PANEL DISCUSSION

# Exploring the Paclitaxel Safety Signal

Experts discuss considerations in trial design, adjudication, and paths forward.

WITH MARIANNE BRODMANN, MD; WILLIAM GRAY, MD; PETER A. SCHNEIDER, MD; RAMON L. VARCOE, MBBS, MS, FRACS, PHD; AND THOMAS ZELLER, MD, PHD

Marianne Brodmann, MD
Substitute Head of the Clinical Division of Angiology
Department of Internal Medicine
Medical University of Graz
Graz, Austria
marianne.brodmann@medunigraz.at
Disclosures: Advisory board for Medtronic, Philips, Spectranetics Corporation, Biotronik, BD/Bard, and Surmodics, Inc.
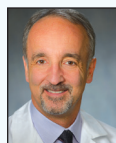
William Gray, MD
System Chief of Cardiovascular Services
Main Line Health
President, Lankenau Heart Institute
Wynnewood, Pennsylvania
grayw@mlhs.org
Disclosures: Consultant to Boston Scientific Corporation, Medtronic, Abbott Vascular, Philips, and Surmodics, Inc.

Peter A. Schneider, MD
Professor of Surgery
Division of Vascular & Endovascular Surgery
University of California, San Francisco
San Francisco, California
peteraschneidermd@gmail.com
Disclosures: Scientific advisory board for Abbott, Medtronic, and Boston Scientific Corporation; consultant to Cardiovascular Systems, Inc., Silk Road Medical, Profusa, Surmodics, Inc., Intact Vascular, Inc., and Cagent Vascular.

Ramon L. Varcoe, MBBS, MS, FRACS, PhD
Vascular Surgeon
Prince of Wales Hospital
Associate Professor of Vascular Surgery
University of New South Wales
Sydney, Australia
r.varcoe@unsw.edu.au
Disclosures: Consultant to Abbott Vascular, Boston Scientific Corporation, Medtronic, Intact Vascular, Inc., Shockwave Medical, Inc., Surmodics, Inc., and Intervene.

Thomas Zeller, MD, PhD
Department of Angiology
Universitäts-Herzzentrum
Freiburg-Bad Krozingen
Bad Krozingen, Germany
thomas.zeller@universitaets-herzzentrum.de
Disclosures: Honoraria from Abbott Vascular, Veryan Medical Ltd., Biotronik, Boston Scientific Corporation, Cook Medical, Gore & Associates, Medtronic, Philips/Spectranetics, Shockwave Medical, BIBA Medical; consultant to Boston Scientific Corporation, Gore & Associates, Medtronic, Veryan Medical Ltd., Intact Vascular, Shockwave Medical, Bayer, Vesper Medical; research, clinical trial, or drug study funds received by institution from Bard Peripheral Vascular, Veryan Medical Ltd., Biotronik, Cook Medical, Gore & Associates, Medtronic, Philips, Terumo, TriReme Medical, Shockwave Medical, Med Alliance, Intact Vascular, B. Braun, Cardiovascular Systems, Inc., Boston Scientific Corporation; stock in QT Vascular.
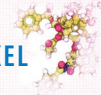
What is the single biggest learning point from the evaluation of paclitaxel delivery devices after the meta-analysis?
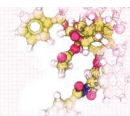
Prof. Varcoe: I am hopeful that clinical trialists, sponsors, and the vascular community will place a greater emphasis on safety outcomes such as long-term mortality. In the past, trial endpoints have been focused on efficacy, an approach that has guided trial design and follow-up protocols. If we take one thing away from this experience, I would like it to be greater scrutiny by regulators toward trial design that prioritizes safety and long-term follow-up.

Dr. Gray: Although it has nothing to do specifically with paclitaxel per se, probably the most significant point learned, and one that will affect future trial conduct, is the value of complete follow-up data. The "missingness" of data in the paclitaxel studies was not unique in the general scope of most studies. However, when combined with the relatively small numbers of total enrolled patients, the inability to get even rudimentary vital statistics data on most withdrawn patients, the lack of rigor in defining crossover rates (and assigning those patients a different status vis-à-vis paclitaxel exposure), and the possible ascertainment bias due to lack of blinding, it fundamentally weakened the meta-analysis to the point where it became, frankly, uninterpretable.

Prof. Zeller: In general, downstream drug loss during balloon insertion and inflation must be reduced to the minimum amount to exclude any potential systemic side effects; this is independent from the type of drug. To achieve the best possible technical and—as a result—clinical outcome, we will need biologicals coated on balloons or stents in the future. As a consequence of the meta-analysis, all drug-based devices will need to prove long-term safety. Therefore, it will be essential to develop drug-releasing devices that guarantee exclusive local drug delivery.

Dr. Schneider: Randomized controlled trials (RCTs) are geared to discern differences in effectiveness, but they are also more subject to different kinds of bias than I anticipated, especially when it comes to vital statistics. In addition, this is the first time in my career that major and sudden changes have been made in practice worldwide based on a potential risk demonstrated by a meta-analysis without clarity as to the mechanism or incidence.

Prof. Brodmann: Patients with peripheral artery disease (PAD) are sick and have to be carefully followed after their endovascular procedures to detect relevant comorbidities and improve life expectancy.

One prevailing theme in discussions on the mortality signal is that of perceived shortcomings in the current class of RCTs for drug delivery technologies. In particular, the overall powering combined with relatively high numbers of loss to follow-up have been highlighted. As some of the key trialists of the generation, to what degree do you think modern PAD trials have adequately evaluated the study devices in terms of both safety and efficacy?
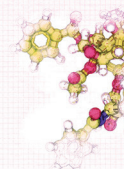
Dr. Schneider: There has been a lot of criticism of these trials; however, it should be noted that, for the most part, they did what they were intended to do. That is, the efficacy of a variety of devices was assessed, and this added much to our understanding of paclitaxel device effectiveness in PAD. No one, myself included, anticipated the need to power the trial for 5-year mortality. Whether the mortality risk of paclitaxel turns out to be real or not, I anticipate that drug delivery devices will be assessed using larger trials, longer-term endpoints, and lengthier follow-up, accompanied by greater efforts to follow every patient.

Prof. Brodmann: Modern peripheral trials are better than older peripheral trials at evaluating device technologies because they do address longer-term outcomes. A benefit of studying drug-coated technologies is that device companies are interested in how the drug might work for a longer time period. But the modern trials are still underpowered, and mortality has not been an endpoint in any of the device trials so far. Furthermore, follow-up in these trials does not last long enough. There is a need to look at adequate coronary trials and new technologies and maybe take some advice from those in terms of how to achieve an adequate trial setting for the peripheral vasculature.

Prof. Varcoe: It was clear from the *Journal of the American Heart Association (JAHA)* systematic review that trial numbers meta-analyzed at longer time points of 2 to 5 years were low, making them prone to type-1 error when analyzed at a summary level. We have seen a correction as additional data have come forward by adjusting for crossovers after analyzing treatment received at a patient level, adding additional data from trials such as LEVANT 1 and 2, IN.PACT Japan, and AcoArt I and from chasing down patients lost to follow-up. This has moved the pooled estimate toward the null and eliminated statistical significance, illustrating how careful we must be in interpreting an underpowered meta-analysis. I believe that as a vascular community, we have done a poor job of providing sufficient trial data for longer time points, which is required to give us confidence in these meta-analysis results. Moreover, safety outcomes have often been an afterthought, as evidenced by the high proportion of missing patients at longer follow-up.

Prof. Zeller: In my opinion, the outcome (mortality signal) of the meta-analysis is artificial. European and Asian RCTs showed opposite outcomes, even resulting in a trend to higher mortality rates in the control groups (eg, AcoArt I). In my opinion, current study

devices have been sufficiently evaluated regarding safety.

Dr. Gray: Regarding efficacy, yes—I believe that the trials are adequately constructed and powered to assess it. But two more relevant extensions of that question are whether the comparators we are powering against are reasonable (eg, percutaneous transluminal angioplasty [PTA]) and whether noninferiority or superiority is tested. Separately and in combination, these will both make the results of the current and recent trials (and the device being tested) more or less useful, and they will almost certainly affect how we conduct trials going forward.

Regarding safety, it's a numbers game. It appears that we are measuring the correct safety outcomes, and thankfully, events are infrequent. However, this poses a statistical dilemma because it is then difficult to power safety endpoints with so few individual events. Accordingly, composite safety endpoints are constructed when possible, allowing for more statistical robustness. As a matter of fact, the number of patients mandated for a given trial is not infrequently the result of the need to gather enough safety events—not for an efficacy endpoint, which might be statistically satisfied with fewer subjects.

Dr. Gray, compared to percutaneous coronary intervention (PCI), the populations in PAD trials are significantly smaller. How were the requisite N sizes determined in each setting (PCI and PAD), and why were those in PAD trials smaller?
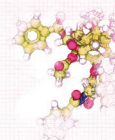

Dr. Gray: When the rates of events—let's say target lesion revascularization (TLR) or target lesion failure—came down in PCI trials as devices improved (bare-metal stents to drug-eluting stents [DESs]), more numbers were required to show statistical differences. For example, in the early BENESTENT trial in 1994 that compared PTA with first-generation bare-metal coronary stents, approximately 500 patients were randomized. In those early PCI days, the assumed rate of restenosis events was 30% in the PTA arm with an assumed treatment effect of 40%.[1] If those numbers sound familiar,

it's because they are roughly the same assumptions that one might use in a drug-coated balloon versus PTA trial. Alternatively, trial designs for coronary PCI today assume event rates around 6% and noninferiority margins of roughly half that.[2] That leads to larger trials in the range of approximately 1,500 patients. When we achieve those types of successes more routinely—the Eluvia DES (Boston Scientific Corporation) in the IMPERIAL trial did approach those same event rates—we will need to conduct larger trials, assuming we are comparing devices with similarly low rates.

Do you think future FDA investigational device exemption (IDE) trials should all follow a specific design with common definitions and endpoints, continue to be developed individually, or somewhere in between?
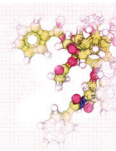
Dr. Schneider: We already have a lot of specific designs and common definitions, especially for claudication trials. For example, assessment of patency after superficial femoral artery (SFA) treatment has been well established using a primary endpoint of freedom from duplex-derived restenosis and clinically driven TLR at 1 year. Hemodynamic measurements are important and should be performed in every patient. There is a strong interest in quality-of-life measures, and these have been increasingly included in trials. We will likely see this become more standardized going forward.

However, I do not believe that specific designs will be of value in most situations, and we will continue to develop most trials on an individual basis. I also don't believe that drug delivery devices should be tested in the same way as devices that do not deliver drugs. Head-to-head device trials are different than a device versus a known and long-term standard such as PTA. Eventually, we want to have a body of evidence large enough for control groups so that performance goals can be constructed.

Dr. Gray: It would make sense to try to standardize definitions and endpoints whenever feasible, especially in light of some potential unforeseen need to amalgamate the data from individual unrelated trials in the same class of devices, as was the case in the mortality question in the recent paclitaxel example.

Profs. Brodmann and Zeller, do you foresee increased uniformity in European-based clinical trial designs, either related to the paclitaxel discussion or changes in CE Mark regulation?
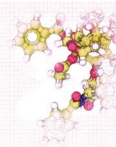
Prof. Brodmann: Yes, I hope that what we have experienced so far in Europe since December 2018 might get us a step closer to uniformity in European-based trials and move us into a more sophisticated approach with regard to CE Mark regulation.

Prof. Zeller: There may be changes in the duration of follow-up requested for the primary safety endpoint, potentially extending it to 3 years and beyond. I believe and fear that European agencies will very much follow FDA recommendations.

Prof. Varcoe, at the Vascular Leaders Forum and the FDA panel hearings, you presented early results from research suggesting that study arms in SFA device trials may experience higher reported mortality rates compared with control arms, even if the devices do not include paclitaxel. How would you briefly summarize this finding and the status of the study, and what are some of the possible reasons for this observed trend in study arms?
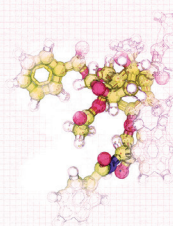
Prof. Varcoe: We performed a systematic review and meta-analysis of RCTs that compared mortality in experimental versus control treatment for the SFA. Importantly, we excluded trials that evaluated a drug-coated device to reduce the possibility that our findings might be related to the antiproliferative drug. We found very similar results to the *JAHA* meta-analysis, where experimental treatment arms had higher mortality rates. This is something clinical trialists have anecdotally described for many years, and it suggests that there may be factors other than paclitaxel that are responsible for this phenomenon.

We hypothesize that there is a contribution from ascertainment bias, where the unblinded clinical trial team is more likely to tenaciously chase down patients who received an experimental therapy but failed to respond to contact efforts; thus, the trial team is more likely to accurately record those patients as "mortality" rather than "lost to follow-up." There is also the increased medical interaction that comes with control

treatments known to have an increased likelihood of TLR. Each repeat revascularization brings more medical touch points with additional opportunities to optimize or enhance medication therapy, control risk factors, and encourage exercise and smoking cessation—all factors known to increase life expectancy.

Dr. Schneider, in evaluating the available data from RCTs after publication of the meta-analysis, what has been learned about confounding elements in the various control arms? Which elements are common to several or all trials, and which might be specific to single trials/possible anomalies?
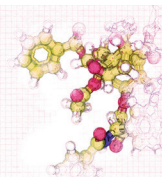
Dr. Schneider: We have learned that studying efficacy is very different from studying vital statistics such as long-term mortality. Because our focus has been on efficacy, specifically patency, when a patient loses patency, there is very little incentive to be followed years later, thus creating an ascertainment bias. When many patients are lost to follow-up, for a wide variety of reasons, it may skew the results in one direction. Another important example is that in the early phase among those who lost patency, medical management was more intensive than in those who remained asymptomatic, and this may have had an impact.

Lifetime paclitaxel use was not recorded in any of the RCTs. A substantial minority of the patients will have had paclitaxel in the contralateral limb, in the ipsilateral limb during follow-up, or by some other mechanism (eg, treatment of arteriovenous graft/fistula or cancer). This prevents us from identifying who received what, which makes causal relationships nearly impossible to clarify.

One of the key challenges in interpreting the data has been the lack of a clear causal link between paclitaxel and mortality. Event adjudication is particularly difficult in this population due to the presence of multiple patient comorbidities as well as the lateness of the effect signal. How has this affected exploration of the signal, and is it possible to design a trial such that event adjudication can adjust for these challenges?

Prof. Varcoe: There are several problems with attempting to adjudicate death events related to paclitaxel. First, there is no identified biologic mechanism of toxicity despite input from toxicologists, hematologists, and biological scientists. Second, in an FDA analysis of an "as-treated" population identified from their IDE trials, they found no predominant cause of death. It is likely that a toxic drug would have a single mechanism of action that would be identified as a cause-of-death "spike" when compared alongside control group deaths. Third, paclitaxel has been used in very high doses for curative breast cancer treatment since it was first FDA-approved in 1992. Long-term data from that group of patients demonstrate that paclitaxel reduces mortality, raising considerable doubt about toxicity. Fourth, when drug dose has been investigated, it has not been found to be associated with increased mortality.[3] A biologic gradient does not exist, which raises further doubts about toxicity and fails to satisfy the Bradford Hill criteria for causality.[4] Therefore, with no clear definition of what a paclitaxel-related death might look like and considerable doubt as to whether toxicity is involved at all, it seems unlikely that trial design will ever be in a position to adjudicate such hypothetical paclitaxel toxicity events.

Prof. Zeller: This challenge could only be overcome if, in multiple 10,000-patient sample sizes, some clusters of increased mortality could be identified. This is very unlikely because the mortality rates for drug-coated technologies and bare devices are both within the expected margins based on historic epidemiologic studies.

Dr. Gray: Competing risks for death (age, diabetes, heart failure, renal failure, etc) that have established and clear causality can never really be separated from paclitaxel in these populations, which makes the lack of a causal link between paclitaxel and very late death all the more important. In my mind, lacking that link—absent some heretofore undescribed plausible effect of paclitaxel becoming evident in the future—further limits the impact of any findings at the meta-analysis level.

Future trials should be large, with long follow-up and enough patients and time to look not only at the

question of paclitaxel exposure but also at dosing as a possible agent of mortality. It is estimated that it would require tens of thousands of patients.

Prof. Brodmann: I think that the severity of PAD, presence of comorbidities, and consecutive negative outcomes concerning high mortality are completely new in the world of device trials; therefore, event adjudication has to be redefined in future trials. This is possible, but the trial needs a completely new setup and physicians who know PAD patients.

Dr. Schneider: Event adjudication is a weakness for all PAD trials and may be a weakness for many other fields as well. Methods of determining cause of death are not standardized, autopsy rates are low, and death may occur that is not witnessed or that takes place in a

setting where medical expertise is not readily available. In general, cause of death in a given patient's story may be multifactorial and unclear as to relevant importance of potential causes.

These things have made exploration of the signal more challenging. It is not so much about designing a trial such that event adjudication can adjust to these challenges but taking a global look at the problem, attempting to standardize adjudication of mortality and other major events, and placing some urgency around the best possible identification of the cause.

At present, it appears that there is no clustering of deaths in the years after treatment. This factor, combined with the absence of a dose response between paclitaxel and mortality risk, makes it so that identifying an underlying biologic mechanism may not be possible.

How has your practice using paclitaxel devices changed throughout 2019 as a result of the meta-analysis and its aftermath? How are you communicating with your patients about the results of the meta-analysis, other studies, and your own professional opinions?
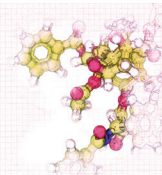
Prof. Brodmann: It has not changed at all. I guess the main reason was that with our tight schedule of follow-up in general at our institution, we had a real sense of the advantages of paclitaxel-coated devices for our patients. As one side effect, we were able to reduce the number of patients on our institutional waiting list for endovascular therapies and saw no warning sign with regard to drug-coated technology.

I have to mention as a personal statement that I'm usually very suspicious of any new approach, especially those concerning treatment options for our patients. Strong evidence is needed to convince me that the new approach is better than the old one. However, looking after the patients we treated with drug-coated technologies has made me a "drug believer."

Prof. Varcoe: The meta-analysis has raised concerns around a safety signal in claudicants who live beyond 2 to 5 years. I have very few young and healthy claudicants in my practice, so my approach has changed very little. I tell patients that a statistical signal has been observed in a specific group of patients, and it may or may not apply to them. We are not sure whether it represents a real danger; however, we do know that drug-coated devices reduce the likelihood that you will have to return for repeat procedures in the future. I then tell them that my approach is to assess their individual risk of receiving a drug-coated device versus not receiving one. In this period of uncertainty, I will try to make the best decision for them at the time.

It's not easy to tell patients that we don't always know their risk, but I find that being honest helps enhance trust in the doctor-patient relationship, and they very much appreciate it.

How do you contextualize various study types when trying to reconcile disparate findings, and how do you prioritize/value an RCT versus a meta-analysis?
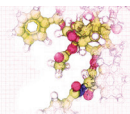
Prof. Zeller: I regard the RCT as being the highest level of evidence, despite all sample size limitations. Meta-analyses rely too much on the appropriate execution, as demonstrated by the Katsanos et al *JAHA* study.

Dr. Gray: A well-constructed and well-conducted RCT will need no meta-analytic treatment. But classically, meta-analyses are thought of as the pinnacle of available scientific data, especially when asking a question with a low frequency rate (eg, mortality in claudi-cants) that is typically not addressed in single RCTs. But I would modify that: summary-level data (as was used in the Katsanos et al *JAHA* paper) can only be, at best, hypothesis generating. Furthermore, authors performing meta-analyses must make certain choices in how the analyses are performed, some of which may be forced by the lack of available data. This can significantly limit the value of the meta-analysis, specifically in terms of how missing data are handled and how crossover is identified and handled.

Where do you place the emerging new registry and data collection models in the data hierarchy, and how do you compare them to other registries?
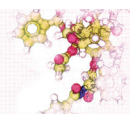
Dr. Schneider: In general, we are taught to place RCTs at the top of the research quality pyramid. However, there are some caveats to this. RCTs are terrific for head-to-head battles on efficacy, but if numerous patients are lost to follow-up, the long-term mortality data will not be reliable. Massive registry data with excellent ascertainment of death may offer a better assessment of mortality than the aforementioned RCT data.

Dr. Gray: I'm not always a fan of "big data," but I believe some of the strongest data we have comes from Medicare, Optum, and other large real-world data sets that have been adjusted or propensity matched. Using these data to construct time-to-event analyses has been very useful and revealing. Although I recognize the limitations (possible selection and ascertainment issues, definitional challenges, differences in inpatient versus outpatient and/or study subjects), they are nevertheless reassuring and directionally consistent with each other.

What are your thoughts on the FDA's most recent communication from August 2019?
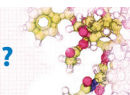
Prof. Brodmann: I feared much more negative advice from the FDA. I think that this communication might allow us to go on with all the trials we are in so far and carry on with our daily practice. Yes, we need additional communication with our patients, but we have been doing this since December 2018.

Prof. Varcoe: I thought it was an appropriate response. The FDA softened its language and put the onus back on doctors to make individual risk-benefit assessments for their patients. It also emphasized the importance of continuing to study this safety signal through ongoing clinical trials.

Dr. Gray: The letter largely reflected the panel deliberations, although it did not give as much emphasis or credence to the large data sets that were also presented to the panel, specifically the Medicare and Optum analyses. Given the multiple limitations that weakened the analysis, this seems to be a reasonable approach. In the end, the March letter effectively and largely shifted the burden of the device use to the practitioner. It would have been great to walk that back a bit more with the August letter.

Dr. Schneider: I am optimistic that we are moving toward a longer-term solution to this challenge and that more data will help us develop clarity. The FDA has acknowledged much of what the frontline physician is facing. There seems to be a mortality signal, but there is no dose response or mechanism. The quality of the data from which the signal was derived is poor because it was designed to determine 1-year patency, not 5-year mortality. The patients must be informed, clinical trials must

be continued, and we must use these devices judiciously. Since the initial meta-analysis, all subsequent developments in the data have suggested that either there is no signal or the mortality signal is much smaller than initially reported.

Prof. Zeller: The wording slightly lessens the severity of the second announcement in terms of leaving the decision of whether or not to use paclitaxel-eluting devices to the discretion of the physician and the respective patient.

What are the keys to next-generation SFA device study designs?
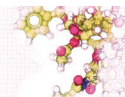
Prof. Zeller: Follow-up compliance and sticking to prescribed medications should be a greater focus to guarantee equal patient cohort sizes during follow-up and exclude confounding parameters potentially affecting predefined endpoint outcomes.

Prof. Brodmann: Here are some of the key areas in my opinion:
- A larger population size than before
- Much more rigid follow-up/patient communication protocols; we need long-term follow-up to prove that what we do prolongs patient survival
- Control arm elements including paclitaxel-coated balloons or new drug-coated devices (sirolimus vs paclitaxel)
- Endpoint definition commonality; mortality should be an endpoint!
- Multinational collaboration is a "must have"
- I would be in favor of multidevice pooling

Prof. Varcoe: I think it's essential that we focus on reducing patient numbers lost to follow-up or withdrawn. High proportions of either raise uncertainty as to the safety of any device being investigated. I would like to see greater efforts being made to contact patients through their family members, primary care physician, and local hospital; have tiered consent forms that enable death to be established even after a patient has withdrawn from the study; and document patient mortality through Medicare databases or death registry data linkage. ∎

1. Serruys PW, de Jaegere P, Kiemeneij F, et al. A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group. N Engl J Med. 1994;331:489-495.
2. Mauri L, Doros G, Rao SV, et al. The OPTIMIZE randomized trial to assess safety and efficacy of the Svelte IDS and RX sirolimus-eluting coronary stent systems for the treatment of atherosclerotic lesions: trial design and rationale. Am Heart J. 2019;216:82-90.
3. Holden A, Varcoe RL, Jaff MR, et al. Paclitaxel and mortality: the dose argument is critical. J Endovascular Ther. 2019;26:467-470.
4. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295-300.

Endovascular NEWS

CONTINUING COVERAGE:

PACLITAXEL IN PAD

evtoday.com/sfa/paclitaxelcontroversy