

Use Caution in Comparing SFA Trial Data

Not all superficial femoral artery device trials are created equal—in fact, none are.

BY WILLIAM A. GRAY, MD, AND MARIO LANDINI

A speaker has just completed a lecture on superficial femoral artery (SFA) intervention outcomes to rousing applause at a vascular symposium you are attending. The lecture was an impressive amalgamation—a comparison and discussion of all the available peer-reviewed, published, prospective, multicenter, core lab–adjudicated, randomized controlled data from someone universally recognized as an expert clinician and leader in the field. It was utterly bereft of any detectable industry bias. The perfect talk—worth the money, time, travel, and effort you expended. Except that it is almost certainly critically flawed in multiple ways and therefore unintentionally misleading. This article will examine how and why.

COMPARISON-CONFOUNDING CRITERIA

In every prospective device trial performed in the United States under the auspices of the FDA, strict inclusion and exclusion criteria guide the selection of patients who would be appropriate candidates for the device being tested. These selection specifications ensure both patient safety and the sanctity of the trial, and they are purposefully semirestrictive in order to reduce variability and therefore potential confounders. Although many of these criteria are similar across SFA trials, they are not always entirely aligned and, in some cases, may purposefully include or exclude certain patients for device-specific purposes. Current examples include the Shockwave DISRUPT PAD III trial (severe calcification is required, not excluded as in most SFA

trials) or the Intact Vascular TOBA II study (dissection is required, not excluded, as in many drug-coated balloon trials).

However, even imagining that the inclusion and exclusion criteria of two trials were identical, the actual enrolled distribution of patients might differ in the various clinical or angiographic criteria based entirely on chance, geography, type of hospital, etc. Rates of chronic total occlusions, lesion length, vessel diameter, degrees of calcification, and rate of diabetes are routinely different between trials. Complicating these disparities further is the possibility that the definitions of these criteria may differ. In the case of ultrasound or angiographic core lab control, the definitions provided to the labs by the sponsor to analyze the data may vary, and different core labs themselves may have unique algorithms applied for analysis even while using similar definitions.

At a very fundamental level, the populations included in each trial are not the same for all the reasons listed, and it is therefore difficult to directly compare outcomes across trials.

ENDPOINT REPORTING

In nearly every SFA trial conducted to gain FDA approval, the primary efficacy endpoint is 1-year primary patency, which is defined as the freedom from both clinically driven target lesion revascularization (CD-TLR) and binary restenosis (as defined by Doppler ultrasound). This standard seems reasonable enough,

TABLE 1. METHODS FOR ENTERING AND CENSORING SUBJECTS

	Entry to the Model	Follow-Up Time Included in the Model	DUS Performed at 361-390 Days
Model 1	All subjects	360 days	Excluded
Model 2	All subjects	390 days	Included
Model 3	All subjects	360 days	Included, DUS day manually set to 360
Model 4	Evaluable subjects	360 days	Excluded
Model 5	Evaluable subjects	390 days	Included
Model 6	Evaluable subjects	360 days	Included, DUS day manually set to 360

Note: Six methods for including and censoring subjects were simulated. These differ based on entry criteria, length of follow-up, and censorship to DUS-triggered events found beyond the duration of follow-up but within the allowed window.

Abbreviations: DUS, duplex ultrasound.

Reproduced from Vardi M, Lei L, Shangguan S, Doros G. Variability in freedom from loss of primary patency results in trials assessing stent implantation in the superficial femoral artery. *J Invasive Cardiol*. 2014;26:614-617. Used with permission from HMP Global.

except that the methods used in reporting these results are problematic, giving rise to variability in the way outcomes are represented.

Although point estimates are available if one digs for them, typically the 1-year endpoint is displayed as a Kaplan-Meier survival curve with a line drawn at 360 days denoting the value for the primary endpoint efficacy result. However, strictly speaking (and without getting too far into the statistical weeds), the Kaplan-Meier method is intended to be a measure of survival—mortality—usually as the result of a treatment or control; when a death occurs, it is registered in a continuous fashion—daily at a minimum. Herein lies the rub when it comes to the use of Kaplan-Meier estimates in displaying the primary patency endpoint in SFA trials: Only one of the two components of the endpoint is measured/reported daily. Because CD-TLR occurs (and is reported) more or less daily over the course of the study, it is reasonably represented using a Kaplan-Meier curve. TLR events occur continuously throughout the time interval represented; there is no “step” in the curve that would suggest a bunching of events related to lack of assessment or reporting.

Binary restenosis is a different story. It is typically measured by duplex only at prespecified intervals: 6 months and 1 year. As a result, the reporting of the loss of patency is only really possible at those time points, even if it actually occurred at some point in between measurements (which it almost invariably does). That would not be a meaningful problem if all subjects in the study came in together to be measured at exactly on or just before the date of the endpoint reporting (in this case day 360) and were included in the results. But that is not the way research (and life) works. There is usually a 60-day window (30 days before and 30 days after the actual date of reporting) to allow for patient scheduling, travel, or testing availability. As such, a subject’s patency is unknown until they have that endpoint duplex examination, and therefore they cannot be counted as a loss of patency. For all practical purposes, a subject is patent until measured and proven otherwise. This means that the 360-day value on the Kaplan-Meier curve reflects only some of the population being studied.

Comparing data between trials, it is possible that there would be differences in the percentage of the

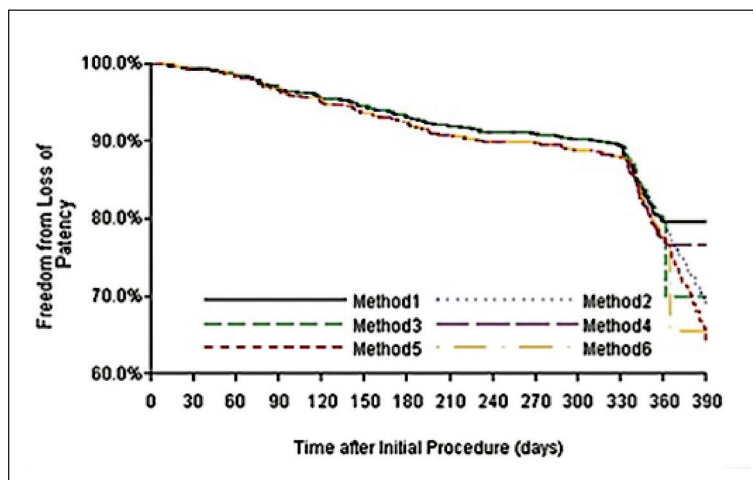


Figure 1. Freedom from loss of primary patency: comparison of six methodologies. Kaplan-Meier estimate for freedom from loss of primary patency curves averaging across results from 1,000 simulations. Reproduced from Vardi M, Lei L, Shangguan S, Doros G. Variability in freedom from loss of primary patency results in trials assessing stent implantation in the superficial femoral artery. *J Invasive Cardiol.* 2014;26:614–617. Used with permission from HMP Global.

population that has been included in the 360-day outcome, and this discrepancy will affect the values reported. For example, Trial A may have measured one-third of its population and Trial B two-thirds of its population, and even if the two devices being tested have the same effectiveness, the patency value reported for Trial A is likely to be higher. A truer reflection would be the 390-day value, and even truer still would be the 2-year data, for which the curve could be analyzed even after the window is closed while still including late assessments of binary restenosis. Modeling of this phenomenon and its effects on patency reporting are nicely illustrated by Vardi et al (Table 1 and Figure 1).¹

Remarkably, as a result of this Kaplan-Meier artifact, the comparison of 1- and 2-year patency values can also be open to interpretation even within the same study. Because the values for patency by Kaplan-Meier given at days 360 and 720 may be “incomplete,” any differences between them may not really reflect the change in patency over the time between years 1 and 2. If the end-of-window 390- and the 750-day values are used instead of the 360- and 720-day numbers,

the reported performance of the device may look quite different.

Try it on a few of the studies already published. Generally, the drop in patency will decrease between 1 and 2 years (ie, the device performance appears to improve).

SUMMARY

Although it is unwise to put much credence into direct comparisons of outcomes for the purposes of deciding which device is “better,” there is value in contextualizing the results of any trial against the backdrop of the studies that came before. Even in the absence of head-to-head trials, clinical evidence must in fact remain the driver for clinical decision-making and technology adoption. Such decisions may be better informed when endpoint and reporting definitions and overall study methodologies are fully disclosed and properly appraised according to their merits and within their limits. ■

1. Vardi M, Lei L, Shangguan S, Doros G. Variability in freedom from loss of primary patency results in trials assessing stent implantation in the superficial femoral artery. *J Invasive Cardiol.* 2014;26:614–617.

William A. Gray, MD

System Chief of Cardiovascular Services
Main Line Health
President, Lankenau Heart Institute
Wynnewood, Pennsylvania
grayw@mlhs.org

Disclosures: Consultant to Medtronic, Boston Scientific Corporation, Gore & Associates, Abbott Vascular, Intact Vascular, Inc., Shockwave Medical, Inc., Philips, and Surmodics, Inc.

Mario Landini

Head of Business Development and Clinical Strategy
Image-Guided Therapy Devices
Philips
Zaventem, Belgium

Disclosures: Full-time employee at Philips.