 3 illustrates how temporal dynamics and endpoint architecture shape long-term interpretation.

Key observations include:

1. Violation of proportional hazards assumptions.
2. Numerical late-phase mortality divergence.
3. Dependence of composite outcomes on rehospitalization components.
4. Analytical reliance on hierarchical win ratio and RMST integration.

These findings underscore the necessity of cautious extrapolation when translating early procedural benefits into long-term therapeutic preference.

In structural heart disease, time is not merely a covariate; it is the decisive variable. In patients with substantial life expectancy, as in low-risk patients, the essential question is not whether the statistical average appears neutral, but in which direction risk ultimately evolves.

At seven years, that answer remains unsettled.

REFERENCES

1. ClinicalTrials.gov. PARTNER 3 trial: safety and effectiveness of the SAPIEN 3 transcatheter heart valve in low risk patients with aortic stenosis (P3) [Internet]. Bethesda (MD): National Library of Medicine (US); 2026. Available in: <https://clinicaltrials.gov/study/NCT02675114>
2. Leon MB, Mack MJ, Pibarot P, et al; PARTNER 3 Investigators. Transcatheter or surgical aortic-valve replacement in low-risk patients at 7 years. *N Engl J Med*. 2026;394(8):773-783. doi: 10.1056/NEJMoa2509766.
3. Domb BG, Sabetian PW. The blight of the type ii error: when no difference does not mean no difference. *Arthroscopy*. 2021;37(4):1353-1356. doi: 10.1016/j.arthro.2021.01.057.
4. Bowers D. Medical statistics from scratch: An introduction for health professionals. 4th ed. Nashville, TN: John Wiley & Sons; 2019.
5. McCoy CE. Understanding the use of composite endpoints in clinical trials. *West J Emerg Med*. 2018;19(4):631-634. doi: 10.5811/westjem.2018.4.38383.
6. Han K, Jung I. Restricted mean survival time for survival analysis: a quick guide for clinical researchers. *Korean J Radiol*. 2022;23(5):495-499. doi: 10.3348/kjr.2022.0061.
7. Faizi N, Alvi Y. Survival analysis. In: Faizi N, Alvi Y, editors. *Biostatistics manual for health research*. Amsterdam: Elsevier; 2023. p. 195-211.
8. Otto CM, Nishimura RA, Bonow RO, et al; Writing Committee Members. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2021;77(4):e25-e197. doi: 10.1016/j.jacc.2020.11.018.
9. Généreux P, Piazza N, Alu MC, et al; VARC-3 Writing Committee. Valve Academic Research Consortium 3: updated endpoint definitions for aortic valve clinical research. *Eur Heart J*. 2021;42(19):1825-1857. doi: 10.1093/eurheartj/ehaa799.
10. García-Villarreal OA. Reoperation rate versus failure rate as quality indicators in transcatheter edge-to-edge repair for mitral regurgitation. *Am J Cardiol*. 2024;231:70-71. doi: 10.1016/j.amjcard.2024.08.036.
11. Centers for Disease Control and Prevention. Life expectancy [Internet]. Atlanta (GA): CDC; 2026. Available in: <https://www.cdc.gov/nchs/fastats/life-expectancy.htm>
12. Schneider L. US life expectancy is rebounding. *JAMA*. 2025;334(10):849. doi: 10.1001/jama.2025.10994.
13. New York State Department of Health. Life expectancy table [Internet]. Albany (NY): NYSDOH; [cited 2026 Feb 12]. Available in: https://www.health.ny.gov/health_care/medicaid/publications/docs/adm/06adm-5att8.pdf

Funding: none.

Disclosure: the author has no conflict of interest to disclose.