

# Surrogate and clinical endpoints for studies in peripheral artery occlusive disease: Are statistics the brakes?

Waliszewski M<sup>1</sup>, Redlich U<sup>2</sup>, Breul V<sup>3</sup> and Tautenhahn J<sup>2</sup>

<sup>1</sup>Medical Scientific Affairs, B. Braun Melsungen AG, Berlin, Germany

<sup>2</sup>Klinikum Magdeburg, Magdeburg, Germany

<sup>3</sup>Medical Scientific Affairs, B. Braun Aesculap AG, Tuttlingen, Germany

Article history: Received: 02.01.2017 Accepted: 27.02.2017 Published: 30.04.2017

## ABSTRACT:

**Background:** The aim of this review is to present the available clinical and surrogate endpoints that may be used in future studies performed in patients with peripheral artery occlusive disease (PAOD). Importantly, we describe statistical limitations of the most commonly used endpoints and offer some guidance with respect to study design for a given sample size. The proposed endpoints may be used in studies using surgical or interventional revascularization and/or drug treatments.

**Methods:** Considering recently published study endpoints and designs, the usefulness of these endpoints for reimbursement is evaluated. Based on these potential study endpoints and patient sample size estimates with different non-inferiority or tests for difference hypotheses, a rating relative to their corresponding reimbursement values is attempted.

**Results:** As regards the benefit for the patients and for the payers, walking distance and the ankle brachial index (ABI) are the most feasible endpoints in a relatively small study samples given that other non-vascular impact factors can be controlled. Angiographic endpoints such as minimal lumen diameter (MLD) do not seem useful from a reimbursement standpoint despite their intuitiveness. Other surrogate endpoints, such as transcutaneous oxygen tension measurements, have yet to be established as useful endpoints in reasonably sized studies with patients with critical limb ischemia (CLI).

**Conclusions:** From a reimbursement standpoint, WD and ABI are effective endpoints for a moderate study sample size given that non-vascular confounding factors can be controlled.

## KEYWORDS:

peripheral artery occlusive disease, clinical endpoints, surrogate endpoints, biometric estimates

## WHAT THIS PAPER ADDS

Studies in patients with PAOD use a plethora of various endpoints which may not be relevant to quality of life of patients. This review provides a roadmap for trial design with an overview of potential clinical and surrogate endpoints, their requirement for sample size, and their relevance to reimbursement.

## INTRODUCTION

Peripheral artery occlusive disease (PAOD) is a manifestation of systemic atherosclerosis in the lower extremities and is associated with a significant mortality and morbidity, especially in the elderly.<sup>1-3</sup> It is of paramount importance that patients with intermittent claudication (IC) and critical limb ischemia (CLI) be recognized as different populations that should be studied separately.<sup>4</sup>

While there are a number of available diagnostic tools such as computed tomography angiography (CTA), magnetic resonance angiography (MRA), ultrasound Doppler or duplex Doppler (USD), and digital subtraction angiography (DSA), other markers are also available that gauge the severity of PAOD.

Before presenting individual endpoints, one should mention that there are initiatives to standardize study endpoints as ambiguous definitions do not permit inter-study comparisons among peripheral endovascular studies.<sup>4-5</sup>

Fleming and Powers describe a relevant primary endpoint as a measure of how the patient “feels, functions and survives”.<sup>6-7</sup> In this context, survival rate, disease exacerbation or absence of clinical events, and subjective health quality are viewed as pertinent indicators of

clinical success. The most commonly used measure of treatment success in the peripheral vascular bed are patency, occlusion rates<sup>8</sup>, and target lesion revascularization<sup>9</sup> (TLR) rates. Even though the ankle-brachial index<sup>2,10</sup> (ABI) is mostly used as a variable for stratification or risk classification, it could potentially be used as a primary endpoint as well. The list of meaningful surrogate endpoints includes walking distance<sup>11</sup> (WD) that may be defined as maximum walking distance (MWD) or pain-free walking distance (PFWD). Recently, transcutaneous oxygen tension measurements (TcPO<sub>2</sub>) in revascularized tissues<sup>12</sup>, and more specifically in patients with CLI<sup>13</sup>, have also been considered as potential endpoints. Moreover, in patients with advanced PAOD with Rutherford class 6, amputation<sup>14</sup> or mortality rates<sup>2</sup> as well as wound healing surrogate endpoints<sup>15</sup>, such as wound area post-revascularization, may serve as pertinent study endpoints. Quality of life endpoints have also been used to assess treatment success in patients with IC in RCTs.<sup>16</sup>

## METHODS

We hypothesize that there are differences in required patient numbers for endpoints of equal clinical importance. The purpose of this review is to describe, evaluate, and compare potential primary study endpoints for PAOD trials in terms of their clinical value, reimbursement relevance, and corresponding study sample sizes that ensure sufficient statistical power.

Furthermore, we hypothesize that there are reimbursement-relevant endpoints that can be chosen for future studies without the need to use large sample sizes. To better explain these complex relationships, a graphic comparison is attempted to gauge the given endpoints and their clinical/reimbursement value based on a semi-quantitative scoring system.

Because of its high relevance, comprehensiveness, and a recent publication date, the systematic review by Simpson et al.<sup>17</sup> was used to select PAOD-related study endpoints to estimate the number of patients required by particular study designs. Simpson and co-workers included patency, restenosis measures, need for re-intervention, generic measures of quality of life (QoL), clinical status, exercise tolerance, walking distance, limb salvage, complications, and adverse events as relevant study endpoints.

## STATISTICAL ANALYSIS

The numbers of patients required per study design were estimated with nQuery/nTerim, version 2.0 (Statistical Solutions Ltd. Cork, Ireland). Depending on the dependent variables (dichotomous, continuous), various biometric algorithms were used. The tested hypotheses were divided in terms of clinical/surrogate endpoints and non-inferiority and superiority designs (test for difference).

## ENDPOINT DESIGNS

### Dichotomous event rates

The most common dichotomous event rate is the patency rate (primary, primary assisted, or secondary) which can also be interpreted as 100% patency rate minus the occlusion rate. As defined by Diehm et al.<sup>4-5</sup>, patency should not be interpreted as a non-occlusion rate but rather as freedom from binary restenosis, i.e. diameter stenosis of less than 50% either measured by quantitative angiography or by ultrasound. Other dichotomous event rates can be amputation or infection rates, which may not be the first choice in the general PAOD population.

Common designs to show a significantly better patency rate are described with the following test hypotheses:

$H_0$ : Patency rate  $\pi_1$  in the treatment group is lower than or equal to the patency rate  $\pi_2$  in the control group

$H_a$ : Patency rate  $\pi_1$  in the treatment group is higher than the patency rate  $\pi_2$  in the control group

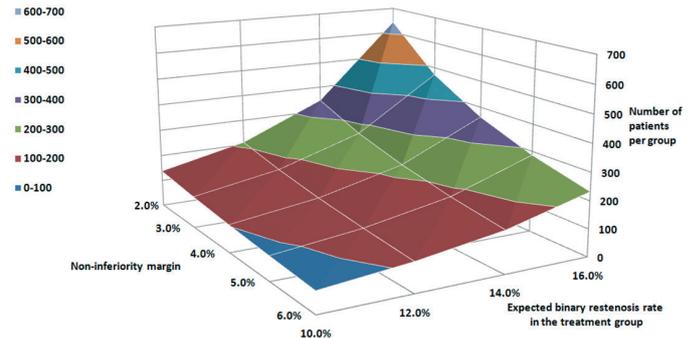
A non-inferiority test hypothesis is formulated as follows:

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

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

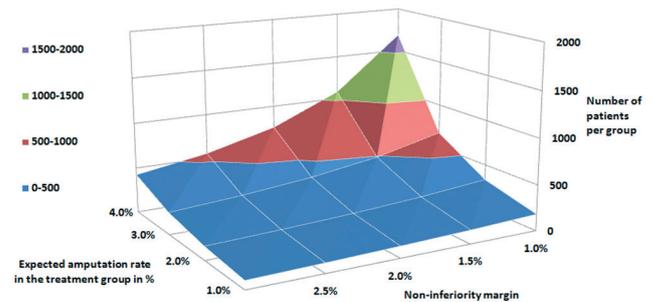
## ANGIOGRAPHIC ENDPOINT/TEST FOR DIFFERENCE

Differences in interventional treatment groups are often described with late lumen loss (LLL) either in-lesion or in-segment, which are obtained from baseline and follow-up angiograms at a pre-specified post-procedural time point.



**Fig. 1.** Sample size estimates for a test-of-difference (superiority) design with a range of binary restenosis rates in both treatment groups (top panel) and sample size estimates for a non-inferiority design with different non-inferiority margins, a higher binary restenosis rate in the control group (by 4%), and expected differences in binary restenosis rates (bottom panel).

**Biometric estimates for a non-inferiority design of amputation rates with 5.0% in the control group, different non-inferiority margins (1.0-3.0%) and expected amputation rates (1.0-4.0%) (Power=80%, one-sided alpha=0.025)**



**Fig. 2.** Sample size estimates for a non-inferiority design with different non-inferiority margins (3.0%-1.0%), a 5.0% amputation rate in the control group and different expected amputation rates in the treatment group (to be defined in minor, major, BTK, and ATK).

$H_0$ :  $LLL_1$  in the treatment group is equal to or higher than the  $LLL_2$  in the control group

$H_a$ :  $LLL_1$  in treatment group is lower than the  $LLL_2$  in the control group

### Angiographic endpoint /non-inferiority

The non-inferiority test hypothesis in terms LLL can be expressed as:

$H_0$ :  $LLL_1$  in the treatment group minus the  $LLL_2$  in the control group is larger than or equal to a non-inferiority margin  $\Delta_{LLL}$

$H_a$ :  $LLL_1$  in the treatment group minus the  $LLL_2$  in the control group is smaller than a non-inferiority margin  $\Delta_{LLL}$

## Surrogate endpoint ankle brachial index (ABI)/ test for difference

Positive changes in the hemodynamic environment, i.e. lower pressure drop across a target lesion, can be assessed with the ankle brachial index (ABI). The most feasible statistical approach is to observe post-interventional/post-surgical ABIs in two treatment groups (superiority design) or ABIs in one treatment group

Tab. I. Clinical Endpoints

CLINICAL ENDPOINT	ADVANTAGES	DISADVANTAGES
Patency rates (freedom from binary restenosis, i.e. $\geq 50\%$ ) <sup>5</sup>	-highly relevant from a clinical perspective -relevant for reimbursement -may also be assessed without fluoroscopy, i.e. Doppler/ Duplex Ultrasound, MRI	-large patient numbers necessary -difficult inter-study comparisons due to various definitions -often interpreted as 100% occlusion rate in vascular reconstructions
Amputation rates (must be grouped into minor=below the ankle, major=above the ankle, BTK=below the knee, ATK=above the knee)	-highly relevant in critical limb ischemia (CLI) -highly relevant for reimbursement	-large patient numbers necessary -not necessary useful in overall PAOD cohorts
Target lesion revascularization rate	-highly relevant from a reimbursement perspective -ideal for cost-efficacy studies	-may require larger study populations unless a larger difference in treatment success is suspected
Major adverse limb events = MALE= TLR + amputation	-highly relevant from a clinical perspective -ideal for cost-efficacy studies	-may require larger study populations unless a larger difference in treatment success is suspected
Mortality rates	-highly relevant from a clinical perspective -commonly accepted endpoint for patients with CLI	-large patient numbers necessary -not necessary useful in general PAOD cohorts
Revision rates	-highly relevant from a clinical and payer's perspective	-may reflect the overall morbidity of the study population
Infection rates	-highly relevant in challenged PAOD populations with graft infection promoting risk factors	-only applicable in challenged populations e.g. those at risks for infections -a very large patient population is necessary
Rutherford category shift	-provides a good overview of the entire study population	-walking ability may also be affected by musculo- skeletal events/conditions

Tab. II. Angiographic Surrogate Endpoints

ANGIOGRAPHIC ENDPOINT	ADVANTAGES	DISADVANTAGES
Late lumen loss in lesion	-independent of reference vessel diameter -well accepted -smaller study populations	-sensitive to projection mismatch -based on unpaired t-test -hemodynamically indifferent
Late lumen loss in segment	-independent of reference vessel diameter -well accepted -smaller study populations	-sensitive to projection mismatch -based on unpaired t-test -hemodynamically indifferent
Percent stenosis (binary restenosis as % of all patients with a percent restenosis of $\geq 50\%$ )	-independent of reference vessel diameter -well accepted -smaller study populations	-sensitive to projection mismatch -based on unpaired t-test -hemodynamically indifferent
Minimal lumen diameter	-intuitive -direct outcome of quantitative angiography	-does not consider reference vessel diameter -hemodynamically indifferent
Mean lumen diameter in lesion	-may be used in non-randomized trials with intra-patient controls in non-target vessels -is based on paired t-test -hemodynamically sensitive	-not well accepted

at different time points (non-inferiority design). The corresponding superiority test hypothesis is as follows:

$H_0$ : ABI in the treatment is equal to or lower than ABI in the control group

$H_a$ : ABI in treatment group is higher than ABI in the control group

### Surrogate endpoint walking distance/ test for difference

Differences in interventional treatment groups can also be described in terms of walking distance (WD), which is ideally obtained prior to the intervention, post-intervention/post-surgery, and at various follow-up intervals. For hypothesis testing, it is irrelevant if the maximum or pain-free WD are used as the primary endpoint:

$H_0$ : WD in the treatment group is equal to or shorter than WD in the control group

$H_a$ : WD in treatment group is longer than WD in the control group

### Surrogate endpoint/test for difference in $TcCO_2$

For a difference in transcutaneous  $O_2$  tension, e.g. in patients with critical limb ischemia (CLI), the following test hypothesis can be formulated:

$H_0$ :  $TcPO_2$  in the treatment group is equal to or lower than  $TcPO_2$  in the control group

$H_a$ :  $TcPO_2$  in the treatment group is higher as compared to the one in the control group

Alternatively, a single-arm study can be performed with measurements before, immediately after, and at a later time point after the treatment to assess oxygen saturation of the tissue distal to the treated lesion. In this scenario, the test hypothesis would be:

$H_0$ :  $TcPO_2$  at time point  $t_n$  is equal to or lower than  $TcPO_2$  at time point  $t_{n-1}$

$H_a$ :  $TcPO_2$  at time point  $t_n$  is higher as compared to  $TcPO_2$  at time point  $t_{n-1}$

where  $t_n$  is the time point after  $t_{n-1}$ .

### Rutherford category shift endpoint/ test for difference

The Rutherford category shift entails that 7 different Rutherford categories (0=no claudication to 6=severe limb ischemia with tissue loss) are translated into a set of ordinal data. When comparing two time points (e.g. pre-treatment vs. follow-up), the signed-rank test may be considered:

$H_o$ : There is no improvement in the Rutherford categories between the post-treatment condition and at the follow-up

$H_a$ : Rutherford categories improve at the follow-up interval as compared to the baseline condition.

### Quality of life endpoint/ test for difference

Patients' responses on quality-of-life (QOL) questionnaires must be transferred to a score either per treatment group or in one treatment group at various time points. This may be translated into the following test hypotheses:

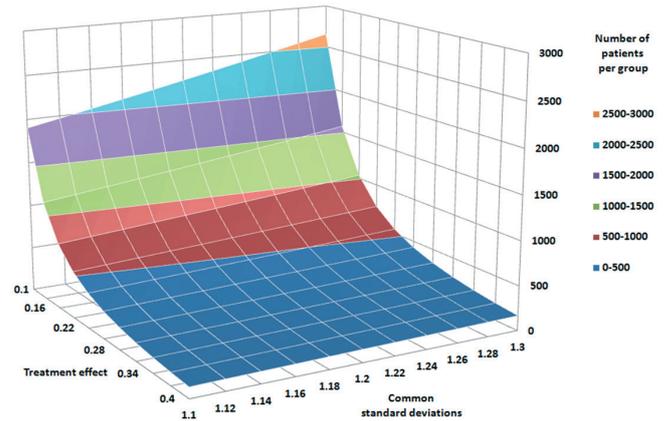
$H_o$ : The  $QOL_{treatment}$  score in the treatment group is equal to or lower than that in the control group

$H_a$ : The  $QOL_{treatment}$  in treatment group is higher than the  $QOL_{control}$  score in the control group

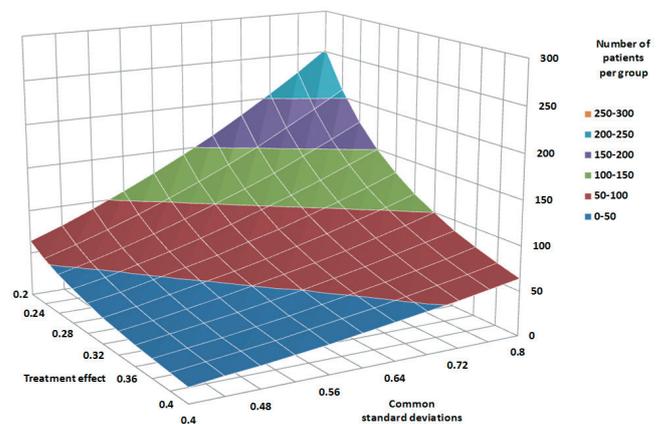
## RESULTS

Tables 1 to 4 summarize the most commonly used endpoints which correspond directly or indirectly to the clinical outcomes measures proposed by Simpson et al.<sup>17</sup> These endpoints, with their respective advantages and disadvantages, are listed in the aforementioned tables. For selected endpoints, the required patient numbers per treatment group with a significance level  $\alpha$  and a power of 80% were calculated.

The most commonly used endpoints in the general PAOD population are patency and restenosis rates<sup>17</sup> since they are directly related to treatment efficacy. In vascular reconstructions, patency is typically defined as 100% minus the occlusion rate, whereas in the endovascular studies the binary restenosis (BR) rate, most commonly defined as a 50% diameter loss, seems to be used most commonly. Figure 1 displays patient population sizes for various BR rate assumptions in the treatment and the control groups. Naturally, the smaller is the expected difference in BR rates, the larger will be the required number of patients per group. For instance, if we assume a BR rate of 30.0% in the control group and 20.0% in the treatment group, we would require 313 patients per group (one-sided  $\alpha=2.5\%$ , power=80%). Alternatively, assuming a non-inferiority design with an BR rate of 20.0% in the control group, an expected BR rate in the treatment group of 16.0%, and a non-inferiority margin of 4.0%, a total of 362 patients per group (one-sided  $\alpha=2.5\%$ , power=80%) would be needed (Figure 1). Overall, with a double digit difference between two event rates, several hundred subjects would be needed per group to reject the null hypothesis, i.e. a benefit in the treatment group.



**Fig. 3.** Sample size estimates for a test-of-difference design with a Rutherford category shift of 1.6 in the control group, various increases in Rutherford category shifts (treatment effect) in the treatment group (0.1-0.4), and different standard deviations (1.1-1.3) identical for both groups (Power=80%, one-sided  $\alpha=0.025$ )

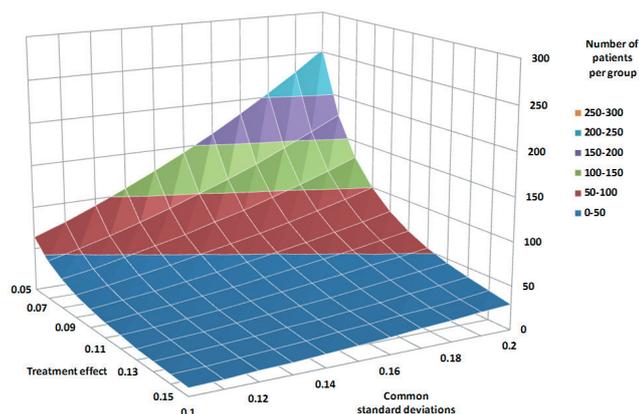


**Fig. 4.** Sample size estimates for a test-of-difference design of late lumen loss (LLL) with 0.4 mm in the treatment group, different standard deviations (0.4-0.7 mm) identical for both groups, and LLL increases (treatment effect) in the control group (0.2-0.4 mm) (Power=80%, one-sided  $\alpha=0.025$ )

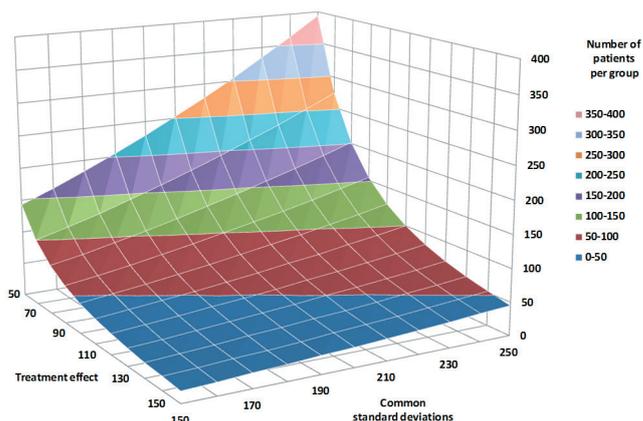
A common clinical endpoint in patients with advanced Rutherford categories (4-6) or CLI is the amputation rate. This endpoint, in turn, can also be included in a composite endpoint such as major adverse limb events (MALE). When limb salvage is of primary interest, a non-inferiority design (see Figure 2) with an amputation rate of 5.0% in the control group, 4.0% in the treatment group, and a non-inferiority margin of only 1.0% will require 1686 patients per group to reject the null hypothesis (one sided  $\alpha=2.5\%$ , power=80%).

The numbers of patients in a trial with Rutherford category shift as the primary endpoint are illustrated in Figure 3. Expected ranges can be obtained from a DCB study that was conducted by Scheinert and coworkers.<sup>18</sup> In this case, the difference in Rutherford categories at follow-up as compared to the post-interventional Rutherford categories are used as an efficacy marker. The 3D curve in Figure 3 has a flat appearance for most of the assumed Rutherford category shift, up to an expected difference of 0.22 in the treatment group. Conversely, large patient numbers are necessary if the expected difference in Rutherford category shift is lower than 0.22 in the treatment group.

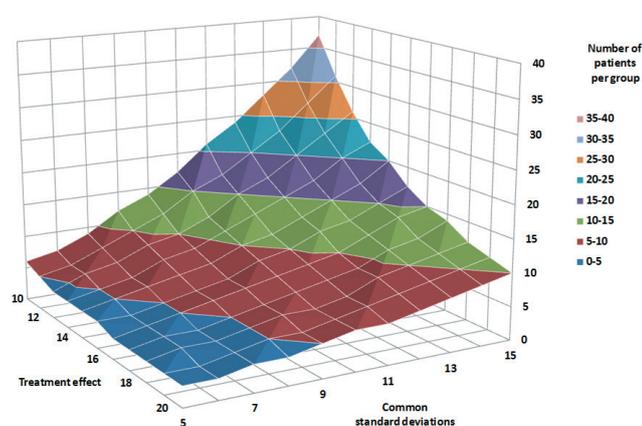
While patency, occlusion, and binary restenosis rates are the most commonly used endpoints in studies with surgical revascularization



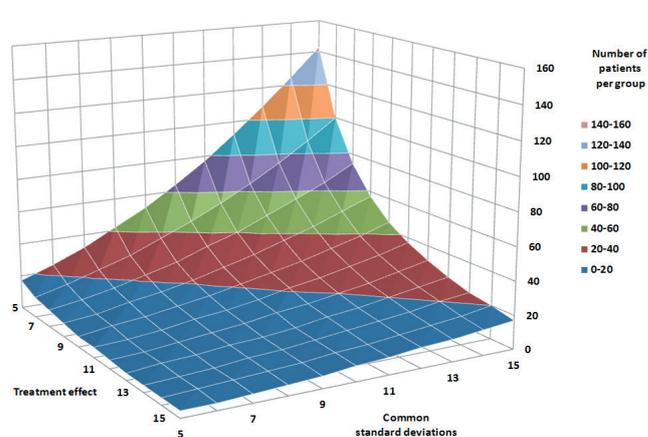
**Fig. 5.** Sample size estimates for a test-of-difference (superiority) design of ankle brachial Index (ABI) with the value of 1.00 in the treatment group, different standard deviations (0.1-0.2) identical for both groups, and expected ABI decreases (treatment effect) in the control group (0.05-0.15) (Power=80%, one-sided alpha=0.025)



**Fig. 6.** Sample size estimates for a test-of-difference (superiority) design for walking distance (WD) with 350 m in the treatment group, different standard deviations (150-250 m) identical for both groups, and expected WD decreases (treatment effect) in the control group (50-150 m) (Power=80%, one-sided alpha=0.025)



**Fig. 7.** Sample size estimates for a test-of-difference (superiority) design for transcutaneous oxygen tension measurements (TcPO<sub>2</sub>) with 20 mmHg m in the control group, different standard deviations (5-15 mmHg) identical for both groups, and expected TcPO<sub>2</sub> increases (treatment effect) in the treatment group (30-45 mmHg) (power=80%, one-sided alpha=0.025)



**Fig. 8.** Sample size estimates for a test of difference (superiority) design of SF-36 physical score with a score of 40 in the control group, different common standard deviations (5.0-15) and expected SF-36 physical score increases (treatment effect) in the treatment group (5-15) (Power=80%, one-sided alpha=0.025)

techniques, angiographic endpoints (Table 2) seem to be favored in studies in interventional radiology. Late lumen loss (LLL), i.e. the loss of lumen diameter at follow-up (typically at 6-8 months) relative to the post-intervention time point, necessitates a blinded core lab. In principle, quantitative angiography analyzes *shadow images* of contrast-filled arterial segments using edge detection algorithms. LLL is the difference between the minimal lumen diameters (MLD) which are measured in the post-procedural and the follow-up angiograms. Based on MLDs, other angiographic endpoints, such as percent stenosis, can be calculated. For an expected LLL of 0.4 mm in the treatment and 0.6 mm in the control group (Figure 4), a total of 194 subjects would be required assuming a LLL standard deviation of 0.7 mm for both groups (power=80%, one-sided alpha=0.025). Mean lumen diameters, i.e. the averaged lumen diameter along a stenotic segment, can also be calculated from angiographic images. Provided that a comparison with an un-diseased vessel segment is possible<sup>19</sup>, intra-patient controls may serve as an alternative when a control group is not feasible. This approach also permits the use of a paired t-test (intra-patient comparisons), which may add statistical power.

Table 3 summarizes potential surrogate endpoints, such as the ankle brachial index (ABI), which can be readily measured in a non-clinical

setting. ABI is a ratio of systolic blood pressure in the dorsalis pedis or posterior tibial arteries to the highest systolic blood pressure value measured on the left or right arm.<sup>20</sup> In this scenario, ABI of less than 0.9 is considered to be the threshold value<sup>21</sup> for further diagnostic explorations. Assuming ABI standard deviation for both groups of 0.125 and expected mean ABIs in the treatment group of 1.00 and 0.95 in the control group, each group must have 100 patients.

For the pain-free walking distance and the maximum walking distance, biometric estimates based on a superiority design (test for difference) are illustrated in Figure 5. Assuming a walking distance (WD) of 350 m in the treatment group and 275 m in the control group with a common standard deviation of 150 m, derived from the literature<sup>22</sup>, 64 patients per group or 128 patients in total would be required to reject the null hypothesis of no difference between both study groups (one sided  $\alpha=2.5\%$ , power=80%).

Below the knee (BTK) clinical studies are mostly concerned with limb salvage and potentially ulcer healing; thus, measurement of transcutaneous oxygen tension (TcPO<sub>2</sub>) may be a clinically relevant surrogate endpoint in such studies. Redlich et al.<sup>12</sup> documented pre-treatment TcPO<sub>2</sub> values of  $20.5 \pm 4.8$  mmHg in diabetic patients

with critical limb ischemia (CLI) and follow-up TcPO<sub>2</sub> values in the range of the 40–45 mmHg. The calculated sample sizes for TcPO<sub>2</sub>-based studies are illustrated in Figure 7. With an expected TcPO<sub>2</sub> of 45 mmHg and a standard deviation of 12.5 mmHg, 8 patients per group would have to be recruited.

The most commonly used questionnaire of quality of life is the EuroQol-5D with either 3 or five levels of responses to five questions regarding mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. In addition, this patient self-assessment tool includes an estimate, on a scale of 0 to 100, of how the patient sees their health status (100=best health status). This questionnaire needs validation on a country specific level, e.g. a German patient would probably be slightly more critical to specify his overall health status as compared to a patient from the South of France.

For simplicity, the sample size calculations for quality-of-life assessments are based on the SF-36 scoring system for which a pre-treatment baseline value of approximately 40 can be assumed for patients with CLI.<sup>10</sup>

Assuming post-treatment scores between 50 and 65 and identical standard deviations of 5.0–15.0 for both groups, the number of subjects can range between less than 10 and 100 per group.

A potentially useful semi-quantitative methodology (reimbursement scoring system) that illustrates the relative reimbursement value (RRV) for various study endpoints and their corresponding population sizes was first introduced in the field of interventional cardiology.<sup>23</sup> This approach can also be proposed for PAOD-related studies (Figure 9). The number of patients for a superiority test hypothesis, e.g., for MLD and LLL, can be displayed in terms of the reimbursement value. The required study sample would be in the range of 25–200 patients (see Figure 4). However, the reimbursement value for angiographic endpoints in PAOD-related studies would be rather low since they are not used in health technology assessments.<sup>17</sup>

An analogous non-inferiority test hypothesis in terms of amputation rates would require 200–1600 patients (Figure 2). In principle, clinical efficacy corresponds to the reimbursement value. In this regard, a cause-effect relationship between the clinical parameter and quality of life (QOL) must be established. A relative reimbursement value (RRV) of 100% means that there is a direct impact of the measured endpoint on QOL. This can be illustrated with the size of a bubble (Figure 9). In our opinion, the reimbursement value for patients with PAOD is directly related to the patients' mobility, which can be measured with the walking distance to accomplish daily chores (RRV=80%).<sup>24</sup> A Rutherford category shift for a given study sample would reflect this aspect as well. However, it averages the net benefit and may not be observed in all patients (90%) as other walking limiting factors, e.g. coronary artery disease, joint replacements, may co-exist. TLR rates are equally important (RRV=90%) and have been used for Markov-state transition model inputs for coronary devices by the National Institute for Health and Clinical Excellence (NICE) in the UK.<sup>25</sup>

ABI measurements (RRV=75.0%) correlate well with clinical benefit. Nevertheless, other patho-physiological or methodological confounding factors<sup>26</sup> may hinder the detection of direct correlation between ABI and the true post-procedural benefit from a QOL

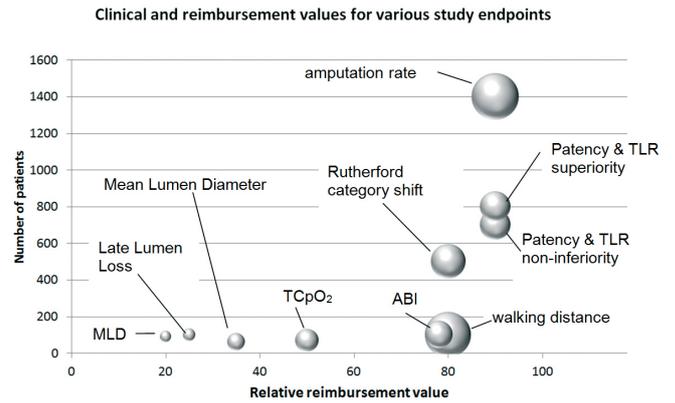


Fig. 9. Reimbursement value for various endpoints as a function of clinical value (size of the bubbles) and the number of patients for a general PAOD population.

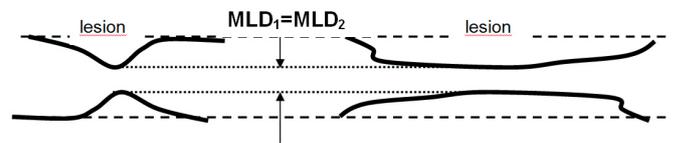


Fig. 10. Hemodynamic indifference despite identical Minimal Lumen Diameters (MLD) in two different lesions with a short (left side) and a long lesion (right side)

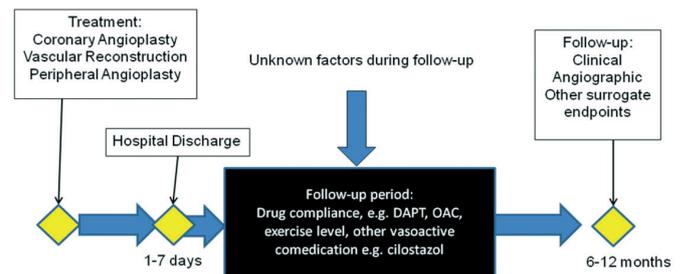


Fig. 11. In Confounding factors during treatment follow-ups

perspective. ABI measurements in target legs, when compared to non-target legs, appear useful to estimate treatment effects. Nevertheless, target and non-target legs must be functioning in the same patient to produce meaningful data for a QOL-relevant walking distance assessment. Major amputations have an important impact on the patients' QOL, especially in the case of CLI. Consequently, major amputation rates are highly relevant from a reimbursement point of view (RRV=90%) since they are associated with a high cost.<sup>27</sup>

An attempt to illustrate these complex relationships in patients with PAOD is presented in Figure 9. When focusing on the most striking difference, it is obvious that the endpoints of walking distance and ABI have a high reimbursement impact and a moderate demand for sample size. In contrast, MLD has a low reimbursement value despite a moderate size of the required sample. Studies with the primary endpoint of amputation rate have a high reimbursement value but require a large number of patients.

## DISCUSSION

To our knowledge, this is the first attempt to provide an overview of patient sample sizes for the most common endpoints used in PAOD-related studies. The input values to estimate the number of patients were obtained from various studies. At an early stage of

Tab. III. Other Surrogate Endpoints

SURROGATE ENDPOINT	ADVANTAGES	DISADVANTAGES
Ankle brachial index (ABI) of target leg	-can be based on paired t-test -can be standardized -hemodynamically intuitive -can be compared to ABI of non-target leg as a control group	-if not standardized, prognostic value may be reduced -may also be affected by atherosclerosis in the upper extremity vasculature or BTK media sclerotic changes -not useful in case of amputations
*Pain-free walking distance	-directly related to quality of life and to Rutherford/Fontaine classifications -smaller study populations	-walking ability may also be affected by musculo-skeletal events
*Maximum walking distance	-directly related to quality of life and to Rutherford/Fontaine classifications	-walking ability may also be affected by musculo-skeletal events
*Claudication onset time (COT)	-can be based on paired t-test -direct outcome for patients' quality of life	-treadmill standard setting are mandatory for all patients -walking ability may also be affected by musculo-skeletal events
*Absolute claudication time (ACT)	-can be based on paired t-test -direct outcome for patients' quality of life	-treadmill standard setting are mandatory for all patients -walking ability may also be affected by musculo-skeletal events
Transcutaneous oxygen tension measurements (TcPO <sub>2</sub> concentration)	-directly related to limb microperfusion -highly relevant to wound healing in patients with CLI -30 mmHg in TcPO <sub>2</sub>	-not well accepted -may be affected by venous insufficiency in the target limb
Regional perfusion index RPI=tcpO <sub>2</sub> (leg)/tcpO <sub>2</sub> (chest)	-standardized tpO <sub>2</sub> -directly related to limb microperfusion -highly relevant to wound healing in CLI patients	-not well accepted -may be affected by venous insufficiency in the target limb
Wound area measurement according to Bahmer [15]	-directly related to wound healing -high impact on quality of life -suitable for cost efficacy studies	-only applicable in CLI patients Rutherford 6

Comment: \*should be based on standard treadmill settings, e.g. 3.2 km/h, 12% inclination or 2% every 2 min

Tab. IV. Quality of Life Endpoints

Surrogate endpoint	Advantages	Disadvantages
EuroQol-5D in the 3L or 5L version	-patient self-assessment -simple overall quality of life status -relevant for reimbursement -recommended by the DEFINE group [4]4	- may be affected by other factors unrelated to PAOD treatment -needs to be validated on a country basis
RAND 36-Item Health Survey	-patient self-assessment -has been used in PAOD populations -relevant for reimbursement	-questionnaire scoring is complex
SF-36 physical score (short form)	-relevant for reimbursement -normally distributed	-more time consuming than EuroQol-5D -only validated for the US population

designing a study protocol, one refers to prior studies with similar target populations and treatments. Subsequently, an affordable sample size is determined, which is pivotal to the trial's cost-efficacy in company sponsored trials. Just because an endpoint is easily measured and requires a reasonable sample size, it does not automatically reflect patient benefits or improvement in QOL. In the remainder of this discussion, we would like to focus on some endpoints which do not appear to be appropriate from a reimbursement standpoint

### Angiographic endpoints

Despite being a direct outcome of quantitative angiography, we would like to advise to exercise caution when using MLD as a primary endpoint. In our opinion, MLD seems to be hemodynamically indifferent (Figure 10) since it does not consider the length of stenosis, as pointed out by studies in the field of interventional cardiology.<sup>23</sup> This, may have a significant impact on the pressure drop across the stenosis, affecting the magnitude of ischemia distal to the stenotic lesion, as shown in coronary lesions by Brosh and coworkers.<sup>28</sup> Despite the fact that the value of MLD is recognized and has been verified by in-vivo animal studies<sup>29</sup>, MLD does not

seem to be an appropriate endpoint. Finally, the theoretical basis provided by the Navier-Stokes equation<sup>30</sup>, which establishes also a length effect on narrowing fluid boundaries in the most general form, should not be ignored.

### ABI measurements

Al-Qaisi et al.<sup>26</sup> gave a comprehensive overview of ABI measurements and their practical value in diagnosing PAOD in patients with various co-morbidities. A plethora of studies have been published on this topic but two potential sources of misleading readings may compromise the prognostic value of ABI measurements. First of all, collateral vasculatures bypassing peripheral lesions may still be able to propagate pressure pulses to the ankle, thereby leading to normal ABI values. In such cases, the authors conclude that ABI cannot reveal the culprit lesion. Furthermore, Al-Qaisi and coworkers point out that strongly calcified peripheral arteries could lead to higher than expected pressure wave amplitudes due to the media sclerosis. This, in turn, would also lead to misleading ABI measurements. Patient with strongly calcified arteries include primarily diabetics, patients with renal insufficiency,

and elderly patients who represent a significant proportion of patients with PAOD.

### Walking distance

From a quality of life standpoint, walking distance should be a preferred endpoint. However, it must be pointed out that due to factors unrelated to the affected limb, such as coronary artery or pulmonary disease, the proportion of patients lost to follow-up may be quite high, which precludes repeated measurements. Matsumura et al.<sup>31</sup> investigated the efficacy of stent implantations in the superficial femoral artery. In that study, the investigators reported that only 10% of all patients were available to determine their walking distance at a one-year follow-up.

### Follow-up interval

Confounding factors during the follow-up period may hinder the detection of the real treatment effect in a properly conducted trial. A well-designed and sufficiently powered study might fail when the primary endpoint is not reached at follow-up. Obviously, throughout the period between hospital discharge and follow-up, confounding factors may preclude the determination of the cause-effect relationship between the initial treatment and the primary outcome (Figure 11). Factors like physical exercise or non-compliance to co-medication are also known to affect the outcome. In a randomized trial, one must assume that these biases are equally distributed across treatment groups. There is a lot of studies that elucidate interactive effects, such as the influence of social support on co-medication compliance.<sup>32</sup> These may also be applicable in the follow-up period. We think that this potential bias may be quite challenging in smaller patient groups of less than 100 subjects.

## REFERENCES

- Selvin E, Erlinger TP. Prevalence of and risk factors for peripheral arterial disease in the United States: results from the National Health and Nutrition Examination Survey, 1999-2000. *Circulation* 2004; **110**(6):738-43.
- Diehm C, Allenberg JR, Pittrow D, Mahn M, Tepohl G, Haberbil RL, Darius H, Burghaus I, Trampisch HJ. German Epidemiological Trial on Ankle Brachial Index Study Group. Mortality and vascular morbidity in older adults with asymptomatic versus symptomatic peripheral artery disease. *Circulation* 2009; **120**(21):2053-61.
- Norgren L, Hiatt WR, Dormandy JA, Nehler MR, Harris KA, Fowkes FG; TASC II Working Group. Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). *J Vasc Surg* 2007; **45** Suppl S:S5-67.
- Diehm N, Pattynama PM, Jaff MR, Cremonesi A, Becker GJ, Hopkins LN, Mahler F, Talen A, Cardella JE, Ramee S, van Sambeek M, Vermassen F, Biaino G. Clinical endpoints in peripheral endovascular revascularization trials: a case for standardized definitions. *Eur J Vasc Endovasc Surg* 2008; **36**(4):409-19.
- Diehm N, Diehm C, Kalka C, Dick F. Peripheral endovascular revascularization trials: high noon for uniform reporting standards. *Acta Chir Belg* 2008; **108**(4):382-5.
- Sullivan EJ. Clinical trial endpoints. U.S. Food and Drug Administration. [www.fda.gov/downloads/Training/ClinicalInvestigatorTrainingCourse/UCM283378.pdf](http://www.fda.gov/downloads/Training/ClinicalInvestigatorTrainingCourse/UCM283378.pdf), accessed: February 18, 2015.
- Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012; **31**(25):2973-84.
- Jensen LP, Lepäntalo M, Fossdal JE, Røder OC, Jensen BS, Madsen MS, Grenager O, Fastang H, Myhre HO, Baekgaard N, Nielsen OM, Helgstrand U, Schroeder TV. Dacron or PTFE for above-knee femoropopliteal bypass: a multicenter randomised study. *Eur J Vasc Endovasc Surg* 2007; **34**(1):44-9.
- Dippel EJ, Makam P, Kovach R, George JC, Patlola R, Metzger DC, Mena-Hurtado C, Beasley R, Soukas P, Colon-Hernandez PJ, Stark MA, Walker C;

## Reimbursement scoring system

To our knowledge, the proposed scoring system is the first attempt to illustrate study population sizes and reimbursement values for a given set of relevant cardiovascular trial endpoints. Although this scoring system is semi-subjective, we still believe that it serves its purpose for early trial design and satisfies reimbursement needs. The fact that WD and ABI seem to be the most useful endpoints is in agreement with a prior NICE technology appraisal for vaso-active drug treatments in patients with PAOD.<sup>33</sup>

## LIMITATIONS

Even though we estimated the required patient numbers for specific study scenarios, this work does not replace a properly conducted sample size estimation for a given design. This reimbursement scoring system needs to be critically assessed and improved. Furthermore, WD and ABI, as the preferred endpoints, require proper inclusion and exclusion criteria to minimize loss to follow-up and other relevant factors that are not related to the treated limb.

## CONCLUSIONS

The primary endpoints of WD and ABI offer the best reimbursement value for a relatively small study sample given that other non-vascular walking limitations can be controlled. Angiographic endpoints such as MLD do not seem useful from a reimbursement standpoint despite their intuitiveness.

- EXCITE ISR Investigators. Randomized Controlled Study of Excimer Laser Atherectomy for Treatment of Femoropopliteal In-Stent Restenosis: Initial Results From the EXCITE ISR Trial (EXCimer Laser Randomized Controlled Study for Treatment of Femoropopliteal In-Stent Restenosis). *JACC Cardiovasc Interv*. 2015; **8**(1 Pt A):92-101.
- Meijer WT, Hoes AW, Rutgers D, Bots ML, Hofman A, Grobbee DE. Peripheral arterial disease in the elderly: The Rotterdam Study. *Arterioscler Thromb Vasc Biol* 1998; **18**(2):185-92.
- Belch JJ, Bell PR, Creissen D, Dormandy JA, Kester RC, McCollum RD, Mizushima Y, Ruckley CV, Scurr JH, Wolfe JH. Randomized, double-blind, placebo-controlled study evaluating the efficacy and safety of AS-013, a prostaglandin E1 prodrug, in patients with intermittent claudication. *Circulation* 1997; **95**(9):2298-302.
- Redlich U, Xiong YY, Pech M, Tautenhahn J, Halloul Z, Lobmann R, Adolf D, Ricke J, Dudeck O. Superiority of transcutaneous oxygen tension measurements in predicting limb salvage after below-the-knee angioplasty: a prospective trial in diabetic patients with critical limb ischemia. *Cardiovasc Intervent Radiol* 2011; **34**(2):271-9.
- Cao P, Eckstein HH, De Rango P, Setacci C, Ricco JB, de Donato G, Becker F, Robert-Ebadi H, Diehm N, Schmidli J, Teraa M, Moll FL, Dick F, Davies AH, Lepäntalo M, Apelqvist J. Chapter II: Diagnostic methods. *Eur J Vasc Endovasc Surg* 2011; **42** Suppl 2:S13-32.
- Bradbury AW, Adam DJ, Bell J, Forbes JE, Fowkes FG, Gillespie I, Ruckley CV, Raab GM; BASIL Trial Participants. Bypass versus Angioplasty in Severe Ischemia of the Leg (BASIL) trial: A survival prediction model to facilitate clinical decision making. *J Vasc Surg* 2010; **51**(5 Suppl):525-685
- Bahmer FA (1999) Wound measurement made truly simple by point counting. *Arch Dermatol* 135(8):991-2
- Bosch JL, van der Graaf Y, Hunink MG. Health-related quality of life after angioplasty and stent placement in patients with iliac artery occlusive disease:

- results of a randomized controlled clinical trial. The Dutch Iliac Stent Trial Study Group. *Circulation* 1999;**99**(24):3155-60.
17. Simpson EL, Kearns B, Stevenson MD, Cantrell AJ, Littlewood C, Michaels JA. Enhancements to angioplasty for peripheral arterial occlusive disease: systematic review, cost-effectiveness assessment and expected value of information analysis. *Health Technol Assess* 2014;**18**(10):1-252.
  18. Scheinert D, Duda S, Zeller T, Krankenberg H, Ricke J, Bosiers M, Tepe G, Naisbitt S, Rosenfield K. The LEVANT I (Lutonix paclitaxel-coated balloon for the prevention of femoropopliteal restenosis) trial for femoropopliteal revascularization: first-in-human randomized trial of low-dose drug-coated balloon versus uncoated balloon angioplasty. *JACC Cardiovasc Interv* 2014;**7**(1):10-9.
  19. Kleber FX, Schulz A, Waliszewski M, Hauschild T, Böhm M, Dietz U, Cremers B, Scheller B, Clever YP. Local paclitaxel induces late lumen enlargement in coronary arteries after balloon angioplasty. *Clin Res Cardiol* 2015;**104**(3):217-25
  20. Hirsch AT, Haskal ZJ, Hertzner NR, et al. Lower ankle/brachial index, as calculated by averaging the dorsalis pedis and posterior tibial arterial pressures, and association with leg functioning in peripheral arterial disease. *Circulation* 2006;**113**(11):e463–e654.
  21. McDermott MM, Liu K, Greenland P, Guralnik JM, Criqui MH, Chan C, Pearce WH, Schneider JR, Ferrucci L, Celic L, Taylor LM, Vonesh E, Martin GJ, Clark E. Functional decline in peripheral arterial disease: associations with the ankle brachial index and leg symptoms. *JAMA* 2004;**292**(4):453-61.
  22. Micari A, Cioppa A, Vadalà G, Castriota F, Liso A, Marchese A, Grattoni C, Pantaleo P, Cremonesi A, Rubino P, Biamino G. 2-year results of paclitaxel-eluting balloons for femoropopliteal artery disease: evidence from a multi-center registry. *JACC Cardiovasc Interv* 2013;**6**(3):282-9.
  23. Waliszewski M, Rittger H. Surrogate and clinical endpoints in interventional cardiology: are statistics the brakes? *Ther Adv Cardiovasc Dis*. 2016;**10**(5):314-26.
  24. Carman TL, Fernandez BB Jr. Contemporary management of peripheral arterial disease: II. Improving walking distance and quality of life. *Cleve Clin J Med* 2006;**73** Suppl 4:S38-44.
  25. NICE guidance MTG1 2010. SeQuent Please balloon catheter for in-stent coronary restenosis. December 2010. <http://www.nice.org.uk/guidance/MTG1>
  26. Al-Qaisi M, Nott DM, King DH, Kaddoura S. Ankle Brachial Pressure Index (ABPI): An update for practitioners. *Vasc Health Risk Manag* 2009;**5**: 833–841.
  27. Peacock JM, Keo HH, Duval S, Baumgartner I, Oldenburg NC, Jaff MR, Henry TD, Yu X, Hirsch AT (2011) The incidence and health economic burden of ischemic amputation in Minnesota, 2005-2008. *Prev Chronic Dis* 8(6):A141.
  28. Brosh D, Higano ST, Lennon RJ, Holmes DR Jr, Lerman A. Effect of lesion length on fractional flow reserve in intermediate coronary lesions. *Am Heart J* 2005;**150**(2):338-43.
  29. Young DF, Cholvin NR, Roth AC. Pressure drop across artificially induced stenoses in the femoral arteries of dogs. *Circ Res* 1975;**36**(6):735-43.
  30. Sherman FS. *Viscous Flow*. 1990. McGraw-Hill.
  31. Matsumura JS1, Yamanouchi D, Goldstein JA, Pollock CW, Bosiers M, Schultz GA, Scheinert D, Rocha-Singh KJ. The United States Study for Evaluating Endovascular Treatments of Lesions in the Superficial Femoral Artery and Proximal Popliteal By using the Protégé EverFlex Nitinol Stent System II (DURABILITY II). *J Vasc Surg* 2013;**58**(1):73-83.
  32. Warner LM, Schüz B, Aiken L, Ziegelmann JP, Wurm S, Tesch-Römer C, Schwarzer R. Interactive effects of social support and social conflict on medication adherence in multimorbid older adults. *Soc Sci Med* 2013;**87**:23-30.
  33. NICE technology appraisal guidance TA223: Cilostazol, naftidrofuryl oxalate, pentoxifylline and inositol nicotinate for the treatment of intermittent claudication in people with peripheral arterial disease. May 2011. <http://www.nice.org.uk/guidance/ta223> -end of text-

Word count: 4100

Page count: 10

Tables: 4

Figures: 11

References: 33

DOI: 10.5604/01.3001.0009.9220

Table of content: <http://ppch.pl/resources/html/articlesList?issueId=9556>

Copyright: Copyright © 2017 Fundacja Polski Przegląd Chirurgiczny. Published by Index Copernicus Sp. z o. o. All rights reserved.

Competing interests: The authors declare that they have no competing interests.



The content of the journal „Polish Journal of Surgery” is circulated on the basis of the Open Access which means free and limitless access to scientific data.



This material is available under the Creative Commons - Attribution 4.0 GB. The full terms of this license are available on: <http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>

Corresponding author: Dr. Matthias W. Waliszewski; Senior Manager Medical Scientific Affairs, B.Braun Vascular Systems, Sieversufer 8, D-12359 Berlin, Germany; Tel: +49 30 568207-171, Fax: +49 5661 71 75-5671; Cell: +49 151 142 667 16; E-mail : [matthias.waliszewski@bbraun.com](mailto:matthias.waliszewski@bbraun.com)

Cite this article as: Waliszewski M., Redlich U., Breul V., Tautenhahn J.; Surrogate and clinical endpoints for studies in peripheral artery occlusive disease: Are statistics the brakes?; *Pol Przegl Chir* 2017; 89 (2): 39-48

