

REVIEW ARTICLE

The Range and Scientific Value of Randomized Trials

Part 24 of a Series on Evaluation of Scientific Publications

Stefan Lange, Stefan Sauerland, Jörg Lauterberg, Jürgen Windeler

SUMMARY

Background: The randomized, controlled trial (RCT) is the gold standard of scientific evidence for the attribution of clinical effects (benefits and harms) to medical interventions. Many different designs for RCTs have been developed in order to counter legitimate critical objections and to better adapt the trials to the continually changing challenges that face clinical research.

Methods: The diversity and adaptability of randomized trial designs are presented and discussed on the basis of a selective literature review and specific illustrative examples.

Results: A wide range of RCT designs enables adaptation to special research tasks and clinical framework conditions. These include (among others) crossover trials, n=1 trials, factorial RCT designs, and cluster-randomized trials. In addition, adaptive designs such as modern platform trials and pragmatic RCTs with simplified clinical questions and less severely restricted patient groups make broad recruitment of patients possible even in routine clinical practice.

Conclusion: Only the randomized allocation of subjects to the treatment and control groups, which is the defining property of RCTs, can adequately ensure that traits of the subjects which might disturb or bias a comparison of two or more medical interventions, will be evenly distributed across groups, regardless of whether these traits are known or unknown. The methodological variants and further elaborations of the RCT that are discussed here will help protect patients by enabling the assessment of the benefits and harms of medical methods and products on the basis of robust evidence even in the present era of rapid innovation.

► Cite this as:

Lange S, Sauerland S, Lauterberg J, Windeler J: The range and scientific value of randomized trials—part 24 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2017; 114: 635–40. DOI: 10.3238/arztebl.2017.0635

It is now consensual that randomized clinical trials (RCTs) are the gold standard for assessing relationships between intervention and outcomes. Many variants and special types of RCTs have been developed to improve the informative value for specific clinical situations and to carry out trials using this randomized design, even though this seems difficult from an organizational perspective. The following article describes a number of such practical possibilities. It is important to remember that randomization refers only to the random assignment to intervention groups; therefore, randomization is neither comparable to using placebos nor equivalent to blinding.

Designs

The classical and most frequent case of a randomized controlled trial (RCT) is a parallel comparison that occurs at the same time (parallel group comparison) of two or more interventions in which allocation to treatment groups is done randomly. While many possible methods can be used to achieve random allocation, electronic methods using random numbers are generally used nowadays.

A very essential element of RCTs is that, prior to inclusion of patients in the study, neither the person responsible for assigning nor the patients know which intervention group the patients are assigned to. This procedure is called allocation concealment and can best be ensured by randomization over telephone or Internet, as is done in modern studies (1). This way provides the guarantee that patients are not selectively included or excluded from the study based on knowledge about their future assigned group.

In studies with a low case number, it is still possible that imbalances between the groups occur for certain patient characteristics despite randomization (2). Theoretically, this is not a problem, as these imbalances are balanced out after a large number of study repetitions. However, if significant prognostic factors should be distributed equally between groups in a specific study, stratified randomization can be used. The presence of multiple factors can be taken into account by using minimization (*Box, example 1*) (3). For this, statistical allocation algorithms are used to ensure that the important prognostic properties are distributed as evenly as possible between the treatment groups at

every time point of patient inclusion. A random component can also be integrated into these algorithms.

Crossover studies are used to determine the short-term effectiveness of interventions (especially of drugs) in treating chronic diseases (*Box, example 2*). Each study participant receives both medications “A” and “B” in a randomized order (i.e., either AB or BA). The two treatment periods are usually separated by a wash-out period, to avoid overlap of the medication effects or side-effects. Crossover studies offer the advantage of making intra-individual comparisons possible. For instance, patients can be asked during which treatment period they felt better. This can lead to a considerable reduction of case numbers under certain circumstances. However, the significance of such studies depends on certain critical requirements. The most important criterion is obvious: at the beginning of the second intervention period, a patient must be able to reach approximately the same baseline state as prior to the first period. Therefore, a major area for such studies has long been asthma in its various forms. In contrast, crossover studies are not suitable for chronic progressive diseases or treatments that are aimed at healing or prolonging survival.

The so-called N = 1 trials (or N-of-1 trials) can be seen as a special type of crossover study (*Box, example 3*). In this case, the same patient is assigned (if possible, in a blinded manner) to several treatments and treatment periods with a random sequence. Comparing the treatments should then provide insight to the best treatment. For patients with a chronic ailment, different interventions can be examined individually. As only one patient is examined, the results can rarely be generalized, but they can nevertheless help to find the optimal treatment for individual patients, for example, in everyday medical practice. In principle, several N = 1 trials can also be meta-analytically combined in order to make generalizations when applicable (4).

Factorial designs “combine” two RCTs in one. They can be used to investigate two interventions (A and B) in parallel; these interventions can also be combined (A + B) (*Box, example 4*). In the simplest case of such a design, with 2×2 factors, patients are randomized to one of four groups (of A + B, only A, only B, or neither A nor B). Comparisons can then be made at the end of the study between patients treated with A and those not treated with A, as well as between those treated with B and those not treated with B. Additionally, the effects of the combination can be evaluated. One important advantage of factorial design is a considerable reduction of case numbers, since the same patients can be used for several questions (partial trial). However, an interpretation problem can arise for the two simple comparisons when the two treatments interact mutually in a relevant manner (that is, by weakening or strengthening the combined effect).

Cluster randomized trials are appropriate when organizational changes or educative measures are to be analyzed, or if for some reason it is difficult or impossible to carry out a comparison intervention at the same

center in parallel and to randomize individual participants (*Box, example 5*). Subject of such studies are, for example, hygiene and preventive measures that are randomized for all hospital departments, nursing homes, or school classes. Cluster randomized trials are also often used in primary care medicine, with certain interventions randomized to individual practices (5). Although outcome measures (for example, avoidance of infections) are determined at the patient level, the cluster nature of data—that is, the dependency on the patients (as the observation units) within a cluster—must still be taken into account for statistical analysis. Also, in a cluster randomized trial, awareness of treatment, and therefore also allocation concealment, can be problematic at the patient level (5, 6).

Adaptive designs allow the study design to be adjusted during the course of the trial (*Box, example 6*). This is primarily used for the adjustment of case numbers of studies, which can be increased or decreased based on interim evaluations. This is especially important if, at the start of the study, the possible effects of the treatment, or certain assumptions that are critical for determining case number (for example, the expected variability), can only be estimated with a high degree of uncertainty. In these cases, the planned size of an RCT might prove to be much too large or too small. An adaptive design makes it possible to carry out an interim analysis of the trial and to adapt the planned case numbers accordingly. Adaptive methods for RCTs can also be used in other ways, for example with regard to the outcome measures or to the patients to be included; however, such adaptations always requires close cooperation with competent biostatisticians (7). It is absolutely necessary that the adaptive designs to be used are first described in detail in the trial protocol. This means that any unplanned interim analyses—provided they are not indicated for safety reasons—should be avoided, as they could place the significance of the trial at risk. Even so, planned interim analyses, which are also intended to serve as an early stopping point if necessary, are not unproblematic, since effects at this point cannot be determined with the desired precision. In addition, premature stopping may lead to a distorted evaluation of effects due to large differences observed during an interim analysis (8, 9). In order to be used efficiently, interim evaluations within the framework of adaptive designs must be based on relatively short-term endpoints. Surrogates are often used, such as that of progression-free survival (PFS) in oncology.

Platform trials are a further development of adaptive designs (*Box, example 7*). In platform trials, several experimental interventions are evaluated against a shared control intervention and/or against each other, using a master protocol (10). However, in contrast to factorial design trials, whether or not combinations have a synergistic effect or mutually weaken each other is not evaluated. In pre-planned interim analyses, the allocation probability is adapted to the individual treatment arms, with the removal of individual arms or the addition of new ones (for example, combinations of

BOX

Examples of randomized controlled trials

1. Minimization

In a study of 300 patients with pulmonary metastases in colorectal cancer, the surgical removal of metastases is compared with active monitoring. To deal with the eight prognostic factors (such as age, sex, number of metastases, and T- and N-stages) and center stratification, minimization was chosen as the allocation mechanism (27, 28). In this case, a simple stratified randomization would not have been possible due to the multiple possible combinations.

2. Crossover studies

In a pilot study, 17 patients with high-risk long-term opioid treatment due to non-cancer-related pain were randomly divided into two groups, who received either a 4-week-period of oral hydrocortisone followed by a 14-day washout and then a 4-week-period of placebo, or the same in reverse treatment order (i.e., placebo, washout, and then hydrocortisone). The aim was to investigate the effects of hydrocortisone treatment on health-related quality-of-life, pain, and perception of pain. In the comparison, some subdomains (for example, effects of pain on everyday activities) showed benefits of hydrocortisone (29).

3. N = 1 trials

For patients with advanced cancer, the effects of methylphenidate treatment on fatigue symptoms was investigated using a series of individual N = 1 studies in a total of 43 patients. Patients could undergo up to three cycles of alternation between 3-day treatment with methylphenidate or with placebo in individually randomized order. No positive effects were demonstrated on an aggregated basis; however, eight patients were identified for whom an individual effect in favor of methylphenidate could be demonstrated using appropriate statistical methods (30).

4. Factorial design

In the HOPE-3 trial, patients were included if they did not have cardiovascular disease but were at intermediate risk for it. Patients were randomly assigned to receive either a cholesterol-lowering treatment, a blood-pressure lowering treatment, a combination of both, or a double placebo. In the study design, it was assumed that the individual treatments reduce the risk by 25–30%, and the combination treatment, by 45–50%. Thus, no relevant interactions between the treatments were expected. Only the cholesterol-lowering treatment resulted in risk reduction (31).

5. Cluster randomized study

The effects of medical training on the blood pressure control in patients in primary care was the main question addressed by a recent study (32). Specifically, long-term blood pressure measurements were evaluated for 103 patients from 22 practices, which were randomly divided into the intervention or control arm. After 5 months, both systolic (−8.2 mm Hg) and diastolic (−4.1 mm Hg) blood pressure measurements were decreased for the entire study population, but no significant differences were observed between the intervention arm and the control arm.

6. Adaptive design

The PHOENIX trial planned to treat 10 900 patients who were undergoing percutaneous coronary intervention (PCI) with cangrelor as compared to clopidogrel for platelet inhibition, with results compared to a composite endpoint. After inclusion of 70% of the patients, an interim analysis was planned. It was estimated that, with an event rate of 5.1% in the control arm and a relative risk reduction of 24%, the study would have a power of 86% to detect reduction. Minor changes in the rate of adverse events and/or the relative risk reduction would have had a considerable impact on the power, which is why there was a possibility to increase the number of cases. Three zones of possible results were defined: an unfavorable zone (relative risk reduction <13.6%), a promising zone ($\geq 13.6\%$ and $\leq 21.2\%$), and a favorable zone ($>21.2\%$); as the interim results fell within the favorable zone, the study was completed with the originally planned number of cases (33).

7. Platform trial

In 2005, the multicenter, open platform trial STAMPEDE began with patients who had advancing or metastatic prostate cancer. Initially, treatment with an androgenic deprivation therapy (ADT) either alone or in five combinations (e.g., ADT with zoledronic acid or docetaxel) was studied (34). Because of lack of efficacy, the COX2 inhibitor arms were discontinued (35), and additional treatment arms (e.g., ADT with abiraterone or metformin) were added (36, 37).

8. Pragmatic trial

In 75 primary care practices, 2799 COPD patients were randomly assigned to usual care or to receiving an additional drug treatment, which also contained an inhaled corticoid. The primary outcome of the unblinded study was the rate of exacerbations in the following year. Exclusion criteria were limited to a necessary minimum—for instance, acute symptoms or specific prior treatments—to avoid contraindications for the drugs. The additional treatment led to a significantly lower rate of exacerbations (38).

individual arms). Platform trials are an efficient alternative in indications with short innovation cycles and smaller target populations. They can also be designed as combined phase 2/phase 3 studies, in which case they are referred to as multi-arm multi-stage (MAMS) RCTs. Umbrella and basket trials can also be included here (11). Both terms are used for the evaluation of so-called targeted therapies in the context of personalized medicine in oncology. For instance, a histopathologic tumor entity (such as non-small cell lung cancer) can be analyzed in subgroups formed by therapies that are directed against the different driver mutations, and this analysis then compared to a common standard therapy (umbrella trial). Alternatively, if distinct types of histopathologic tumor entities are being analyzed, a common targeted therapy is examined for these tumor entities (basket design). Basket studies are, however, mostly carried out uncontrolled; a rationale for this is not really evident (12).

To counter the (sometimes justified) objection that RCTs form artificial scenarios, characterized for example by narrow criteria for inclusion and exclusion and many control examinations, pragmatic RCTs have gained a lot of interest in recent years (*Box, example 8*) (13, 14). The pragmatic element comes from the fact that the trials specifically and directly address the practical questions relevant for routine clinical practice, unfettered by possible side questions. Limiting the inclusion and exclusion criteria to only a few, easily detectable ones ensures that patient recruitment is wide-spread, even in everyday clinical practice. Further, focusing on a few endpoints that are patient-relevant and well-established enhances patient willingness to participate and increases the practical significance of the trials. Such trials can also be supported by registers (15, 16). The often extensive release of accompanying measures and therapies supports proximity to everyday clinical practice and acceptance. This goal-oriented and cost-effective approach is extremely useful for many questions of care and, as shown by multiple examples, is also easy to implement. However, this simplicity and ease-of-use does have its price. For instance, lack of adherence to strict guidelines produces statistical “noise”, which can significantly increase the necessary patient numbers (17). The low degree of standardization of procedures and data collection can also lead to implementation and interpretation problems. Further, forgoing the analysis of additional data prevents follow-up of additional questions; despite the fact that taking into account these additional data makes it more difficult to carry out trials, it is often exactly what makes clinical trials interesting in the first place for many medical scientists.

Effort and effectiveness of RCTs

The goal of obtaining a causal inference from a clinical trial with respect to the effectiveness of medical treatments is most efficiently achieved by RCTs, with the prerequisite that the same quality of standards is valid for all trial forms (following Good Clinical Practice

[GCP]). This is because the costs for preparation of the study protocol, quality assurance of observed medical interventions, and data collection and validation (including secure recording of adverse events) should not differ between different types of studies. Using randomization is by far the easiest and most reliable way to form structurally equivalent groups that permit a scientifically fair comparison between interventions. In contrast, non-RCTs require a much larger number of characteristics and data to be collected in order to try to statistically control bias by confounding influences in the analysis (for example, selection bias due to confounding by indication). Moreover, non-RCTs often yield significantly more heterogeneous results (18), which in turn means that larger sample sizes—and thus increased effort—are required. These are also reasons why not using randomization does not provide a solution for the comparison of rare diseases (19).

From a broader perspective, RCTs also lead to greater efficiency of research and supply of care; for instance, they are the only way to obtain the assurance of significance necessary for clinical guidelines. Thus, after decades, the randomized WHI trial could clarify whether hormone replacement therapy for postmenopausal women is beneficial (20). It is significant that, after evaluating non-RCT data (for example, from patient registries), researchers usually conclude that RCTs are necessary for the final clarification of a clinical benefit of interventions (21, 22).

RCTs in a meta-epidemiological approach—does it make sense?

Results from meta-epidemiological comparisons of RCTs and non-RCTs (mostly observational studies) that appear to suggest equivalence on the same clinical questions are sometimes presented as an argument against the assumed effort of carrying out RCTs. Even if it could be proven that both trial forms are empirically comparable and give similar results, it would still be wise to choose the much more efficient approach of an RCT. Why is that?

Comparisons of the relevant methodological reviews lead to very heterogeneous results. That is, some studies suggest that non-RCTs result in larger effect estimates, while others suggest that they result in smaller effect estimates. Combining these reviews into a meta-review (and overlooking the fact that this is actually inadmissible, due to heterogeneity) reveals no relevant differences, as shown by Anglemeyer and colleagues (23). Furthermore, when better quality and more sophisticated non-RCTs are evaluated, there is a decrease in the differences between RCTs and non-RCTs; in other words, these non-RCTs are closely approaching the RCTs in data quality and control of confounding factors (24). However, this degree of quality is rarely found in non-RCTs and is also very difficult to verify from publications, meaning that the results of conventional non-RCTs (which have a very high degree of bias potential as compared to standard RCTs) cannot be considered as valid.

Ultimately, meta-epidemiological empirical design comparisons do not provide clear answers—even if they show differences, these can be interpreted in various ways. For instance, differences could be due to confounding factors or to the otherwise poor quality of non-RCTs. Additionally, they could also be due to the different settings and study populations of the RCTs and non-RCTs, which would also introduce a systematic bias in comparison of study designs.

Conclusions

In order to arrive at robust, causally interpretable statements about the benefits and harms of (medical) interventions, studies with a nonrandomized allocation require an incomparably higher effort, since controlling for confounding variables is provided by randomization almost free of charge.

As we have shown, there are numerous ways to carry out RCTs in a targeted and valid manner. The necessary infrastructure is also available at universities with the coordinating centers for clinical trials. Developments such as platform and pragmatic trials impressively demonstrate that the RCT instrument is continually adapted to relevant questions, by introducing changes or using highly dynamic research framework conditions. RCTs are neither hostile to innovation (short innovation cycles are a popular counter-argument [25]) nor do they fundamentally contradict the desire for “real world evidence” (26). Therefore, RCTs should not only be maintained as a gold standard for clinical intervention studies and assessments of safety and efficacy, but should also become more important in Germany through targeted research funding to answer patient-relevant questions.

Conflict of interest statement

The authors declare that no conflict of interest exists.

Manuscript received on 6 April 2017, revised version accepted on 12 June 2017.

Translated from the original German by Veronica A. Raker, PhD.

REFERENCES

- Savovic J, Jones H, Altman D, et al.: Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012; 16: 1–82.
- Blair E: Gold is not always good enough: the shortcomings of randomization when evaluating interventions in small heterogeneous samples. *J Clin Epidemiol* 2004; 57: 1219–22.
- Altman DG, Bland JM: Treatment allocation by minimisation. *BMJ* 2005; 330: 843.
- Zucker DR, Ruthazer R, Schmid CH: Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; 63: 1312–23.
- Chenot JF: Cluster-randomisierte Studien: eine wichtige Methode in der allgemeinmedizinischen Forschung. *Z Evid Fortbild Qual Gesundheitswes* 2009; 103: 475–80.
- Kleist P: Studiendesigns mit unvollständiger Aufklärung der Versuchspersonen. *Schweiz Ärztezeitung* 2010; 91: 994–7.

KEY MESSAGES

- Randomized controlled trials remain the gold standard for determining causal efficacy of medical interventions and their benefit assessment.
- The main problem of the internal validity of nonrandomized studies is that confounding factors can bias the results; such confounding factors might not be controlled even by statistical adjustment due to unequal distribution of patient characteristics in the comparison arms.
- The variety of RCT designs is large and is constantly being further developed in order to meet the requirements of different research contexts as well as high innovation dynamics.
- Pragmatic RCTs counter the argument that trials are not close enough to everyday life. Within this framework, questions that are immediately relevant for everyday clinical practice can be addressed.
- The question whether RCTs and nonrandomized trials are equally suitable for clarifying clinical questions about the effectiveness and harms of interventions cannot be answered meta-epidemiologically by comparing the results of the different trials types. Such comparisons are associated with irresolvable problems of interpretation.

- Food and Drug Administration. Adaptive design clinical trials for drugs and biologics – Draft guidance [online]. 02.2010 www.fda.gov/downloads/drugs/guidances/ucm201790.pdf (last accessed on 18 February 2017).
- Bassler D, Montori VM, Briel M, et al.: Reflections on meta-analyses involving trials stopped early for benefit: is there a problem and if so, what is it? *Stat Methods Med Res* 2013; 22: 159–68.
- Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM: Problems of stopping trials early. *BMJ* 2012; 344: e3863.
- Berry SM, Connor JT, Lewis RJ: The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA* 2015; 313: 1619–20.
- Woodcock J, LaVange LM: Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 2017; 377: 62–70.
- Renfro LA, Sargent DJ: Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. *Ann Oncol* 2017; 28: 34–43.
- Roland M, Torgerson DJ: What are pragmatic trials? *BMJ* 1998; 316: 285.
- Tunis SR, Stryer DB, Clancy CM: Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290: 1624–32.
- Sacristan JA, Soto J, Galende I, Hylan TR: Randomized database studies: a new method to assess drugs' effectiveness? *J Clin Epidemiol* 1998; 51: 713–5.
- Lagerqvist B, Frobert O, Olivecrona GK, et al.: Outcomes 1 year after thrombus aspiration for myocardial infarction. *N Engl J Med* 2014; 371: 1111–20.
- Greenfield S, Kravitz R, Duan N, Kaplan SH: Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med* 2007; 120: 3–9.

18. Ioannidis JP, Haidich AB, Pappa M, et al.: Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001; 286: 821–30.
19. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bewertung und Auswertung von Studien bei seltenen Erkrankungen: Rapid Report; Auftrag MB13–01 (online). 05.09.2014 (IQWiG-Berichte; Band 241). [www.iqwig.de/download/MB13–01_Rapid-Report_Studien-bei-seltenen-Erkrankungen.pdf](http://www.iqwig.de/download/MB13-01_Rapid-Report_Studien-bei-seltenen-Erkrankungen.pdf) (last accessed on 18 February 2017).
20. Rossouw JE, Anderson GL, Prentice RL, et al.: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women’s health initiative randomized controlled trial. *JAMA* 2002; 288: 321–33.
21. Angus DC: Whether to intubate during cardiopulmonary resuscitation: conventional wisdom vs big data. *JAMA* 2017; 317: 477–8.
22. Sarno G, Lagerqvist B, Frobert O, et al.: Lower risk of stent thrombosis and restenosis with unrestricted use of ‘new-generation’ drug-eluting stents: a report from the nationwide Swedish Coronary Angiography and Angioplasty Registry (SCAAR). *Eur Heart J* 2012; 33: 606–13.
23. Anglemyer A, Horvath HT, Bero L: Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014; 4: MR000034.
24. Furlan AD, Tomlinson G, Jadad AA, Bombardier C: Methodological quality and homogeneity influenced agreement between randomized trials and nonrandomized studies of the same intervention for back pain. *J Clin Epidemiol* 2008; 61: 209–31.
25. Bundesverband Medizintechnologie (BVMed): 5-Punkte-Plan zur Nutzenbewertung von Medizintechnologien. Berlin: 2014. www.bvmed.de/de/versorgung/nutzenbewertung/5-punkte-nutzenbewertung (last accessed on 18 February 2017).
26. Sherman RE, Anderson SA, Dal Pan GJ, et al.: Real-world evidence—what is it and what can it tell us? *N Engl J Med* 2016; 375: 2293–7.
27. Treasure T, Fallowfield L, Lees B: Pulmonary metastasectomy in colorectal cancer: the PulMiCC trial. *J Thorac Oncol* 2010; 5: 203–6.
28. Surgical & Interventional Trials Unit (SITU) DoSIS, Faculty of Medical Sciences UCL. PulMiCC Newsletter Issue 001 (online). 03.2015. www.ucl.ac.uk/surgical-interventional-trials-unit/documents/trials_doc/pulmicc_doc/pulmicc_open/PULMiCC_news_docs/PulMiCC_Newsletter__Issue_001__March_2015_.pdf (last accessed on 18 February 2017).
29. Nenke MA, Haylock CL, Rankin W, et al.: Low-dose hydrocortisone replacement improves wellbeing and pain tolerance in chronic pain patients with opioid-induced hypocortisolemic responses. A pilot randomized, placebo-controlled trial. *Psychoneuroendocrinology* 2015; 56: 157–67.
30. Mitchell GK, Hardy JR, Nikles CJ, et al.: The effect of methylphenidate on fatigue in advanced cancer: an aggregated N-of-1 Trial. *J Pain Symptom Manage* 2015; 50: 289–96.
31. Yusuf S, Lonn E, Pais P, et al.: Blood-pressure and cholesterol lowering in persons without cardiovascular disease. *N Engl J Med* 2016; 374: 2032–43.
32. Weltermann B, Kersting C, Viehmann A: Hypertension management in primary care. *Dtsch Arztebl Int.* 2016; 113: 167–74.
33. Bhatt DL, Stone GW, Mahaffey KW, et al.: Effect of platelet inhibition with Cangrelor during PCI on ischemic events. *N Engl J Med* 2013; 368: 1303–13.
34. James ND, Sydes MR, Clarke NW, et al.: Systemic therapy for advancing or metastatic prostate cancer (STAMPEDE): a multi-arm, multistage randomized controlled trial. *BJU Int* 2009; 103: 464–9.
35. James ND, Sydes MR, Mason MD, et al.: Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. *Lancet Oncol* 2012; 13: 549–58.
36. Sydes MR, Parmar MK, Mason MD, et al.: Flexible trial design in practice—stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. *Trials* 2012; 13: 168.
37. STAMPEDE: Systemic therapy in advancing or metastatic prostate cancer: Evaluation of drug efficacy. A multi-arm multi-stage randomised controlled trial. Version: 15.0 (online). 29.03.2016 www.stampetrials.org/87548/87552/STAMPEDE_Protocol_v15.0_clean.pdf (last accessed on 18 February 2017).
38. Vestbo J, Leather D, Diar Bakerly N, et al.: Effectiveness of fluticasone furoate-vilanterol for COPD in clinical practice. *N Engl J Med* 2016; 375: 1253–60.

Corresponding author:

Dr. med. Dipl.-Psych. Jörg Lauterberg
 IQWiG – Institute for Quality and Efficiency in Health Care
 Im Mediapark 8,
 50670 Köln, Germany
 Joerg.Lauterberg@iqwig.de