

Title page

Mortality in trials on trans-catheter aortic valve implantation vs surgical aortic valve replacement: a pooled analysis of Kaplan-Meier derived individual patient data

Authors: **Fabio Barili**¹, M.D. Ph.D. F.E.S.C., **Nick Freemante**², PhD., **Alberto Piloizzi Casado**¹, M.D., **Mauro Rinaldi**³, M.D. Ph.D., **Thierry Folliguet**⁴, M.D., **Francesco Musumeci**⁵, M.D., **Gino Gerosa**⁶, M.D., **Alessandro Parolari**⁷, M.D. Ph.D.

Institutions:

¹ Department of Cardiac Surgery, S. Croce Hospital, Cuneo, Italy

² Institute of Clinical Trials and Methodology, University College London, London, UK.

³ Department of Cardiac surgery, AOU "Città della Salute e della Scienza di Torino", University of Turin, Turin, Italy

⁴ Department of Cardiac Surgery, Hôpital Henri Mondor, Paris, France.

⁵ Department of Heart and Vessels, Cardiac Surgery Unit and Heart Transplantation Center, S. Camillo-Forlanini Hospital, Rome, Italy.

⁶ Department of Cardiac Surgery, University of Padua, Padua, Italy.

⁷ Unit of Cardiac Surgery and Translational Research, IRCCS Policlinico S. Donato, Italy

Corresponding Author: **Fabio Barili, M.D., PhD, F.E.S.C.**

Department of Cardiac Surgery, S. Croce Hospital

Via M. Coppino 26, 12100 Cuneo, Italy

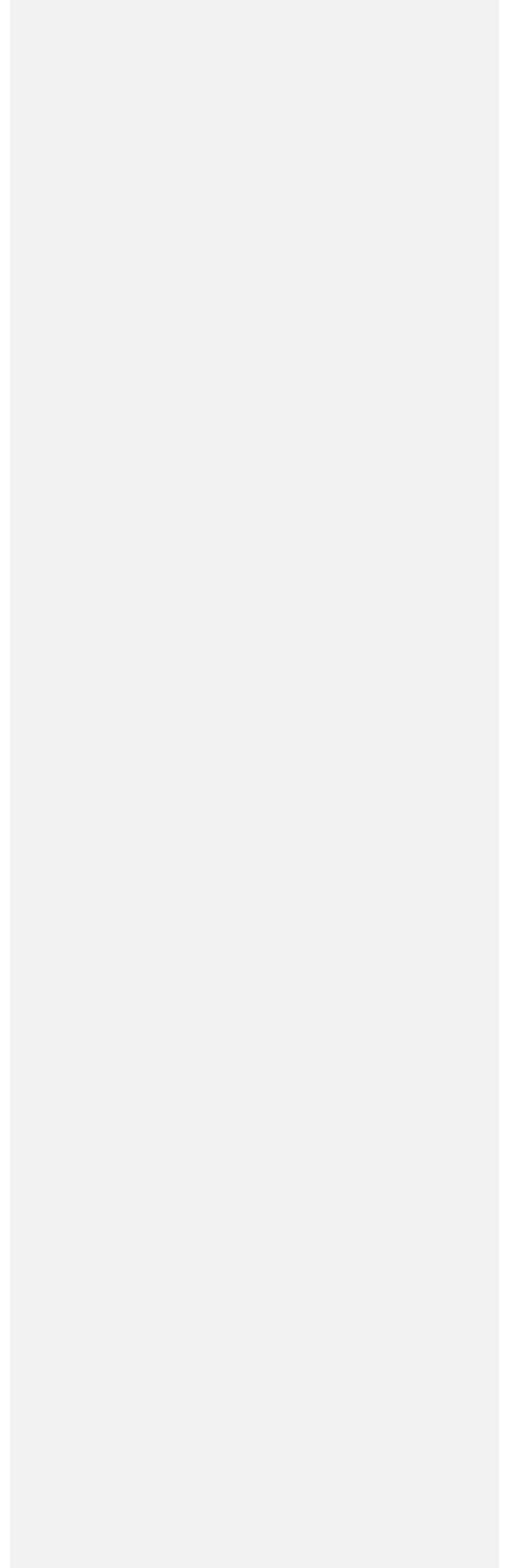
Tel: +39 0171642571

Fax: +39 0171642064

Email: fabarili@libero.it

barili.f@ospedale.cuneo.it

Word count of text: xxx



ABSTRACT

Aims. We designed a pooled analysis of Kaplan-Meier-estimated individual patient data from trials comparing TAVI and SAVR to evaluate their effects on long-term all-cause mortality, to examine the potential time-varying effect and to model their Hazard Ratio over time.

Methods. Trials comparing TAVI vs. SAVR were identified through Medline, Embase, and Cochrane databases. The primary outcome was death from any cause at follow-up. Enhanced secondary analysis of survival curves was performed reconstructing the data from published Kaplan-Meier curves as described by Guyton and Colleagues [5]. Comparison between reconstructed individual patient time-to-event data was performed with Kaplan-Kaplan Meier estimates and grouped frailty Cox model. Landmark analysis was employed in order to overcome the proportional hazards assumption failure.

Results. We identified six eligible trials including 6367 participants, who were randomly assigned to undergo TAVI (3252 patients) or SAVR (3115 patients). The adjusted HR for all-cause mortality estimated with grouped frailty Cox model was 0.94 (95% CI 0.85 -1.06, p-value 0.27). However, the assumption of hazard-proportionality was not fulfilled and landmark analysis was performed. In the first year after implantation, incidence of mortality was significantly lower in TAVI group (Log-rank p-value <0.001) and the HR adjusted by risk profile was 0.85 (95%CI 0.73 – 0.99, p-value 0.003). No difference in all-cause mortality incidence between TAVI and SAVR was evident in the timeframe 12 - 40 months (adjusted HR 0.93, 95%CI 0.77 – 1.12; p-value 0.43). Landmark analysis of all-cause mortality after 40 months demonstrated a reversal of HR (HR 1.30; 95%CI 1.01-1.67; p-value 0.04) favoring SAVR on TAVI.

Conclusions. Mortality in trials on TAVI vs SAVR is affected by treatments with a time-varying effect. TAVI is related to better survival in the first months after implantation while, after 40 months, it was associated with an increased risk of death.

Commented [NF1]: If there is a departure from Constancy then the Cox model is biased so you should not report it. The p value from a log rank test is however sound and fine to use that.

TEXT

Introduction.

Transcatheter aortic valve implantation (TAVI) has been recognized as an appropriate option for the treatment of high and intermediate-risk patients with symptomatic aortic valve stenosis and recent guidelines have established that the choice between TAVI and surgical aortic valve replacement (SAVR) in those risk categories should be assessed by the interdisciplinary Heart Team, based on patient's characteristics and comorbidities. The demonstration of safety and feasibility of the procedure as well as good perioperative and 2-year results have led to broad the indication of implantation also on low-risk profile and the recent results at 1-year in the PARTNER 3 and EVOLUT low risk cohorts confirmed at least the non-inferiority of TAVI previously shown in other risk categories.

All-cause mortality is not only the main quality index but also a standard outcome among trials comparing TAVI and SAVR. However, to date, all trials are individually underpowered to evaluate all-cause mortality, being designed to examine composite primary outcomes. Moreover, existing meta-analyses give limited information as they are focused on fixed time-point, such as 30 days or 1 year, and cannot intercept the potential effect of the treatments on longer expectancy of life in intermediate or low risk profiles. Also the use of hazard ratio as effect size in meta-analysis should be critically evaluated, as the necessary assumption of constant proportional hazards cannot be assessed and visual inspection of the Kaplan-Meier curves can suggest a time-varying HR of TAVI versus SAVR. The use of landmark analysis in PARTNER 1A can indirectly confirm that constant proportional hazards assumption is not valid at least at longer follow-up.

In order to overcome these limitations, we designed a pooled analysis of Kaplan-Meier-estimated individual patient data (IPD) from trials comparing TAVI and SAVR to evaluate their effects on long-term all-cause mortality, to examine the potential time-varying effect and to model the Hazard Ratio over time.

Materials and Methods

Data Sources and searches.

We performed a systematic review of the literature to identify eligible studies published between January 1st, 2007 and June 30th, 2019. The search was done by two independent researchers on July 5-15, employing Medical Subject Headings (MeSH) and free-text terms for identifying relevant references. The electronic databases employed for search were MEDLINE, Embase, the Cochrane Central Register of Controlled Trials (CENTRAL). The search algorithm is detailed in [Table E1](#). In addition, we also checked websites (www.clinicaltrials.gov, www.acc.org, www.cardiosource.com, www.escardio.org, www.tctmd.com) for unpublished data.

Inclusion and exclusion criteria for study eligibility.

We included randomized trials with random allocation to TAVI or SAVR that reported at least 1-year follow-up all-cause mortality and that graphed Kaplan-Meier curves of all-cause-mortality in Text or Appendix or whose survival curves was presented in selected international meetings. We excluded trials that compared different TAVI devices of different SAVR devices, trials comparing aortic valve prosthesis with medical therapy, trials that analyzed perioperative outcomes and trials that did not reported Kaplan-Meier curves of all-cause mortality.

Outcomes.

The outcome of the meta-analysis was death from any cause at follow-up (at least 1-year follow-up) with hazard ratio (HR) as effect size. The longest available follow-up report was selected for each included trial. The number of events in the two arms were extracted from text and employed to estimate Kaplan Meier-derived IPD data. [Hazards Ratios were estimated from Kaplan Meier curves-derived Individual Patient Data \(KMd-IPD\)](#) with Cox-semi parametric model and fully parametric models [and compared with those reported](#)

(when available) in the text. We pooled data from intention-to-treat (ITT) populations, choosing data from as-treated population when ITT data were not available.

Data extraction and data analysis.

Data extraction was performed following the PRISMA guidelines for systematic reviews and meta-analysis. Two independent investigators (A.P.C. and F.B.) performed literature search and identified trials that fulfilled pre-specified inclusion criteria. Eligible trials were then reviewed in duplicate and disagreement was addressed by a third investigator (A.P.). Extracted data from Text and Appendix were: Trial name, year of publication, number of participated centers, recruitment period, maximum available follow-up, trial design, intention-to-treat and as treat groups number, age, sex, Society of Thoracic Surgeons Predicted Risk Of Mortality (STS) risk, chronic kidney disease stage 4 or 5, peripheral vascular disease, prior cerebrovascular event, prior coronary artery bypass graft, prior percutaneous coronary intervention, known atrial fibrillation (AF) or flutter, prior pacemaker, transcatheter heart valve system, number of patients treated with transfemoral and transthoracic approach, associated percutaneous cardiac intervention (PCI) in the TAVI group, associated coronary artery bypass grafting (CABG) in the SAVR group, associated procedures in the SAVR group.

Individual patient data (IPD) estimated from the Kaplan-Meier curves were the event (all-cause mortality 1, censored 0) and time-to-all cause mortality (months). Data extraction from each available Kaplan-Meier curve was performed as described by Guyot and Colleagues [13]. In the first step, KM curves reproduced in each paper were digitized using a dedicated software (Plot Digitized 2.6.2 for Macintosh). The pdf. files were read into the software, the axes defined and mouse-clicks used to select points to read off from the curve, resulting in text file with KM coordinates. The KM data reconstruction algorithm developed in R language (R 3.6.0; R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>) was employed to derive the individual patient data. Derived KM curves were graphically checked with the original ones and the same comparisons of the original studies were performed. KM data from different studies were stored together in the study database.

Risk of bias and quality assessment /certainty of evidence.

The Cochrane risk of bias assessment tool was employed to evaluate the risk of bias among included trials by two reviewers (A.P.C. and F.B.). The evaluated items were allocation sequence generation, allocation concealment, blinding of participants and investigators, completeness of outcome data, and selective outcome reporting. Blinding was considered adequate if outcomes assessors were blinded, as the nature of the treatment does not makes blinding of patients and physician relevant.

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was employed to grade the certainty of evidence (very low, low, moderate, high).

Statistical analysis.

Continuous variables were presented as mean and standard deviation, categorical data were presented as number and percentage.

Kaplan Meier-derived IPD data were pooled together. Statistical methods for time-to-event data were employed to analyze all-cause mortality at follow-up. Kaplan-Meier estimator and log-rank test were employed to estimate and compare the unadjusted incidence of all-cause mortality at follow-up in the two treatment arms. Grouped frailty semi-parametric (Cox) model was employed to estimate HRs in the pooled dataset and for the single trials, accounting for heterogeneity among trials with a random-intercept parameter. Grouped frailty Cox models was also stratified by risk-profile (defined as high, intermediate and low). Proportionality of hazards of the Cox models was checked with the Grambsch-Therneau test and diagnostic plot based on Shoenfeld residuals [22]. Landmark analyses were planned in where visual inspection of Kaplan-Meier curves suggested a time-dependent variance in the HR of TAVI vs SAVR violation regardless of the results of the diagnostic test for constancy.

Time-dependency of covariates was approached with landmark analysis, applying Kaplan-Meier analysis and Cox regression to evaluate survival in the treatment groups (TAVI/SAVR) at time-frames. The cutoffs were chosen on the basis of visual inspection of the scaled Shoenfeld residuals and of the Kaplan Meier curves. Moreover, the time-varying Hazard Ratio for death of TAVI vs SAVR was modeled with fully

parametric generalized survival models (Royston-Parmar models) with baseline smoother and time-varying variables based on b-splines.

Quality assessment of Kaplan-Meier derived IPD data was performed graphically checking the derived Kaplan-Meier curves with the original ones. Moreover, the accuracy was evaluated comparing the estimated and reported (when available) HRs.

The extent of heterogeneity among studies was also assessed performing a random-effect meta-analysis with KM-derived HRs as effect size.

Analyses were performed with R language (R 3.6.0; R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>).

We adhere to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement (PRISMA checklist, Supplemental material).

Results

Trials characteristic and risk of bias.

The literature search identified xxx citations that were evaluated for eligibility in title and abstract and, after excluding duplicates, 8 trials were checked for further assessment. The STACCATO trial was excluded as no data on follow-up longer than 30-day were reported. The EVOLUT Low Risk trial was excluded lacking Kaplan-Meier curves of all-cause mortality for data extraction. The NOTION trial was included as the Kaplan Meier curve of all-cause mortality was reported in a presentation at ACC meeting 2009. Six trials (PARTNER 1A, PARTNER 2A, PARTNER 3A, NOTION, US CoreValve High Risk and SURTAVI) fulfilled the pre-specified inclusion criteria and were included in the meta-analysis (Figure 1).

Table 1 summarized trials' design and baseline characteristics of the study groups. The six eligible studies were multicenter randomized trials and the longer available follow-up was published between 2015 and 2019. The shortest follow-up period was 1-year whereas the longest was 5 years. Valve Academic Research Consortium (VARC) or VARC-2 definitions were applied in all the included studies.

The risk of bias was rated as low in all the trials as assessed by the Cochrane Collaboration's tool. Risk of bias assessments for all the included trials are detailed in Supplementary material (Table E2).

Commented [NF2]: This is very surprising with the level of loss to follow up (and the differential nature of this) plus the use of as treated analyses as the primary analyses etc.

Patients and procedural characteristics.

Overall, 6367 patients were randomly assigned to undergo TAVI (n=3252) or SAVR (n=3115). Two RCTs included patients considered at high risk, two at intermediate risk and two at low-risk. The mean age of the study group ranged between 73 and 84 and 44.1% were female. Baseline characteristics are detailed in Table 1.

In the 6 trials, both balloon-expanding (Edwards SAPIEN, SAPIEN XT and SAPIEN 3) and self-expanding TAVI devices (Medtronic CoreValve) were under study. The TAVI approaches were varied, however the most common access was transfemoral.

All the eligible trials were funded by valve manufactures.

Quality assessment of estimated IPD data.

Visual comparison between original reported Kaplan-Meier curves and estimated KM curves demonstrated no major graphical differences. Hazards Ratios estimated from KMD-IPD were compared to HRs in the paper, when available. NOTION trial and SURTAVI Trial did not calculate TAVI vs SAVR HRs, while comparison between reported and estimated HRs was possible for PARTNER 1A, PARTNER 2A, PARTNER 3 and CoreValve US Pivotal Trials. As shown in Figure E1, HRs estimated from Kaplan Meier-derived IPD data were not different to those reported in the trials, confirming a high accuracy of the IPD-deriving method.

Analysis of Heterogeneity among trials.

The analysis of heterogeneity for HR of all-cause mortality for TAVI vs SAVR identified no observed heterogeneity (Figure E2), p-value=0.68, and the percentage of heterogeneity on total variability (I^2) of 0%, suggesting the observed variability in study estimates may be due to chance.

Analysis of all-cause mortality at follow-up.

Crude mortality rates at 5 years for TAVI versus SAVR were 22.5% vs 21.9% . The Kaplan-Meier estimates of survival at 2 and 5 years were respectively $82.9\% \pm 0.5\%$ and $51.0\% \pm 1.3\%$.

Figure 2 shows the Kaplan-Meier estimates for the incidence of all-cause mortality, based on estimated 146448 patient-months follow-up (median follow-up 24 months). The difference between TAVI and SAVR curves was not significant (Log-rank p-value 0.3). The unadjusted HR for all-cause mortality estimated with Cox semiparametric model was 0.95 (95% CI 0.86 – 1.01, p-value 0.3) and it was confirmed after stratification for risk profile (HR 0.95, 95% CI 0.85 -1.05, p-value 0.33). The random effect's variance, which represents the parameter for heterogeneity, was significant ($\theta = 0.11$, p-value < 0.001)

However, the assumption of hazard-proportionality was not fulfilled based on the analysis of Shoenfeld residuals (Figure E4) and the Grambsch-Therneau test for time-invariant effect (p-value 0.003); hence models accounting for time-varying HRs were needed.

Commented [NF3]: Just make the point that the constancy assumption was not met, and do not report the HR.

Landmark analysis of all-cause mortality.

The cutoffs selected by visual inspection of the scaled Schoenfeld residuals and the Kaplan-Meier curve were 12 and 40 months. Hence Kaplan-Meier curves and Cox-derived HRs were estimated at 3 timeframes, 0-12 months, 12-40 months and over 40 months.

Figure 3 shows the Kaplan-Meier estimates of all-cause mortality by landmark analysis. In the first year after implantation, incidence of mortality was significantly lower in TAVI group (Log-rank p-value <0.001) and the HR of TAVI vs SAVR stratified by risk profile was 0.85 (95% CI 0.74 – 0.99, p-value 0.004). The estimate of random parameter for heterogeneity was significant ($\theta = 0.16$, p-value < 0.001). The assumption of hazard-proportionality was violated (Grambsch-Therneau test p-values 0.003). The proportionality hazard assumption was respected modeling the time-varying effect of TAVI/SAVR with restricted cubic splines with 4 knots (coefficients of the Cox model: -0.36 ± 0.08 , $p < 0.001$; 2.44 ± 0.63 , $p < 0.001$; -4.37 ± 1.25 , $p < 0.001$. Grambsch-Therneau test p-values 0.88), confirming that in the first year the HR of TAVI/SAVR cannot be considered constant and widely varies over time.

No difference in all-cause mortality incidence between TAVI and SAVR was evident in the timeframe 12 - 40 months (Risk-stratified HR 0.93, 95%CI 0.77 – 1.12; p-value 0.43); the proportional hazards assumption in the Cox model was not violated as demonstrated by the Grambsch-Therneau test (p-value 0.38). The estimate of random parameter for heterogeneity was not significant ($\theta = 0.001$, p-value 0.34).

Landmark analysis of all-cause mortality after 40 months demonstrated a reversal of HR (Risk-stratified HR 1.30; 95%CI 1.01-1.68; p-value 0.04) favoring SAVR on TAVI. The estimate of random parameter for heterogeneity was not significant ($\theta = 0.024$, p-value 0.08). Also in this timeframe, the proportional hazards assumption of the Cox model was not violated (Grambsch-Therneau test p-values 0.66).

Evaluation of HR trend over time for TAVI vs SAVR

In order to overcome limitations related to the proportional hazards assumption and estimate the trend over time of HR for all-cause mortality of TAVR vs SAVR, we also employed a fully parametric model.

Figure 4 shows the HR trend over time of TAVI vs SAVR estimated by fully parametric generalized survival models. The graph confirms the lack of constant HR over time. As hypothesized by visual inspection of the Kaplan-Meier curve, in the first months after implantation TAVI demonstrated a significant better survival than surgery. After reaching a nadir in the first month, this advantage decreased over time. There was no advantage of one treatment between 12 and 40 months, although a non-significant trend over an inversion of HR is evident. After 40 months, HR favors SAVR over TAVI, which is associated with an increased risk factor for all-cause mortality, in line with the landmark analysis.

Discussion

The main outcome of this meta-analysis of Kaplan-Meier derived individual patient data from clinical randomized trial comparing TAVI vs SAVR is that the statistically significant survival advantage associated to TAVI is only limited to the first year after implantation, while TAVI demonstrated a significant worse survival compared to surgery after 40 months, independently on risk profile, although this assessment is limited by the lack of individualised risk information available from these digitised data analyses. The evidence for a reversal of HR after 40 months favouring SAVR leads to some key considerations on both potential wide extension of indication to TAVI and limitations of existing meta-analyses of published data.

The shift of risk category toward low-risk profile is requesting a critical appraisal of outcomes that should be evaluated in TAVI vs SAVR trials and the time-span needed, as low-risk individuals have a higher expectancy/quality of life to be taken into account and new devices should at least not negatively affect them compared to the gold standard. An aortic valve prosthesis does not only relieve stenosis but also adds a question mark on durability over-time, which becomes the very new outcome after safety and effectiveness. Health technology assessment is focused on device durability and guidelines recommend the evaluation of structural valve deterioration and reoperation as a main issue. In high-risk profile, short and mid-term results can likely cover the expectancy of life, while the focus is necessarily moving toward follow-up longer than 5 years for younger patients at low-risk. However, at our knowledge, no follow-up greater than 5 -years is still available for TAVI devices.

Within this visual space limited by the 5-years horizon line, our meta-analysis adds a novel world of caution to the wide extension of the indications to TAVI. The demonstration of a time-varying effect of TAVI on all-cause mortality with a significant survival disadvantage after 40 months adds a new key message to the single trials that are individually underpowered to compare all-cause mortality but also contradicts the survival advantage associated to TAVI over two year follow-up recently underscored in meta-analyses. TAVI confirmed to have a better survival in the first year, reaching a HR = 0.6 in the first month and decreasing its positive effect in the next months, while no TAVI advantage has been demonstrated in the second year. To note, the estimated HR by Cox model (0.85, 95%CI 0.73-0.99) is similar to that previously reported (XXX, 95%CI XXX-XXX); nonetheless the evident violation of hazards proportionality assumption that was firstly

Commented [NF4]: Several points come to mind here Fabio; first the trials are too short term to make sensible decisions particularly in the lower risk setting. The FDA has required 10 year follow up, although was prepared to make a decision on 1 year. Second, the trials are quite artificial and the TAVI procedure is highly managed in that setting, delivered only by experts into a highly selected population of ideal TAVI candidates (witnessed by the differential loss to procedure in the trial which is highly significant). The OBSERVANT study by Barbanti gives much more pessimistic results for TAVI which in confidence were replicated in the 2011 / 2012 cohorts for GARY too (which is a trial quality dataset with proper risk stratification and information on appropriate exclusion for SAVR) although the differences have asymptoted to the trial result for a 1 year follow up for the 2013-5 cohort. I hope these data will be presented in Lisbon. So what you see here is the best that can be achieved with this technology arguably, but without the same scrutiny and management of the SAVR intervention (which is however much more mature). Third, there remains uncertainty on the safety of some of the TAVI devices eg the widespread recent recall of the SAPIAN 3 balloon device.

Commented [NF5]: See above; just don't report the HRs where the assumptions are violated; they are biased.

demonstrated should lead not to considered that model and shift toward time-varying algorithms. Novel evidences of disbenefit of TAVI at 5 years are emerging in the real world setting, confirming the relationship between TAVI and increased all-cause mortality (HR 1.38; 95%CI 1.12-1.69) and also increased risk major adverse cardiac and cerebrovascular events. Further data on larger registries could also amplified this relationship, as in the RCT enrolled patients are highly selected with a preference for TAVI characteristics and the implanting sites are highly experienced in TAVI, leading toward the best expected results in favor of TAVI compared to the real world results. Hence, feasibility and safety of a new device implantation cannot replace durability and long-term outcomes, and the indications for TAVI in patients with high expectancy of life cannot prescind from long-term evaluation of outcomes.

Other key points emerging from our results are the intrinsic biases of meta-analyses on time-to-event outcomes. Summarizing Cox-estimated HRs reported by single studies intrinsically hypothesized that the hazards proportionality assumption holds also in the summary effect, although the constancy of HR over time could be altered in pooled data. The evident violation of hazards proportionality assumption in the first year and in the total group (Table 2; also confirmed cutting of the follow-up at 2 years) invalidates not only the Cox model but also the summary effect estimated in the forest plot of Figure E3 and the similar effect previously shown in aggregate meta-analyses. Meta-tools were not developed to capture the potential time-effect on the summary effect and should not be employed in a context where the time-varying effect cannot be checked, being the summary effect strictly dependent on it. The same concerns raise on the estimation of heterogeneity, as the lack of heterogeneity showed in the random-effect model (Figure E3, $I^2=0\%$, $p = 0.51$) and confirmed in previous meta-analyses is not confirmed by the more specific random-effect Cox model ($\theta = 0.16$, p -value < 0.001). Also in this case the estimation of heterogeneity within time-dependent effect sizes (HRs) with a random-effect meta-analysis potentially leads to misleading estimates that cannot be checked. Summarizing, with time-to-event data, summarizing the effect sizes (HR) is limited compared with pooling data and estimating a pooled HR.

Limitations.

Our analysis of pooled KM-derived IPD data has some intrinsic limitation. The duration of follow-up is limited up to 5 years and only 3 studies describe 5-year outcomes; as underscored in the text, longer follow-up is urgently required to establish the longer term comparative effects between TAVI and SAVR. The longer follow-up is available for older devices and results should be validated also in trials with newer ones, that potentially could demonstrate better follow-up for improvement in valve design and technical aspects. It was possible to estimate KM-derived IPD data only for all-cause mortality, which is the only standardized outcome reported in all studies, and IPD meta-analysis on valve durability, cerebrovascular events and valvular-related mortality can clear the potential role of devices in affecting expectancy and quality of life. Moreover, this analysis can be stratified only for risk-profile by STS score and EuroSCORE and the potential impact of singular comorbidities on both heterogeneity and outcomes can not be extrapolated. From a methodological point of view, we have not registered the study on PROSPERO (International prospective register of systematic reviews) and publication bias and selective outcome reporting are other potential limitations.

Conclusions.

This pooled analysis of Kaplan-Meier derived individual patient data compared the all-cause mortality between TAVI and SAVR in RCTs. TAVI is related to better survival in the first months after implantation while, after 40 months, it represent a significant risk factor for all-cause mortality. Our data support caution on the wider increase of indications to TAVI without further evaluation at longer follow-up.

Commented [NF6]: Might be worth mentioning the ARTS trial as a beacon of what we should be achieving and sadly are not generally, with a full 10 year follow up and a sound design.

Commented [NF7]: Not sure what you are saying here? We really need these outcomes agreed.

Commented [NF8]: And that largely confounded at the trial level? Don't have individual risk information which is certainly a limitation.

Tables

Table 1. Baseline characteristics of the trials.

Table 2. Unadjusted and Adjusted HRs of TAVI vs SAVR for all-cause mortality by landmark analysis.

	Unadjusted model			Model stratified by risk profile		
	Hazard Ratio	P value	GTt p	Hazard Ratio	P value	GTt p
0 – 12 Months	0.85 [0.73-0.99]	0.03 [§]	0.002	0.85 [0.73-0.99]	0.04 [§]	0.003
12 - 40 Months	0.94 [0.78-1.13]	0.50	0.38	0.93 [0.77-1.12]	0.43	0.38
40 - 60 Months	1.32 [1.03-1.70]	0.03 [§]	0.65	1.31 [1.01-1.68]	0.04 [§]	0.66

GTt Grambsch-Therneau test p-value for testing hazards proportionality assumption; *Reference level SAVR;

Reference level: Low risk profile. [§] p-value <0.05

Appendix Tables

Table E2. Cochrane risk of bias tool for randomized controlled trials.

Study Author	Year	Trial name	Was randomization adequate?	Was allocation concealment adequate?	Were care providers masked?	Were patients masked?	Were outcome assessors masked?	Was overall attrition	Did the study use	Selective reporting?	Other bias?	Risk of Bias
Mack	2015	PARTNER 1A	Yes	Yes	No	No	Yes	No	Yes	No	No	Low
Gleason	2018	CoreValve US Pivotal High Risk	Yes	Yes	No	No	Unclear	No	Yes	No	No	Low
Thyregod	2019	NOTION	Yes	Yes	No	No	No	No	Yes	No	No	Low
Leon	2016	PARTNER 2A	Yes	Yes	No	No	Yes	No	Yes	No	No	Low
Reardon	2017	SURTAVI	Yes	Yes	No	No	Unclear	No	Yes	No	No	Low
Mack	2019	PARTNER 3	Yes	Yes	No	No	Yes	No	Yes	No	No	Low

Commented [NF9]: Give the rate of loss per randomised group and the actual percent rather than the 20% attrition rate which is quite inadequate here.

Abbreviations: ITT = Intention-to-treat analysis, NOTION = Nordic Aortic Valve Intervention, PARTNER = Placement of Aortic Transcatheter Valves, SAVR = Surgical aortic valve replacement, SURTAVI = Surgical Replacement and Transcatheter Aortic Valve Implantation

