

# Surrogate and clinical endpoints in interventional cardiology: are statistics the brakes?

Matthias Waliszewski and Harald Rittger

## Abstract

**Background:** Randomized controlled trials are the gold standard for demonstrating safety and efficacy of coronary devices with or without accompanying drug treatments in interventional cardiology. With the advent of last-generation drug-eluting stents having enhanced technical attributes and long-term clinical benefits, the proof of incremental angiographic or long-term clinical efficacy becomes more challenging. The purpose of this review is to provide an overview of the most common and alternative study endpoints in interventional cardiology and their potential reimbursement value. Moreover, we intend to describe the statistical limitations in order to demonstrate differences between potential treatment groups. Furthermore, careful endpoint recommendations for a given patient number are offered for future study designs.

**Methods:** The number of patients per treatment group was estimated for various study designs such as noninferiority test hypotheses with hard clinical endpoints and various surrogate endpoints. To test for differences in various surrogate endpoint scenarios, the corresponding patient group sizes were explored. To evaluate these endpoints in terms of their reimbursement impact, preferred endpoints for technical appraisals in interventional cardiology at the National Institute of Health and Care Excellence (NICE) were used.

**Results:** Even with the most stringent experimental control to reduce bias-introducing factors, studies with hard primary clinical endpoints such as the occurrence of major adverse cardiac events (MACE) or target-lesion revascularization (TLR) rates remain the gold standard, with numbers reaching into the 300–700 patient range per group. Study designs using loss in fractional-flow reserve (FFR) or stent-strut-coverage rates can be statistically formulated; however, the clinical ramifications for the patient remain to be discussed. Nonrandomized study designs with inpatient angiographic controls in nontarget vessels may merit further thoughts and explorations.

**Conclusions:** From a reimbursement impact, the primary endpoints MACE and TLR are the best choices for a moderately sized study population of 500 patients per group. Angiographic endpoints, in particular minimal lumen diameter (MLD), are not useful in this context. The emerging endpoints such as loss in FFR or stent coverage require smaller patient populations. However, their impact on reimbursement-related decisions is limited.

**Keywords:** clinical endpoints, interventional cardiology, reimbursement impact, sample size estimates, surrogate endpoints

## Background

The Benestent trial [Serruys *et al.* 1994] is one of the most important landmark studies in interventional cardiology that was conducted more than 20 years ago. It revealed that stenting was superior

to plain old balloon angioplasty (POBA). Its ‘statistical engine’ was a continuity-corrected Chi-squared test (double-sided  $\alpha = 5\%$ , 80% power), based on an event rate of 30% in the POBA group and 18% in the bare-metal stent (BMS) group.

*Ther Adv Cardiovasc Dis*

2015, Vol. 10(5) 314–326

DOI: 10.1177/  
1753944716656150

© The Author(s), 2016.  
Reprints and permissions:  
[http://www.sagepub.co.uk/  
journalsPermissions.nav](http://www.sagepub.co.uk/journalsPermissions.nav)

Correspondence to:  
**Matthias Waliszewski, PhD**  
Medical Scientific Affairs,  
B.Braun Melsungen AG,  
Sieversufer 8, Berlin  
12359, Germany  
[matthias.waliszewski@bbraun.com](mailto:matthias.waliszewski@bbraun.com)  
**Harald Rittger, MD**  
Klinikum Fürth, Fürth,  
Germany

Expected differences in event rates of 12% are in today's clinical practice a rare privilege for statisticians. Since then, improvements in procedural outcomes, comedication regimens and devices, led to very small differences between two treatment groups. Therefore, clinical trials based on these small, expected differences are very challenging to conduct. If one ventures to compare two drug-eluting stents (DESs) in terms of binary-restenosis rates using the same statistical hypothesis as in the Benestent trial, a minimum of 984 patients per group would have to be recruited (double-sided  $\alpha = 5\%$ , 80% power, binary-restenosis rate = 5%, 50% reduction). Given bias-introducing factors, for example, comedication compliance with or without center effects, this study appears to be an effortful task.

This review provides sample-size estimates for standard designs in interventional cardiology such as 'test for differences' or 'noninferiority' designs of hard clinical and angiographic endpoints. In addition, other potential endpoints with the aim to minimize the number of patients for exploratory purposes are presented and discussed. The proposed endpoints are then gauged according to their reimbursement value.

## Methods and results

### Statistical analysis

All sample-size estimates were calculated with nQuery/nTerim version 2.0 (Statistical Solutions Ltd., Cork, Ireland). Test hypotheses were divided in terms of endpoints (clinical and surrogate) and test hypotheses (noninferiority and test for difference). For simplicity reasons we will refer to the 'test for difference' as a 'superiority' design.

### Reimbursement impact

Due to the plethora of reimbursement systems worldwide, a grossly simplified approach was chosen to determine the level of impact for selected study endpoints in the field of interventional cardiology. The National Institute of Health and Care Excellence (NICE) in the UK publishes technology appraisals in a number of evaluation pathways for major indications. Within the framework of coronary artery disease (CAD), NICE appraised a total of five technologies for DESs [NICE, 2010a], BMSs [NICE, 2011], bio-absorbable scaffolds [NICE, 2014a], drug-coated

balloons [NICE, 2010b] and CAD-relevant comedication [NICE, 2003]. These appraisals were investigated for the frequency of study endpoints that constituted the basis for their relative reimbursement impact in this review.

### Surrogate endpoint study designs

*Angiographic or test for difference (superiority).* In the most commonly used study design, an angiographic benefit in the treatment group is shown by a significantly lower late lumen loss [(LLL) in lesion or in segment]. This corresponds to the following test hypothesis:

$H_0$ : LLL in the treatment group is equal to or higher than the LLL in the control group.

$H_a$ : LLL in the treatment group is lower than the LLL in the control group.

*Angiographic or noninferiority.* Another common design is defined with a noninferiority test hypothesis

$H_0$ : LLL in the treatment group is higher than or equal to the LLL in the control group plus a noninferiority margin  $\Delta_{LLL}$ .

$H_a$ : LLL in treatment group is lower than the LLL in the control group plus a noninferiority margin  $\Delta_{LLL}$ .

*Alternative angiographic endpoints.* A single-group design with inpatient control measurements such as reference vessel mean lumen diameters in a nontarget vessel may also be used when a patient control group is not feasible. This, however, may only be justified when lumen changes in the entire coronary vasculature are not expected. Kleber and coworkers investigated the effect of positive remodeling in target lesions as compared with nontarget-vessel-reference vessel diameters as an inpatient control [Kleber *et al.* 2014]. They observed target-lesion lumen enlargement when compared with nontarget vessel mean lumen diameters within the same patient:

$H_0$ : The mean lumen diameter in the treated segment (or lesion) of the target vessel is lower or equal to the mean lumen diameter in the control segment (or lesion) of a nontarget vessel within the same patient.

$H_a$ : Mean lumen diameter in the treated segment (or lesion) of the target vessel is larger than the mean lumen diameter in the control segment (or lesion) of a nontarget vessel within the same patient.

*Clinical endpoint/superiority.* To show significantly lower event rates [e.g. target-lesion revascularization (TLR) or major adverse cardiac event (MACE)] the following test hypothesis is typically applied:

$H_0$ : The event rates  $\pi_1$  in the treatment group is higher or equal to the event rate  $\pi_2$  in the control group.

$H_a$ : The event rate  $\pi_1$  in the treatment group is lower than the event rate  $\pi_2$  in the control group.

*Clinical endpoint/non-inferiority.* Another common design is defined by the noninferiority test hypothesis:

$H_0$ : The event rate  $\pi_1$  in the treatment minus the event rate  $\pi_2$  in the control group is larger than or equal to the non-inferiority margin  $\Delta$ .

$H_a$ : The event rate  $\pi_1$  in the treatment minus the event rate  $\pi_2$  in the control group is smaller than the non-inferiority margin  $\Delta$ .

*Fractional-flow reserve/superiority.* Despite its lack of universal acceptance, a less common study design can be formulated based on a difference in fractional-flow reserve (FFR) between FFR values after the intervention and at the follow-up interval. Absolute FFR values can be obtained from the relevant literature for sample size estimates [Pijls *et al.* 2000; De Bruyne *et al.* 2012; Johnson *et al.* 2014]:

1.  $\Delta FFR_{\text{control group}} = FFR_{\text{control group post PCI}} - FFR_{\text{control group 6 months}}$
2.  $\Delta FFR_{\text{treatment group}} = FFR_{\text{treatment group post PCI}} - FFR_{\text{treatment group 6 months}}$ 
  - where  $\delta_{FFR}$  can be defined as 0.05 or 0.10 depending on the expected treatment effect.
  - $H_0$ :  $\Delta FFR_{\text{treatment group}} - \Delta FFR_{\text{control group}} \geq \delta_{FFR}$

that is, the investigational device is inferior to the device in the control group.

$H_a$ :  $\Delta FFR_{\text{treatment group}} - \Delta FFR_{\text{control group}} < \delta_{FFR}$

that is, the investigational device is not inferior to the device in the control group.

*Vasomotility and mean lumen diameter.* A potential endpoint to study endothelial cell regardless of smooth-muscle cell functionality can be defined in terms of drug-induced mean lumen diameter changes at baseline  $t_0$ , that is, prior to intracoronary drug administration and at time points  $t_1$  after, for example, acetylcholine (vasoconstrictive) and  $t_2$  after nitroglycerine (vasodilatory) intracoronary administration into the same target lesion [Brugaletta *et al.* 2012]. Potential test hypotheses are:

$H_0$ : The mean lumen diameter in the treated lesion at  $t_0$  is not significantly different as compared to the mean lumen diameter at  $t_1$  or alternatively at  $t_2$ .

$H_a$ : Mean lumen diameter in the treated lesion at  $t_0$  is different from the mean lumen diameter at  $t_1$  or alternatively different from the mean lumen diameter at  $t_2$ .

*Ratios of uncovered stent struts.* A percentage of uncovered struts between two treatment groups could also be defined as a potential safety endpoint [Adriaenssens *et al.* 2014]. Moreover, ratios of uncovered struts could conceptually be determined at different time points within one group to document the time course of stent coverage (e.g. 4 weeks and 3 months). The rationale of this approach would be that if there are less uncovered struts as compared with a control group, the risk of stent thrombosis (ST) would be reduced. Recently, this design was used by Karjalainen and coworkers, who used a noninferiority margin of 5% and a standard deviation for their measurements of 5% [Karjalainen *et al.* 2015].

The corresponding test hypotheses can be formulated as follows with the difference in uncovered struts and the following definitions:

$\Delta_{\text{uncovered struts}} = \text{percentage difference of uncovered stent struts between two treatment groups (BMS versus DES) or time points (4 weeks versus 3 months) where } \delta_{\text{strut coverage}} \text{ is the noninferiority margin.}$

$H_0$ :  $\Delta_{\text{uncovered struts}} \geq \delta_{\text{strut coverage}}$

$H_A$ :  $\Delta_{\text{uncovered struts}} < \delta_{\text{strut coverage}}$

**Table 1.** Angiographic surrogate endpoints.

Angiographic endpoint	Advantages	Disadvantages
Late lumen loss in lesion (in-lesion LLL)	Independent of differences in reference vessel diameters between treatment groups Well accepted Smaller study populations	Sensitive to projection mismatch Based on unpaired <i>t</i> test Hemodynamically indifferent
Late lumen loss in segment (in-segment LLL)	Independent of differences in reference vessel diameters between treatment groups Well accepted Smaller study populations	Sensitive to projection mismatch Based on unpaired <i>t</i> test Hemodynamically indifferent
Percent stenosis	Independent of differences in reference vessel diameters between treatment groups Well accepted Smaller study populations	Sensitive to projection mismatch Based on unpaired <i>t</i> test Hemodynamically indifferent
Binary restenosis rate	Correlates with LLL	May not have a clinical relevance Oculostenotic reflex may introduce bias in the target-lesion revascularization rate
Minimal lumen diameter	Intuitive	Dependent of differences in reference-vessel diameter between treatment groups Hemodynamically indifferent
Mean lumen diameter in lesion	May be used in nonrandomized trials with inpatient controls in nontarget vessels Is based on paired <i>t</i> test Hemodynamically sensitive	Not well accepted Not a direct outcome of QCA, i.e. lesion length must be considered

QCA, quantitative coronary angiography.

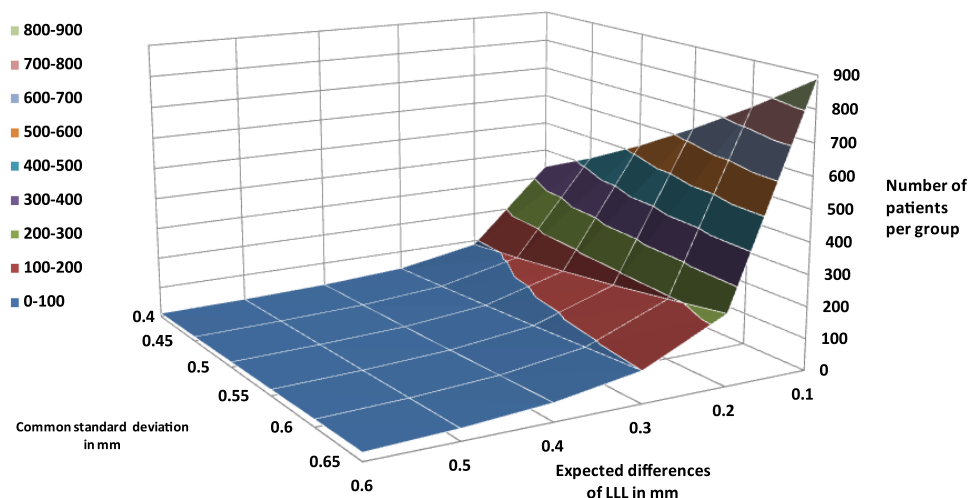
## Results

The most common angiographic endpoints are listed in Table 1. The most frequently used surrogate endpoint in interventional cardiology is late lumen loss (LLL) within the lesion or within a defined segment. In general, the larger the standard deviations of LLL, the more patients have to be recruited (Figure 1). With an expected difference in LLL of at least 0.3 mm, a superiority design would require less than 50 patients per treatment group independent of the assumed LLL standard deviations. However, the necessary number of patients can quickly increase into the 200–300 range if the expected LLL difference is in the range of 0.1 mm (top panel, Figure 1). The corresponding sample size three-dimensional plots are different for noninferiority and superiority designs (Figure 1: top *versus* bottom panel). While in the superiority design, a plateau of less than 100 patients per treatment group can be expected with a treatment effect of more than

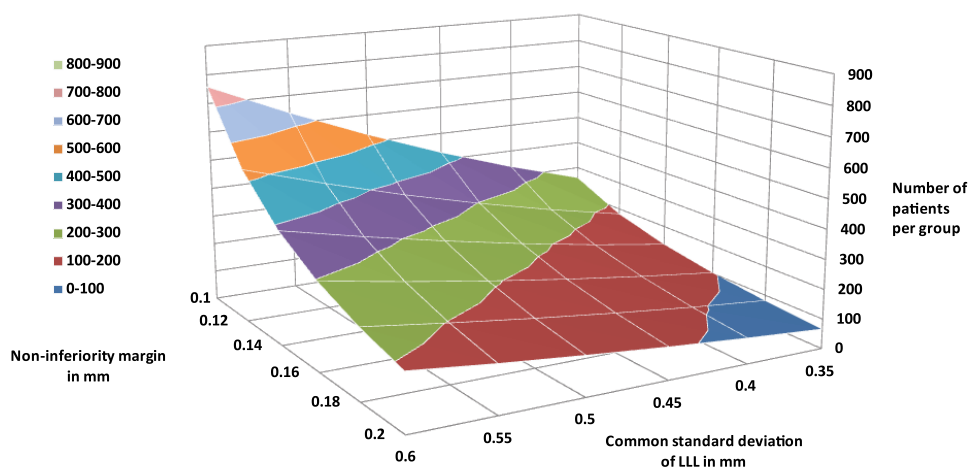
0.3 mm, the noninferiority surface curve has a more gradual incline towards smaller noninferiority margins and higher standard deviations. In the noninferiority study design, 200–300 patients per group for a noninferiority difference of 0.14–0.16 mm and LLL standard deviations of 0.5 mm are needed. Other angiographic surrogate endpoints are percent stenosis (PS); binary-restenosis rate (BR), minimal lumen diameter (MLD) and in-lesion mean lumen diameter.

Other surrogate endpoints (Table 2) are also described in the literature, such as FFR [Pijls *et al.* 2000; De Bruyne *et al.* 2012; Johnson *et al.* 2014], which can be defined as a pressure drop across a lesion. Due to the novelty of the ‘loss in FFR’ as a primary endpoint, sample-size estimates were calculated for various common standard deviations and noninferiority margins (Figure 2). With a common standard deviation of 0.1 and a noninferiority margin of 0.4, 100 patients per group are needed.

**Sample size estimates for a superiority design of late lumen loss with different expected differences and standard deviations (Power=90%, alpha=0.025, one-sided)**



**Sample size estimates for a non-inferiority design of late lumen loss with different non-inferiority margins and standard deviations (Power=90%, alpha=0.025)**



**Figure 1.** Sample-size estimates for a late lumen loss [LLL] superiority design with various common standard deviations and expected difference (top panel), and sample-size estimates for a noninferiority design with different noninferiority margins and commonly observed LLL standard deviations.

Furthermore, the percentage of uncovered struts as a primary endpoint has gained popularity since rapidly covered stent struts may have an advantage in terms of the reduced risk of stent thrombosis (ST) [Tahara *et al.* 2011]. This would translate to a reduced duration of dual antiplatelet therapy (DAPT), a benefit to the patient. Figure 3 illustrates

the number of patients per group in a noninferiority design for covered stent struts with various noninferiority margins and common standard deviations.

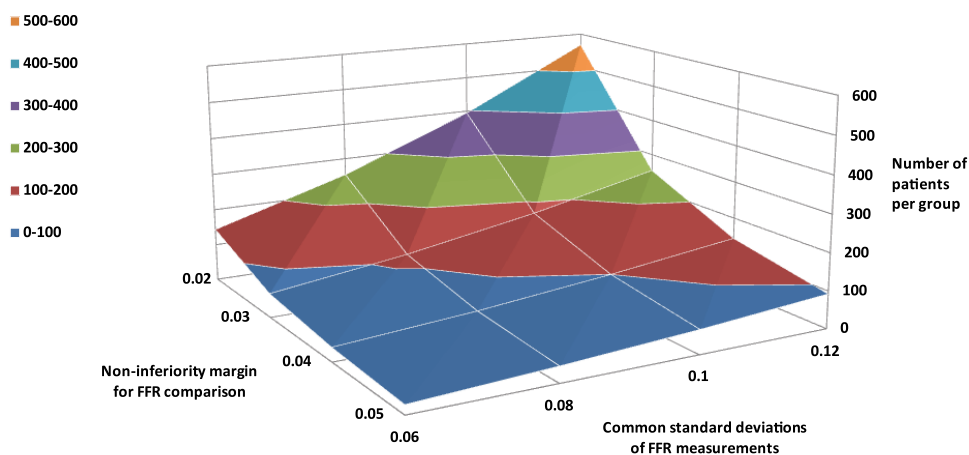
Table 3 lists clinical endpoints such as the TLR rate, MACE rate or the target-vessel failure (TVF) rate. ST and stroke rates may serve as

**Table 2.** Other surrogate endpoints.

Surrogate endpoint	Advantages	Disadvantages
Loss in FFR	Based on paired <i>t</i> test Smaller study populations Small measurement errors (small measurement standard deviations) Hemodynamically intuitive	Not well accepted 'Grey zones', i.e. nonlinear relationship between degree of stenosis and increase in perfusion (FFR > 0.90) Indifferent in terms of lesion location (proximal/distal) Provides only a binary answer (significant stenosis yes/no)
Percent uncovered struts assessed by OCT	Smaller study populations Interdevice comparisons or intradevice comparisons for different time points possible	Hemodynamically indifferent Only useful to interpret 'stent strut coverage' from a safety point view Not an efficacy marker
In lesion mean lumen and vasomotility	Only measurement for endothelial and smooth muscle cell functionality	Only useful for specific devices (BVS, DCB)

FFR, fractional flow reserve; OCT, optical coherence tomography; BVS, bioresorbable vascular scaffold; DCB, drug-coated balloon.

**Sample size estimates for a non-inferiority designs with FFR non-inferiority margins (0.02-0.05) and various standard deviations for FFR measurements (0.06-0.12) (Power=80%, one-sided alpha=0.025)**



**Figure 2.** Sample-size estimates for a noninferiority design with difference in fractional-flow reserve (FFR) as the primary endpoint for various noninferiority margins and common standard deviations in FFR measurements.

safety endpoints for selected patient populations. Figure 4 illustrates the needed patient numbers for TLR in a noninferiority design. These estimates also hold true for MACE and TVF rates. Given an 8% expected event rate in the treatment group and a noninferiority margin of 4%, the group size would need to be 440 patients. The largest patient groups are necessary if the ST rate is chosen in a noninferiority design. Given the already-low incidence rates in the sub-1% range and noninferiority margins ranging from

0.2–0.6%, thousands of subjects are required for a properly designed trial (Figure 5).

The summarized study-population sizes are complemented with the postulated reimbursement value (Table 4). Basis for the reimbursement impact were the aforementioned NICE appraisals in the field of CAD. In the searched appraisals, MACE was identified as the most frequent endpoint for studying the clinical evidence for a particular technology. Myocardial infarction (MI) as a

Sample size estimates for a non-inferiority design of % uncovered struts , different non-inferiority margins (2.0-3.0%) and various standard deviations % uncovered struts(3.0-6.0%) (Power=80%, one-sided alpha=0.025)

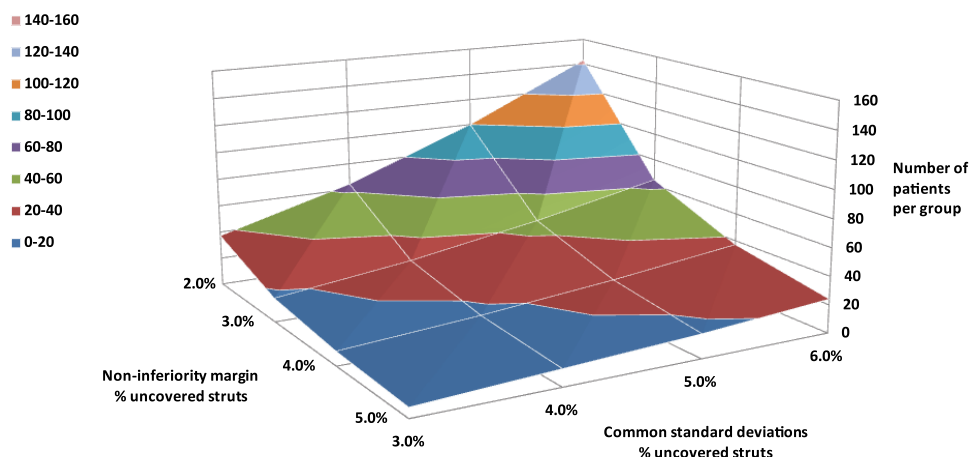


Figure 3. Sample-size estimates for a noninferiority design with percentage of uncovered stent rates as the primary endpoint, various noninferiority margins and common standard deviations.

Table 3. Clinical endpoints.

Clinical endpoint	Advantages	Disadvantages
MACE rate or TVF Rate	Highly relevant for reimbursement	In observational studies, MACE/TVF may be difficult to determine Larger population >500 patients
TLR (Re-PCI + CABG) rate or TVR (Re-PCI, CABG) in target vessel	Highly relevant for reimbursement	In observational studies, target lesion re-PCI may be difficult to determine
Cardiac death rate	Highly relevant for reimbursement	Meaningful test hypotheses require patient numbers >1000 patients per group Cause of death sometimes unclear
MI rate	Should be associated with TLR	Not always attributable to the treated target lesion or vessel
ST rate	Highly desirable for interdevice comparisons, e.g. DES	Meaningful test hypotheses require patient numbers >5000 patients per group
Stroke rate	Highly relevant for reimbursement	Meaningful test hypotheses require patient numbers >2000 patients per group May not be an outcome of the coronary revascularization procedure

CABG, coronary-artery bypass graft; DES, drug-eluting stents; MACE, major adverse cardiac event; TLR, target-lesion revascularization; Re-PCI, repeat percutaneous coronary intervention; TVR, target-vessel revascularization; TVF, target-vessel failure; MI, myocardial infarction; ST, stent thrombosis.

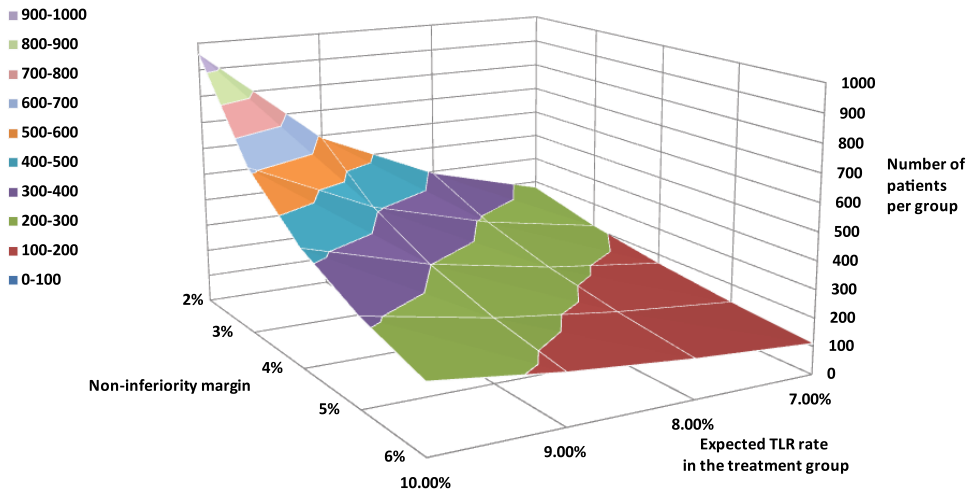
reimbursement-relevant endpoint was used in 17.6% of all referenced endpoints, followed by TLR, TVR and cardiac death, with 11.8%. Based on these findings, their relative impact was rated with 100% in the case of MACE, 80% in case of MI, and 70% for TLR/TVR and cardiac death (Table 4).

## Discussion

### Late lumen loss

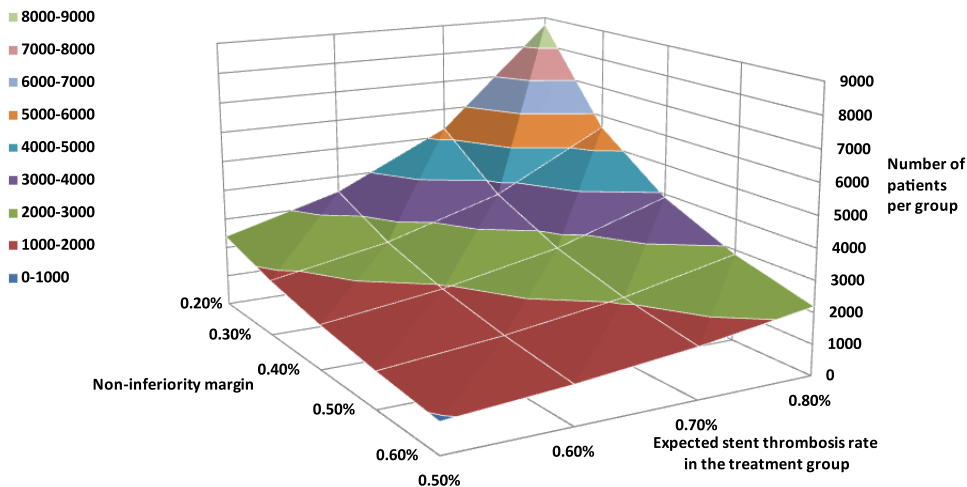
LLL, PS and MLD are based on edge-detection algorithms applied in coronary quantitative angiography (QCA), which requires a calibration based

**Sample size estimates for a non-inferiority design of TLR with 12.0% in the control group and different non-inferiority margins (2-6%) and expected TLR rates in the treatment group (7-10%) (Power=80%, one-sided alpha=0.025)**



**Figure 4.** Sample-size estimates for a noninferiority design in terms of target-lesion revascularization (TLR) as the primary endpoint, with various noninferiority margins and expected TLR rates in the treatment group.

**Sample size estimates for a non-inferiority design of stent thrombosis with 1.0% in the control group, different non-inferiority margins (0.2-0.6%) and expected stent thrombosis rates (0.5-0.8%) (Power=80%, one-sided alpha=0.025)**



**Figure 5.** Sample-size estimates for a noninferiority design with stent-thrombosis (ST) rates as the primary endpoint with various noninferiority margins and expected ST rates in the treatment group.

on the outer diameter of a contrast-dye-filled guiding-catheter tip (6F = 2.0 mm). Defined on this reference calibre, all diameters obtained from the shadow images of an angiogram are the foundation

to compute LLL, PS and MLD. The major methodological advantage is that a blinding relative to the treatment group can be elegantly performed when the device type cannot be recognized in the



**Table 4.** Study population sizes and postulated reimbursement value.

Endpoint	Average patient size per treatment group	Use of the endpoint by NICE in all pathways <sup>a</sup>	Postulated reimbursement value
MACE <sup>1</sup>	500	5 (29.4%)	100%
MI <sup>1</sup>	>500	3 (17.6%)	80%
TLR, TVR <sup>1</sup>	500	2 (11.8%)	70%
Cardiac death <sup>1</sup>	3000	2 (11.8%)	70%
Binary restenosis rate <sup>1</sup>	300	1 (5.9%)	50%
Late lumen loss <sup>1</sup>	100–200	1 (5.9%)	30%
ST rate	5000	0 (0.0%)	20%
FFR <sup>2</sup>	50–150	0 (0.0%)	20%
Mean lumen diameter	100–200	0 (0.0%)	0%
Minimal lumen diameter	100–200	0 (0.0%)	0%

\*As of February 2016.  
<sup>1</sup>Based on NICE appraisals TA152, IPG492, TA71, TA236 and MTG1.  
<sup>2</sup>Based on NICE appraisal MIB2.  
FFR, fractional flow reserve; MACE, major adverse cardiac event; MI, myocardial infarction; NICE, National Institute of Health and Care Excellence; ST, stent thrombosis; TLR, target-lesion revascularization; TVR, target-vessel revascularization.

angiogram. True blinding (investigator *versus* QCA operator) is therefore apparent for device studies, such as drug-coated balloon (DCB) *versus* POBA or BMS *versus* DES. In trials comparing DCB and DES, a systematic QCA error may concur given that a stent per definition has a sharper radiopaque edge, whereas a nonimplant angiogram may have a higher level of haziness, thereby leading to smaller MLDs. Based on the pioneering work by Mauri and coworkers, the mean LLL correlates well with binary-restenosis [Mauri *et al.* 2005b] rates but mean LLL is not a significant determinant of TLR [Mauri *et al.* 2005a]. Without questioning the scientific value of LLL in interventional trials that led to the paramount level of evidence we have at our disposal today, it must be mentioned that LLL has limited value from a reimbursement viewpoint. In addition, LLL may not be applicable when bio-absorbable vascular scaffolds (BVS) are studied due to their different time-dependent mechanical or biological behaviour in terms of recoil, vessel remodelling and neointimal changes [Bangalore and Mauri, 2009]. In these devices, after the degradation of the stent or scaffold postimplantation, late recoil may play a pivotal point that cannot be quantified with the standard angiographic LLL follow ups of 6–9 months. The major finding of our sample-size calculations revealed the ‘hopelessness’ to run head-to-head comparisons in terms of LLL, especially when the treatment effects are on a low level, in the 0.1–0.2 mm range.

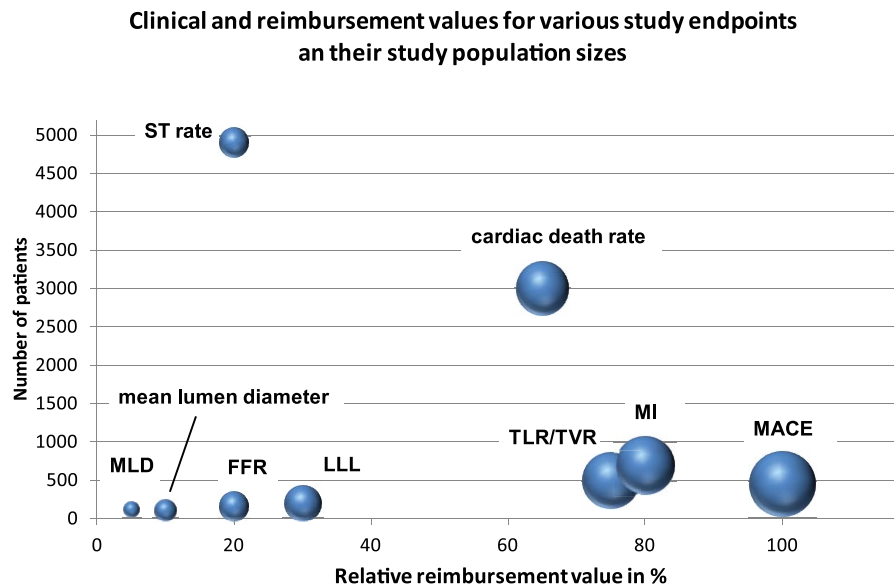
#### Minimal lumen diameter

The advantages and disadvantages of LLL measurements also apply to the MLD. However, it is our opinion that the MLD seems to be hemodynamically indifferent because it does not consider the lesion length that impacts on the translesion pressure drop. This has been established in the human domain by Brosh and coworkers using FFR [Brosh *et al.* 2005], and also in experimental animal models [Young *et al.* 1975]; the latter demonstrated that there is a lesion-length-dependent pressure gradient across artificially produced stenoses. We should also not ignore the Navier–Stokes equation [Sherman, 1990] that provides the theoretical basis for pressure gradients across narrowed fluid boundaries defined by the diameter and length of the stenosis.

#### Fractional-flow reserve endpoints

By intuition, not every pressure drop across a lesion manifests itself as hemodynamically or clinically important. However, due to the averaging algorithm during pullback of the FFR wire, measurement variations are very small and in the  $\pm 0.12$  mmHg range [Pijls *et al.* 2000].

Even though a difference in FFR as an endpoint seems intuitive, it ought to be pointed out that a mere pressure drop independent of the lesion location may not have the merit to determine clinical efficacy. Also, the fact that FFR cannot be



**Figure 6.** Number of patients *versus* relative reimbursement value and clinical value (size of the bubbles) for various endpoints.

FFR, fractional-flow reserve; LLL, late lumen loss; MACE, major adverse cardiac event; MI, myocardial infarction; MLD, minimal lumen diameter; ST, stent thrombosis; TLR, target-lesion revascularization; TVR, target-vessel revascularization.

interpreted in a linear fashion, that is, a  $\Delta$ FFR from 0.95 to 0.90 mmHg, is certainly less severe than a difference between 0.80 to 0.75 mmHg. Only when the lesion locations have the same impact on the coronary circulation, for example, with a lesion inclusion criterion such as the proximal left-anterior descending artery, may it be worthwhile to further explore FFR endpoints. Needless to say, the absolute FFR values at follow up should be above the established cut-off value of 0.80 [De Bruyne *et al.* 2012]. A recently published randomized trial comparing DCB and DES treatment in *de novo* lesions by Shin and coworkers revealed FFR values of  $0.85 \pm 0.08$  in DCB *versus*  $0.85 \pm 0.05$  in the stent group, which provides additional support that an FFR-based endpoint can be feasible even when the balloon dilatation is compared with a stent implantation, whereas the latter would certainly provide more postprocedural lumen gain [Shin *et al.* 2015].

#### Strut coverage

Strut coverage, by definition, can only provide an estimate on the temporal healing characteristics. It is, by definition, not an efficacy marker for vessel patency. Furthermore, it appears important that the formation of a functional endothelium can neither be expected nor truly documented with optical coherence tomography (OCT). In the current interventional cardiology community, there is an

unmet need to reduce the duration of DAPT. Nevertheless, given that a completely covered stent has a lower risk for ST, an estimate on how fast stent struts can be covered seems desirable. A novel healing-response endpoint that includes the ratios of uncovered and malapposed stent struts was proposed by the US National Institute of Health and elaborated by Räber and coworkers, which seems to include more healing properties than the mere number of uncovered struts [Räber *et al.* 2015]. Unfortunately, there is no defined cut-off value for stent-strut coverage that translates to a no-risk situation for ST. There are literature reports, however, that state when 90–95% of stent struts are covered, the risk of ST without DAPT is sufficiently reduced [Tahara *et al.* 2011]. Nonetheless, it is debatable if the percentage of covered stent struts is an adequate measure to assess the risk of late ST and its continued need for DAPT. In the case of late ST, neoatheroma formation may be the culprit without underlying mechanical factors such as strut coverage.

#### Clinical endpoints

One side aspect of this review was to investigate other attractive endpoints to keep patient numbers low in early proof-of-concept trials. Clinical endpoints are most likely not the first approach when conducting early studies. However, it is undisputable that from a clinical standpoint and

in terms of patient benefits, MACE, TVF or TLR rates are the most meaningful measures and the 'gold standard' for documenting efficacy. In this context, the importance to standardize clinical endpoints in coronary stent trials remains unchanged [Cutlip *et al.* 2007]. Recently, a large noninferiority margin of 4.5% with event rates of 7% was used to demonstrate noninferiority between resorbable scaffolds and DES by Ellis and coworkers [Ellis *et al.* 2015].

TLR, as the preferred efficacy endpoint was suggested by Silber and Herdeg, inasmuch as it establishes a direct cause-and-effect relationship between the lesion treatment and its failure [Silber and Herdeg, 2008].

#### Follow-up period

This review is an attempt to provide a first glance at study population sizes for preferred endpoints in patients with CAD. When biometric planning for a trial begins, comparable target populations and their expected outcomes for a specific treatment are elucidated first. If the number of patients is not manageable, study-budget-related questions gain importance. The provided estimates in this review may serve as a first guidance to eliminate possible trial designs that cannot be conducted with a reasonable number of patients. Often neglected, confounding factors may be introduced during the course of the follow up. By definition, once the study patients are discharged to the point of the follow up, there is a black box of confounding factors which may not be fully described. The patients' reluctance to adhere to their DAPT may increase the risk of ST [Iakovou *et al.* 2005] or the level of exercise can have an impact on the recurrence of MI. Without the intention to delve into the post-interventional psychological aspects, there are also established relationships between compliance with comedication and psychological support for the patient [Warner *et al.* 2013].

#### Reimbursement

To have a clinical or angiographic benefit of one treatment group is the first step towards demonstrating clinical efficacy. However, which incremental benefit is worth additional spending in a particular healthcare system? We will have difficulties explaining a LLL improvement of 0.1 mm to a health-insurance provider, unless these angiographic results translate into clinical improvements. To our knowledge, there is no accepted methodology

available that gauges the reimbursement value for various endpoints as a function of the number of patients. A first step on this terrain is attempted with the endpoints illustrated in Figure 6 on the basis of the sample-size estimate in Figures 1–5. We are aware that this is a first crude attempt to relate endpoints, their sample-size estimates and their frequency of use in NICE appraisals to formulate a relative reimbursement value in %.

Given that TLR directly correlates with the treatment success or efficacy, as previously suggested [Räber *et al.* 2015], the reimbursement value of 70% seems adequate, whereas MACE remains the gold standard. In addition, TLR rates were the most useful endpoints of the main Markov transition-model inputs in a cost-efficacy study conducted by NICE [NICE, 2010b] and Bonaventura and coworkers, comparing DES and DCB angioplasty [Bonaventura *et al.* 2012]. The reimbursement value for the rate of ST as a primary endpoint is rather limited (20%) inasmuch as it is not an efficacy endpoint but a safety endpoint with a low incidence rate of 0.5–2.0% at 1 year. Despite this semiquantitative and certainly somewhat subjective attempt to portray these complex relationships of reimbursement value, clinical importance and study-population sizes, it seems obvious that angiographic endpoints are not the favourites in this context. Cardiac death is also of high interest from a reimbursement point of view. Nevertheless, there is the ethical dilemma due to fact that there are no additional treatment costs when this endpoint is reached.

In this review, we refrained from quality-of-life endpoints, even though they have been used in various patient subsets, such as patients undergoing aortic-valve replacements [Tully *et al.* 2015]. In multimorbid patients, by nature, multiple factors contribute to the overall health status so that causal relationships between the primary treatment and the overall health status are difficult to determine.

#### Limitations

The estimated number of patients for a given test hypothesis are based on available literature references. They do not replace a properly conducted sample-size calculation for a particular design. Furthermore, most endpoints, such as those related to safety and efficacy, are not interchangeable for a given target population and the objective of a trial.

To gauge reimbursement values, each endpoint was based on NICE appraisals only and should be expanded to other cost-benefit analyses.

## Conclusions

In terms of reimbursement value, the primary endpoints MACE and TLR remain the best choice for a moderately sized study population of 500 patients per group. The angiographic endpoint MLD does not reflect all aspects of the hemodynamic environment distal of the lesion, and appears to be of low-reimbursement impact. Even though it is desirable to refrain from angiographic endpoints from a reimbursement standpoint, other surrogate endpoints such as difference in FFR or strut coverage with smaller study populations may merit further explorations for proof-of-concept studies. Nevertheless, the emerging endpoints, such as loss in FFR or stent coverage, without clearly established clinical benefits, are not useful for reimbursement purposes. An inpatient angiographic endpoint such as the mean lesion diameter in the target lesion as compared with a nontarget lesion may be useful if randomization is not possible.

## Acknowledgements

We would like to express our gratitude to Viktor Breul (Medical Scientific Affairs, Aesculap AG, Tuttlingen, Germany) for his highly appreciated statistical support and expertise to verify the statistical basis of this review. We also wish to acknowledge Dr Christian Sperling (Medical Scientific Affairs, B. Braun Melsungen AG, Berlin, Germany) for his shared expertise.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Conflict of interest statement

The authors declare that there is no conflict of interest.

## References

Adriaenssens, T., Dens, J., Ughi, G., Bennett, J., Dubois, C., Sinnaeve, P. *et al.* (2014) Optical coherence tomography study of healing characteristics of paclitaxel-eluting balloons vs. everolimus-eluting stents for in-stent restenosis: the SEDUCE (safety and efficacy of a drug eluting balloon in coronary artery

restenosis) randomised clinical trial. *EuroIntervention* 10: 439–448.

Bangalore, S. and Mauri, L. (2009) Late loss in a disappearing frame of reference: is it still applicable to fully absorbable scaffolds? *EuroIntervention* 15: F43–F48.

Bonaventura, K., Leber, A., Sohns, C., Roser, M., Boldt, L., Kleber, F. *et al.* (2012) Cost-effectiveness of paclitaxel-coated balloon angioplasty and paclitaxel-eluting stent implantation for treatment of coronary in-stent restenosis in patients with stable coronary artery disease. *Clin Res Cardiol* 101: 573–584.

Brosh, D., Higano, S., Lennon, R., Holmes, D., Jr. and Lerman, A. (2005) Effect of lesion length on fractional flow reserve in intermediate coronary lesions. *Am Heart J* 150: 338–343.

Brugaletta, S., Heo, J., Garcia-Garcia, H., Farooq, V., van Geuns, R., De Bruyne, B. *et al.* (2012) Endothelial-dependent vasomotion in a coronary segment treated by ABSORB everolimus-eluting bioresorbable vascular scaffold system is related to plaque composition at the time of bioresorption of the polymer: indirect finding of vascular reparative therapy? *Eur Heart J* 33: 1325–1333.

Cutlip, D., Windecker, S., Mehran, R., Boam, A., Cohen, D., van Es, G. *et al.* and Academic Research Consortium (2007). Clinical end points in coronary stent trials: A case for standardized definitions. *Circulation* 115: 2344–2351.

De Bruyne, B., Pijls, N., Kalesan, B., Barbato, E., Tonino, P., Piroth, Z. *et al.* (2012) FAME 2 trial investigators. Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease. *N Engl J Med* 367: 991–1001.

Ellis, S., Kereiakes, D., Metzger, D., Caputo, R., Rizik, D., Teirstein, P. *et al.* (2015) Everolimus-eluting bioresorbable scaffolds for coronary artery disease. *N Engl J Med* 373: 1905–1915.

Iakovou, I., Schmidt, T., Bonizzoni, E., Ge, L., Sangiorgi, G., Stankovic, G. *et al.* (2005) Incidence, predictors, and outcome of thrombosis after successful implantation of drug-eluting stents. *JAMA* 29: 2126–2130.

Johnson, N., Tóth, G., Lai, D., Zhu, H., Açar, G., Agostoni, P. *et al.* (2014) Prognostic value of fractional flow reserve: linking physiologic severity to clinical outcomes. *J Am Coll Cardiol* 64: 1641–1654.

Karjalainen, P., Varho, V., Nammias, W., Mikkelsen, J., Pietilä, M., Ylitalo, A. *et al.* (2015) Early neointimal coverage and vasodilator response following biodegradable polymer sirolimus-eluting vs. durable polymer zotarolimus-eluting stents in patients with acute coronary syndrome –HATTRICK-OCT trial. *Circ J* 79: 360–367.

- Kleber, F., Schulz, A., Waliszewski, M., Hauschild, T., Böhm, M., Dietz, U. *et al.* (2014) Local paclitaxel induces late lumen enlargement in coronary arteries after balloon angioplasty. *Clin Res Cardiol* 104: 217–225.
- Mauri, L., Orav, E., Candia, S., Cutlip, D. and Kuntz, R. (2005a) Robustness of late lumen loss in discriminating drug-eluting stents across variable observational and randomized trials. *Circulation* 112: 2833–2839.
- Mauri, L., Orav, E. and Kuntz, R. (2005b) Late loss in lumen diameter and binary restenosis for drug-eluting stent comparison. *Circulation* 111: 3435–3442.
- NICE (2010a) NICE guidance TA152: Drug-eluting stents for the treatment of coronary artery disease. Available at: <http://www.nice.org.uk/guidance/ta152>
- NICE (2014a) NICE guidance IPG492: Bioresorbable stent implantation for treating coronary artery disease. Available at: <http://www.nice.org.uk/guidance/ipg492>
- NICE (2011) NICE guidance TA71: Guidance on the use of coronary artery stents. Available at: <http://www.nice.org.uk/guidance/ta71>
- NICE (2003) NICE guidance TA236: Ticagrelor for the treatment of acute coronary syndromes. Available at: <http://www.nice.org.uk/guidance/TA236>
- NICE (2010b) NICE guidance MTG1: SeQuent Please balloon catheter for in-stent coronary restenosis. Available at: <http://www.nice.org.uk/guidance/MTG1>
- NICE (2014b) NICE guidance MIB2: The Pressure Wire fractional flow reserve measurement system for coronary artery disease. Available at: <http://www.nice.org.uk/advice/mib2>
- Pijls, N., De Bruyne, B., Bech, G., Liistro, F., Heyndrickx, G., Bonnier, H. *et al.* (2000) Coronary pressure measurement to assess the hemodynamic significance of serial stenoses within one coronary artery: validation in humans. *Circulation* 102: 2371–2377.
- Räber, L., Onuma, Y., Brugaletta, S., Garcia-Garcia, H., Backx, B. Iñiguez, A. *et al.* (2015) Arterial healing following primary PCI using the Absorb everolimus-eluting bioresorbable vascular scaffold (Absorb BVS) *versus* the durable polymer everolimus-eluting metallic stent (XIENCE) in patients with acute ST-elevation myocardial infarction: rationale and design of the randomised TROFI II study. *EuroIntervention* 13 August 2015 [Epub ahead of print]
- Serruys, P., de Jaegere, P., Kiemeneij, F., Macaya, C., Rutsch, W., Heyndrickx, G. *et al.* (1994) A comparison of balloon-expandable-stent implantation with balloon angioplasty in patients with coronary artery disease. Benestent Study Group. *N Engl J Med* 331: 489–495.
- Sherman, F. (1990) *Viscous flow*. New York: McGraw-Hill.
- Shin, E., Ann, S., Balbir Singh, G., Lim, K., Kleber, F. and Koo, B. (2015) Fractional flow reserve-guided paclitaxel-coated balloon treatment for de novo coronary lesions. *Catheter Cardiovasc Interv* 1 October 2015. [Epub ahead of print]
- Silber, S. and Herdeg, C. (2008) [Drug-eluting stents for diabetic patients. A critical appraisal of the currently available data from randomized trials]. *Herz* 33: 196–205.
- Tahara, S., Chamié, D., Baibars, M., Alraies, C. and Costa, M. (2011) Optical coherence tomography endpoints in stent clinical investigations: strut coverage. *Int J Cardiovasc Imaging* 27: 271–287.
- Tully, P., Roshan, P., Rice, G., Sinhal, A., Bennetts, J. and Baker, R. (2015) Change in quality of life after transcatheter aortic valve implantation and aortic valve replacement surgery in Australian patients aged  $\geq 75$  years: the effects of EuroSCORE and patient operability. *J Geriatr Cardiol* 12: 30–36.
- Warner, L., Schüz, B., Aiken, L., Ziegelmann, J., Wurm, S., Tesch-Römer, C. *et al.* (2013) Interactive effects of social support and social conflict on medication adherence in multimorbid older adults. *Soc Sci Med* 87: 23–30.
- Young, D., Cholvin, N. and Roth, A. (1975) Pressure drop across artificially induced stenoses in the femoral arteries of dogs. *Circ Res* 36: 735–743.