Journal of the American College of Cardiology
© 2010 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 55, No. 6, 2010
ISSN 0735-1097/10/$36.00
doi:10.1016/j.jacc.2009.10.027

EDITORIAL COMMENT

# When Can Noninferior Be Superior?

## The Multidimensional Nature of Clinical Decision-Making Calls for Innovative Approaches to Clinical Trials*

E. Magnus Ohman, MD, Robert M. Califf, MD

*Durham, North Carolina*

The use of drug-eluting stents (DES) has dramatically changed the face of interventional cardiology with a substantial reduction in repeat revascularization procedures. However, there has not been a reduction in the major cardiovascular events such as death and myocardial infarction (MI) with the first-generation drug-eluting stents (sirolimus-eluting stent [SES] and paclitaxel-eluting stent [PES]) up to 4 years after implantation (1). Also observed in these first waves of randomized trials was a low, but a statistically significant higher rate of late stent thrombosis (after 1 year when compared with bare-metal stents [BMS]), and this difference may be altered by continuous exposure to aspirin and a P2Y12 inhibitor. There was a considerable uptake of the first-generation DES based on the larger effect on restenosis, but the rate of uptake was blunted because of concerns about late stent thrombosis and awareness of the absence of effect on death and MI. This was followed by the second-generation of DES (everolimus-eluting stent [EES] and zotarolimus-eluting stent [ZES]) being approved by the Food and Drug Administration in 2008. Several third-generation DES are also now being tested.

See page 543

We have now entered into a global discussion about comparative effectiveness. The U.S. government currently uses the following definition: "Comparative effectiveness research is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions. The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances. To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations (2)."

In this context, with regard to coronary stents, many questions exist: When should a stent be chosen in preference to medical therapy or surgery? If one chooses a stent, which type is superior? Does the comparative balance of benefit and risk vary as a function of patient or angiographic characteristics? Furthermore, the patient and his/her family, the primary care and noninvasive physicians, and the hospital and health system in addition to the interventional cardiologists all have a stake in decision making.

Typically, the majority of interventional cardiologists pick a device based on personal experience with the device, combined with physical characteristics of the device in terms of delivery to more complex lesions, findings from prospective comparative randomized clinical trials, and perceived effects in patients with particular characteristics from either substudies or individual studies. Much of the literature has focused on diabetes and complex lesion morphology as key differentiating factors. In the U.S., very few interventional cardiologists use price point for decisions on selection for individual patients, whereas health care systems often preferentially stock devices that are at more favorable pricing as long as the DES selected are felt to be similar to the predicated devices (first-generation stents).

Given the rapid shift of American cardiology practices to an employee status within health systems, and the dominance of this model in other parts of the world, the challenge of deciding when a device is "similar" is quite daunting and critical to the health of our patients. Although the word "similar" has no statistical meaning, statistical testing for "similar outcomes" involves a methodology that has been called noninferiority or equivalence testing. In the former, the hypothesis is that the new stent should not perform worse than pre-set criteria; therefore, the implicit hypothesis is that the device is not worse than the standard device. The term "equivalence" is usually considered in a 2-sided analysis (as opposed to noninferiority that only uses 1-sided testing). Thus, a stent can be found to be non-equivalent because it is either worse or better.

As the hurdle for noninferiority testing is strict and requires prior determination of boundaries of acceptable difference between devices, it typically requires a larger number of patients to be randomized. Importantly, with both equivalence and noninferiority hypothesis testing, one could pre-specify that one would test for superiority if the trial met the noninferiority criteria. In this sequential testing approach, one can initially test for noninferiority, and if this hurdle is met, without loss of statistical power (alpha), a formal test for superiority can be done. Implicit in the

*Editorials published in the *Journal of the American College of Cardiology* reflect the views of the authors and do not necessarily represent the views of *JACC* or the American College of Cardiology.

From the Duke University Medical Center, Duke University, Durham, North Carolina. For full disclosures for Drs. Ohman and Califf, please see www.dcri.duke.edu/research/coi.jsp.

hypothesis testing using a 1-sided noninferiority testing strategy is that one does not hypothesize that the stent is better unless one pre-specifies this as an analysis.

The issue becomes even more complex when one considers that it is not only possible, but likely that different components of a classic cardiovascular composite end point are influenced differently by 1 stent compared with another. The testing of a composite for noninferiority leaves little room to further examine differences in specific components, although this could be theoretically possible.

There is substantial knowledge of the performance of the majority of currently available DES from numerous angiographic studies. An interesting picture has emerged with a relative order of in-stent late loss. This is an important parameter that is directly associated with subsequent rates of target vessel revascularization (TVR). Pocock et al. (3) performed an analysis from 11 multicenter randomized stent trials comparing DES with BMS from 5,381 patients. They found that in-stent late loss (ILL) was highly predictive of the subsequent need for target lesion revascularization during follow-up and suggested that ILL could be reliably used as a surrogate end point for target lesion revascularization. As already discussed, however, this approach is insufficient to inform decision making because it does not address the risk of death and MI that weigh more heavily than the risk of restenosis.

Moreno et al. (4) have identified that the SES appears to have the lowest ILL (in millimeters) with a range of approximately 0.10 to 0.20 mm on follow-up angiographic analysis. Similarly, EES has values of 0.10 to 0.16 mm (5,6). PES have had an ILL ranging from 0.30 to 0.40 mm in the early trials (4). Finally, ZES has had late loss of 0.34 to 0.62 (7,8). Thus, of the first-generation DES, PES has had the highest ILL, whereas ZES has had the highest ILL loss among the second-generation stents. For comparison, the ILL among BMS in the early comparative trials has ranged from 0.63 to 1.05 mm, suggesting that the ILL of ZES is closer to some BMS in overall comparison. Importantly, in the prospective randomized trial (ENDEAVOR II) (8), the ILL was 1.03 mm with BMS and 0.61 with ZES, which was a statistically significant difference ($p < 0.001$).

In this issue of the *Journal*, Leon et al. (9) publish the results of the ENDEAVOR IV (Randomized Comparison of Zotarolimus-Eluting and Paclitaxel-Eluting Stents in Patients with Coronary Artery Disease) trial, which was a prospective, randomized comparison of ZES versus PES for single de novo coronary artery lesion. The primary hypothesis was a noninferiority test with regard to 9-month target vessel failure, which was defined as cardiac death, MI, or TVR. The ZES was found to be noninferior with a target vessel failure rate of 6.6% for ZES versus 7.1% for PES ($p < 0.001$). This conclusion of noninferiority with regard to the composite end point was driven by the combination of fewer periprocedural MIs with ZES (0.5% vs. 2.2%, $p = 0.007$), but with a numerically higher rate of in-segment restenosis with ZES 15.3% versus 10.4%, ($p = 0.24$). Based on these findings, Leon et al. (9) conclude that ZES has similar clinical safety and efficacy as PES among single de novo coronary lesions.

From the decision maker's perspective, this trial raises the question of whether ZES and PES are really similar or whether the significant reduction of periprocedural MIs, which is highly significant, should be considered a superior attribute of ZES? The study had not set up an a priori hypothesis testing for superiority for any individual end points; thus, in statistical terms, the most prudent conclusion is that ZES is indeed noninferior and that a reduction in MI observed is of interest but can only be hypothesis-generating and should be tested in another trial. Tantalizing in this regard is the fact that in the ENDEAVOR III (Randomized Comparison of the Endeavor ABT-578 Drug Eluting Stent with Bare Metal Stent for Coronary Revascularization) trial, MI was also significantly reduced (0.6% vs. 3.5% with ZES compared with SES, $p = 0.04$) but this was from a much smaller experience of only of 436 patients (7). In both the ENDEAVOR III and IV trials, the rate of in-segment restenosis was numerically higher but statistically so only in the ENDEAVOR III trial. Thus, the combined trial experience to date suggests that ZES may indeed have a positive effect on periprocedural MI but an untoward effect on late segment loss and/or restenosis.

The EES is another second-generation stent that also has been evaluated in large-scale clinical trials. Similarly, the pivotal SPIRIT III (Clinical Evaluation of the Xience V Everolimus Eluting Coronary Stent System in the Treatment of Patients with de novo Native Coronary Artery Lesions) trial was also a 1-sided noninferiority study but with a pre-specified superiority testing if noninferiority was met (5). The ILL was 0.16 mm for EES compared with 0.30 mm for PES ($p = 0.002$). The composite major cardiac events were reduced at 1 year (6.0% vs. 10.3%, $p = 0.02$), which was confirmed at the 2-year follow-up (10) with a significant 32% reduction in target vessel failure (10.7% vs. 15.4%, $p = 0.04$), and a 45% reduction in major cardiovascular events (cardiac death, MI, target lesion revascularization, 7.3% vs. 12.8%, $p = 0.004$). Thus, both the second-generation stents appear to have trend toward superiority for the hard clinical end points but EES, when compared to ZES, appears to have a more favorable ILL.

How do the interventional cardiologist and the multiple other decision makers now choose a device for the individual patient? Both stents are by statistical study design noninferior, both for ILL as well as for target vessel failure, defined as death, MI, and TVR. Both show tantalizing benefits, such as reduction in MI, which incidentally was not shown in the BMS versus DES trials up to 4 years. Some physicians will be swayed by the larger difference in the composite benefit observed with EES, whereas others will be swayed by the reduction in MI by ZES, but none of this is conclusive. Part of the issue at hand is the relatively small sample size of these second-generation stent trials. Furthermore, they both were designed as noninferiority

trials and, in the case of the SPIRIT trials, with a prespecified idea that possibly EES might be superior. No such sequential testing was part of the statistical approach for ZES.

Our belief is that the best approach to this conundrum is to have more appropriately sized clinical end point-based investigation to better capture the overall intent of comparative research, namely improvement in significant clinical end points such as death, MI, and the need for subsequent revascularization. Whereas these trials should be designed to either capture superiority or true equivalence, the complexity of the possible findings should not be dismissed because rational and intelligent people often will value different parts of a composite end point in different ways. For example, the significance of periprocedural MI in the absence of an obvious procedural complication continues to be debated and some major debate continues about the relative value of preventing a restenosis event precipitating TVR and an episode of stent thrombosis.

From the preceding discussion, it will also be apparent that there will be a strong interest to use ZES as the comparator in future trials as it has the highest ILL of DES. We would like to discourage manufacturers and researchers from this approach, but rather test it against the 2 devices that so far appear to have either the lowest late loss (SES) or the stent with the better clinical outcomes (EES compared with the other first-generation stent, PES). Alternatively, a pragmatic trial design would compare a new stent against any approved device, but the sample size with such a heterogeneous control group would be substantially larger. Although this may be attractive from a clinical perspective, it may not be either practical and/or easy to understand, as physicians will likely select certain lesion or patient characteristics, for 1 of the control DES versus the other, making any comparison and subgroup analysis difficult to evaluate.

Our recommendation for any third- and fourth-generation DES is that the first hurdle should be that any new DES meets the stricter criteria of equivalence for ILL. By defining equivalence, one will narrow the range to which ILL could be expected: for example, ILL of one-half of the 0.54 mm, thus preserving at least 50% of the late loss when compared with BMS. Such a trial would be larger than a noninferiority trial, but it would also possibly point toward potential superiority. Once such a phase-2 program has been defined, a larger phase-3 program driven purely by clinical end points and with superiority testing versus noninferiority followed by superiority testing for the third-generation DES might be considered. This stepwise approach would help guide the clinical trials' methodology and allow the most appropriate clinical trials to be used, which ultimately will help physicians as DES technology matures. With increased certainty, our patients can also be more certain that they ultimately would get the best stent for their clinical condition and not have the difficult decision of trading off a reduction in MI versus later restenosis.

**Reprint requests and correspondence:** Dr. E. Magnus Ohman, Duke Heart Center, Duke University Medical Center, Cardiology, Room 8676A HAFS Building, Duke Hospital Erwin Road, Box 3126, Durham, North Carolina 27710. E-mail: ohman001@mc.duke.edu.

## REFERENCES

1. Stone GW, Moses JW, Ellis SG, et al. Safety and efficacy of sirolimus- and paclitaxel-eluting coronary stents. N Engl J Med 2007;356:998–1008.
2. U.S. Department of Health and Human Services. Comparative Effectiveness Research Funding. Available at: http://www.hhs.gov/recovery/programs/cer/draftdefinitionhtml. Accessed September 2009.
3. Pocock SJ, Lansky AJ, Meharan R, et al. Angiographic surrogate end points in drug-eluting stent trials. J Am Coll Cardiol 2008;51:23–32.
4. Moreno R, Fernandez C, Sanchez-Recald A, et al. Clinical impact of in-stent late loss after drug-eluting coronary stent implantation. Eur Heart J 2007;1583–91.
5. Stone GW, Midei M, Newman W, et al. Comparison of an everolimus-eluting stent and a paclitaxel-eluting stent in patients with coronary artery disease. JAMA 2008;299:1903–13.
6. Serruys PW, Ong ATL, Piek JJ, et al. A randomized comparison of a durable polymer everolimus-eluting stent with a bare-metal coronary stent: the SPIRIT first trial. Eurointervention 2005;1:58–65.
7. Kandzari DE, Leon MB, Popma JJ, et al. Comparison of zotarolimus-eluting and sirolimus-eluting stents in patients with native coronary artery disease. J Am Coll Cardiol 2006;48:2440–7.
8. Fajadet J, Wijns W, Laarman GJ, et al., on behalf of ENDEAVOR II Investigators. Randomized, double-blind multicenter study of the Endeavor zotarolimus-eluting phosphorylcholine-encapsulated stent for treatment of native coronary artery lesions: clinical and angiographic results of the ENDEAVOR II trial. Circulation 2006;114:798–806.
9. Leon MB, Mauri L, Popma JJ, et al. A randomized comparison of the Endeavor zotarolimus-eluting stent versus the TAXUS paclitaxel-eluting stent in de novo native coronary lesions: 12-month outcomes from the ENDEAVOR IV trial. J Am Coll Cardiol 2010;55:543–54.
10. Stone GW, Midei M, Newman, et al. Randomized comparison of everolimus-eluting and paclitaxel-eluting stents. Circulation 2009;119:680–6.

**Key Words:** drug-eluting stents ■ target lesion revascularization ■ zotarolimus-eluting stent.