Does it help that efficacy has been proven once we start discussing (added) benefit?

Armin Koch* and Yvonne Ziert

Institut für Biometrie, Medizinische Hochschule Hannover, Carl-Neuberg-Straße 1, 30625 Hannover, Germany

Received 20 January 2014; revised 20 March 2015; accepted 21 March 2015

Since the introduction of benefit assessment to support reimbursement decisions in Germany there seems to be the impression that totally distinct methodology and strategies for decision making would apply in the field of drug licensing and reimbursement. In this article, the position is held that, while decisions may differ due to differing mandates of drug licensing and reimbursement bodies, the underlying strategies are quite similar. For this purpose, we briefly summarize the legal basis for decision making in both fields from a methodological point of view, and review two recent decisions about reimbursement regarding grounds for approval. We comment on two examples, where decision making was based on the same pivotal studies in the licensing and reimbursement process. We conclude that strategies in the field of reimbursement are (from a methodological standpoint) until now more liberal than established rules in the field of drug licensing, but apply the same principles. Formal proof of efficacy preceding benefit assessment can thus be understood as a gatekeeper against principally wrong decision making about efficacy and risks of new drugs in full recognition that more is needed. We elaborate on the differences between formal proof of efficacy on the one hand and the assessment of benefit/risk or added benefit on the other hand, because it is important for statisticians to understand the difference between the two approaches.

Keywords: Assessment of clinical trials; Drug licensing; Reimbursement.

1 Introduction

The Act on the Reform of the Market for Medicinal Products (AMNOG) passed by the German government in 2011 brought a huge change to the German drug market. It provoked discussions between stakeholders regarding its impact on the drug market, the availability of new pharmaceutical products, and reimbursement (Der Deutsche Bundestag, 2010a). Already in 2004, the German Institute for Quality and Efficiency in Health Care (IQWiG) was established and has ever since undertaken evaluations of the added benefit of certain medicines already on the market. Since the introduction of AMNOG this act applies to all pharmaceutical products with a new active agent launched after January 1st, 2011 in Germany. It describes a multistage process, which includes settling the prices of such new drugs. The first step in this process is the assessment of added benefit, which is determined by the Federal Joint Committee (G-BA) with support from the IQWiG.

The discussions of the stakeholders especially focus on the additional requirements of this new process and the underlying methods for decision making. Differences to the process of drug licensing in Germany and the European Community are challenged, which is well-established and, for obvious reasons, has to precede the process of reimbursement (Sattelmeier et al., 2013). On a national level the Federal Institute for Drugs and Medical Devices (BfArM) and the Paul Ehrlich Institute (PEI) share

*Corresponding author: e-mail: Koch.Armin@mh-hannover.de
as national competent authorities the responsibility for the drug licensing process with the European Medicines Agency (EMA) on a European level.

In the recent past a number of authors have held the view that decision making in drug licensing and reimbursement differ (Hasford et al., 2010). This position has been refuted (Bender et al., 2010), but has been reiterated on various occasions.

In this paper, the mandates of “drug licensing” and “reimbursement” in Germany are described from a methodological point of view, first. We exclude those situations where differences between drug licensing and reimbursement decisions were based solely on formal reasons (e.g. chosen comparator or endpoint in drug licensing studies was not considered appropriate for reimbursement decision making), but eventually discuss the methodological implications. Instead, our focus is on those situations, where decision making was mainly based on the outcome of clinical trials. Therefore two examples from cardiovascular disease have been selected, where the same pivotal studies and the chosen endpoints were accepted in both assessments. For reasons explained below, examples from cardiovascular disease are optimally suited to compare the decision making strategies in drug licensing and reimbursement. We close with a discussion focusing on additional challenges in both fields and giving some recommendations about how the decision making processes could be better aligned. The quantification of the added benefit, which is as well a task of the reimbursement bodies in Germany, is not discussed here.

Throughout this paper, we use the term “reimbursement” whenever issues of added benefit are discussed irrespective of whether the legal framework of the AMNOG applies, or (added) benefit is discussed independently from this act. This is contrasted to the decision making during drug licensing, where, as will be explained in due course, formal proof of efficacy and a principally positive decision about the benefit/risk ratio has to be made. We feel that statisticians should be aware of the differences between the two processes, because of the increasing importance of statistics (and statisticians) in the discussion. According to our understanding, there is a huge chance for statisticians to appropriately involve in decision making beyond the question, whether some effect is “significant” or not. A combination of both, knowledge from biostatistics, and experience in clinical epidemiology and evidence-based medicine, is required if statisticians want to be involved beyond pure methodological questions.

2 Legal background and implications

2.1 Drug licensing

The legal basis for drug licensing in Germany is the Drug Law (AMG; Der Deutsche Bundestag, 2013). According to §25 (2) of the AMG the competent higher federal authority may refuse to grant the marketing authorization if (own translation):

1. The documents submitted […] are incomplete.
2. The medicinal product has not been tested in accordance with the current state of scientific knowledge […].
3. The medicinal product has not been manufactured according to established pharmaceutical rules or does not show appropriate quality.
4. The therapeutic efficacy attributed to the medicinal product by the applicant is lacking or is insufficiently substantiated by the applicant in accordance with the confirmed state of scientific knowledge.
5. The benefit/risk profile is unfavorable […].

This is for obvious reasons in line with the respective European legislation (Directive 2004/27/EC of the European Parliament and of the Council of 31 March 2004), where it is stated that “a marketing authorization shall be refused if, […] it is clear that: (a) the risk-benefit balance is not considered...
to be favorable; or (b) its therapeutic efficacy is insufficiently substantiated by the applicant; or (c) its qualitative and quantitative composition is not as declared.”

Albeit in instances single arm clinical trials may be sufficient to substantiate efficacy of a new drug or an already licensed drug in a certain new indication (new disease), controlled and preferably randomized clinical trials are the gold standard to evaluate the efficacy and safety of new drugs (Committee for Proprietary Medicinal Products, 1998, 2000). Phase III clinical trials in late stage drug development are supposed to provide the confirmatory evidence for efficacy and the evaluation of a positive benefit/risk ratio to substantiate the licensing decision.

It is important to note the different wording used in points four and five, suggesting qualitatively different requirements for the assessment of efficacy of a drug on the one hand, and its benefit/risk profile on the other. According to point four in the AMG list of criteria, or point (b) of the respective paragraph in the European legislation the efficacy of a new drug has to be substantiated. In statistical terms this would mean that within an adequate and well-controlled clinical trial a “statistically significant” difference regarding a primary outcome variable compared to placebo, or at least noninferiority (respectively equivalence) compared to an active comparator can be demonstrated (the three principal approaches to demonstrate efficacy of a new drug are discussed in ICH-E9 (Committee for Medicinal Products for Human Use, 1998)).

In contrast to the assessment of efficacy, the assessment of the benefit/risk profile needs no substantiation (see point 5, in the AMG list of criteria, or (a) of the European legislation, respectively), but includes evaluation and consideration. In practice the assessment of the benefit/risk profile is the result of a discussion, where different efficacy and safety endpoints are taken into account and even advantages that are not related to efficacy or safety (e.g. more convenient administration of the drug) can sometimes be put into perspective.

It is current practice that clinical trials are primarily planned for demonstrating significant effects regarding predefined primary efficacy endpoints and not to substantiate efficacy and safety of a drug and this is, as outlined above, in line with legal requirements (Koch, 2011). Because formal proof of efficacy of a new drug is the primary aim, rigorous statistical principles (e.g. addressing multiplicity issues, type-I and type-II error), as described in various international guidelines for clinical trials (e.g. ICH, EMA-guidelines), are of high relevance for the planning and assessment of a clinical trial.

In practice the assessment of a clinical trial regarding its efficacy and benefit/risk profile within the process of drug licensing is done in two hierarchically ordered steps. First, efficacy of a new drug is assessed and secondly, the benefit of drug treatment is balanced against risks and this includes a detailed assessment of the safety profile of the new drug. This means that the assessment of efficacy is the entry-card for further evaluations. Until now, the statistical principles and the disease/indication specific discussion about what is needed for planning a clinical trial in order to substantiate efficacy are the main focus of regulatory guidelines. Initially planned as a measure to harmonize requirements for drug licensing in the European member states, these guidelines are of relevance for the pharmaceutical industry, as well.

The second step, the evaluation of the benefit/risk profile, is characterized by interpreting safety signals very carefully and balancing these against the outcome of a broad range of variables describing positive effects of the drug under investigation. This process is specific for every drug and disease and concepts of statistical inference do not usually play a major role in the regulatory discussion. Obviously, formal planning of the safety assessment at the start of the phase III development is in most instances difficult because only phase III clinical trials are sufficiently informative about clinical safety. This is the justification for not planning the assessment of safety to the same degree as this is the case for the evaluation of efficacy; it will be the exception that, unless a certain risk can be identified (e.g. from other members of a class or early-phase trials), specific hypotheses can be reasonably formulated upfront.

Specific guidelines for benefit/risk assessment are still missing although there are attempts in the regulatory context, for example the European Medicines Agency’s Benefit-risk methodology project.
published a work package to report on applicability of current tools and processes for regulatory benefit/risk assessment in 2010.

2.2 Reimbursement

Since the introduction of AMNOG by the German government all pharmaceutical products with a new active agent launched to the German drug market after January 1st, 2011 have had to pass the so-called AMNOG-procedure. The legal basis for this procedure is §35b and §130b of the Social Security Code (SGB) V. The objective of the AMNOG procedure is to determine the reimbursement amount for a drug in due consideration of its added benefit. Basically, the AMNOG procedure consists of the following two phases: During the first phase, the additional benefit of a new drug is assessed. In case the G-BA arrives at a positive conclusion about additional benefit for a drug, the price, which the statutory health insurances will have to pay, is negotiated (discount negotiation). In case the G-BA does not conclude additional benefit for a drug, this drug is classified into a fixed-price group or, if this is not possible, the maximum price is the price of the comparative therapy according to the G-BA. It is also possible that the pharmaceutical company withdraws the new drug from the German market without any price negotiations. In the following, the first phase of the process is described in more detail.

When a new drug is launched to the German drug market, or four weeks after the licensure of a new indication the pharmaceutical company has to submit a Benefit Dossier to the G-BA ((1) §35b SGB V), whose format and methods have to follow the Rules of Procedure of the G-BA (Der Deutsche Bundestag, 2010b) and the IQWiG method paper (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013). In the dossier the applicant has to demonstrate that the new drug has additional benefit regarding patient-relevant endpoints, especially mortality, quality of life, and morbidity, compared to an appropriate comparator determined by the G-BA (“appropriate comparative therapy”). Usually, the G-BA contracts the IQWiG for giving a recommendation regarding the added benefit of the new drug within three months. After receiving the IQWiG’s recommendation the G-BA has to decide on the extent of the added benefit of the new drug within three months.

Although the G-BA does not have to follow the recommendations of the IQWiG, the institute plays a key role in the assessment of added benefit of new drugs. In order to understand the basic principles of how added benefit is determined by the IQWiG, it is necessary to have a closer look at its underlying methods and procedures. Beside the evaluation of formal aspects of the dossier, which are defined by the Rules of Procedure of the G-BA (Der Deutsche Bundestag, 2010b), the IQWiG assesses the reliability of the presented treatment effects regarding their extent and consistency. To accomplish this task the IQWiG follows international standards of evidence-based medicine, which are set in its method paper. The key component of the IQWiG’s evaluation is the independent assessment of the extent and the probability of the added benefit and whether this is in line with the conclusion of the pharmaceutical company. Thereby the extent and the probability of additional benefit of a new drug is determined by the following three steps (IQWiG, 2013):

1. The probability of the existence of a treatment effect is evaluated for every endpoint separately (qualitative statement). Based on the level of evidence the probability is rated as hint, indication, or proof.
2. For all endpoints for which at least a hint for a treatment effect has been certified, the extent of the effect is determined (quantitative statement). The following quantitative statements are possible: major additional benefit, important additional benefit, slight additional benefit, additional benefit not quantifiable, no additional benefit, and benefit smaller than the benefit of the comparator.
3. A conclusion regarding the additional benefit is drawn, which incorporates the level of evidence and extent of treatment effects of all endpoints.
The method paper of the IQWiG exactly defines what kinds of studies are needed to fulfill a certain evidence level. Additionally, it is exactly defined what size of the treatment effect for which type of patient-relevant endpoint (e.g. overall mortality, side effects, quality of life) corresponds to which kind of quantitative statement and this decision is based on the respective boundary of a 95%-confidence interval to also reflect the precision of the estimate. It is thus very transparent, how the information in the dossier is evaluated. Nevertheless, the final conclusion of the IQWiG regarding the overall added benefit, which compromises the different evidence levels of all endpoints, is a result of a weighing up process comparable to the evaluation of the benefit/risk profile within the process of drug licensing.

3 Where decision making for drug licensing and reimbursement should coincide

In the past “benefit” was being used synonymously to “efficacy”, the specific prespecified positive effect of a drug to cure or reduce the burden of disease or malfunction of the body (see e.g. Victor et al., 1991). According to the definition of the IQWiG, “benefit” is defined as positive effects of a drug that are causally related to patient-relevant endpoints (IQWiG, 2013). The same reference precisely defines a causal relationship as the availability of a sufficient degree of certainty that observed effects are caused by the experimental intervention alone (and are not merely a chance finding, or the consequence of confounding effects). Patient-relevant endpoints mainly include variables describing mortality, morbidity, or, health-related quality of life.

The strong emphasis on endpoints that are directly relevant to the patient (in contrast to surrogate endpoints that require validation and are in many instances not directly perceptible by the patient) is often seen and criticized as the main difference between decision making in drug licensing and reimbursement. In drug licensing various established surrogate endpoints are accepted, as long as there is validation and no negative evidence that the link between the surrogate and the clinical endpoint is no longer valid. Examples are blood pressure lowering, HbA1c lowering, and lipid lowering that are all accepted for reducing the risk of cardiovascular morbidity and mortality. Concerns with the cardiovascular safety of the anti-diabetic drug rosiglitazone and the Food and Drug Administration (FDA) requirement for cardiovascular outcome data when licensing new anti-diabetic drugs are an example for the regulatory reaction on such negative evidence (U.S. Department of Health and Human Services, Food and Drug Administration, 2008; Nissen and Wolski, 2007).

In order to compare decision making strategies in both fields we had to exclude situations with different requirements regarding endpoints. For this reason, we chose examples in the field of cardiovascular disease for further investigation. Large-scale clinical trials investigating treatment efficacy based on mortality, or relevant cardiovascular morbidity should provide good examples of situations, where regulatory guidance documents recommend primary and secondary endpoints for clinical trials that qualify for the assessment of reimbursement claims, as well. And in fact, among others, the trial of clopidogrel versus acetylsalicylic acid (aspirin) in patients at risk of ischaemic events (CAPRIE) and the study of platelet inhibition and patient outcomes (PLATO) have been used as pivotal studies for formal proof of efficacy and the assessment of a positive benefit/risk ratio (as is required in drug licensing) and for the assessment of added benefit (during the reimbursement discussion) (IQWiG, 2006; IQWiG, 2012).

Obviously, added benefit can be justified by various approaches, namely by establishing a causal relationship between treatment with the experimental drug and superior efficacy, superior safety, or both, as well as improvements in health-related quality of life. Again, in order to focus the discussion, we restrict ourselves further on the situation, where added benefit is justified by superior efficacy of the experimental drug as compared to control.
Table 1  CAPRIE (CAPRIE Steering Committee, 1996): Treatment effects overall and in the pre-stratified subgroups for the primary endpoint.

| Population | Clopidogrel group % [no./total no.] | Aspirin group % [no./total no.] | Risk difference % | 95%-confidence interval for risk difference | p-value\(^a\) |
|------------|----------------------------------------|-------------------------------|-----------------|----------------------------------------|---------------|
| Primary endpoint: death from vascular causes, myocardial infarction, ischemic stroke |
| Full\(^b\),\(^c\) | 9.8 [939/9599] | 10.7 [1021/9586] | -0.9 | [-1.8; -0.1] | 0.028 |
| (1) Previous stroke | 13.4 [433/3233] | 14.4 [461/3198] | -1.0 | [-2.7; 0.7] | 0.236 |
| (2) Previous myocardial infarction | 9.3 [291/3143] | 9.0 [283/3159] | 0.3 | [-1.1; 1.7] | 0.679 |
| (3) Previous peripheral arterial occlusive disease\(^d\) | 6.7 [215/3223] | 8.6 [277/3229] | -1.9 | [-3.2; -0.6] | 0.004 |

\(^a\)Two-sided p-values are derived from tests for superiority.  
\(^b\)Population CAPRIE.  
\(^c\)Population licensing.  
\(^d\)Population reimbursement.

4 Examples of reimbursement decisions in cardiovascular disease

In the following section, two reimbursement decisions in the field of cardiovascular disease are reviewed in detail and differences and similarities to the preceding licensing process are described. In both instances, subgroup findings played a key role in the reimbursement decision.

4.1 CAPRIE: Similar decision making strategies can lead to different outcome

CAPRIE (CAPRIE Steering Committee, 1996) is a randomized, double-blind, multi-center clinical trial in 19,185 patients with atherosclerotic vascular disease. Patients were randomized to aspirin or clopidogrel with the primary objective to demonstrate superior efficacy of clopidogrel over aspirin regarding a reduced number of events in a composite primary endpoint including ischemic stroke, myocardial infarction, and vascular death. With an event rate of 9.8% in the clopidogrel group compared to the event rate of 10.7% in the aspirin group for the primary endpoint (p-value = 0.028), the study met its primary objective. However, despite global superiority, results in prespecified strata of patients requiring treatment with platelet inhibitors, because of (1) previous myocardial infarction, (2) previous stroke, and (3) previous peripheral arterial occlusive disease (PAOD) were qualitatively and quantitatively different. They clearly indicated superior efficacy in the PAOD-group, only (see Table 1), suggesting that overall superiority was achieved mainly by the fact that a weighted average has been calculated (loosely speaking) from a strongly positive treatment effect in PAOD (risk difference –1.9% in favor of clopidogrel), a weak trend in patients with previous stroke (risk difference –1.0% in favor of clopidogrel) and a borderline neutral effect in patients with previous myocardial infarction (risk difference 0.3% in favor of aspirin).

In consequence, neither the discussion about licensing, nor the decision making concerning reimbursement was straightforward: while clopidogrel was licensed for the full study population in the end, only for the PAOD-group a positive decision on added benefit and reimbursement was made.
Particularly because results in the prestratified subgroup of patients with myocardial infarction favored the treatment with aspirin, the licensing decision had to take into account that sound efficacy data were available for treating these patients with aspirin. In fact, the positive licensing decision was based on concluding that noninferiority has been convincingly demonstrated and that (if a priori chosen) nobody had requested a noninferiority margin smaller than 1.7% (the actual upper boundary of the treatment effect in this subgroup). Unfortunately, all post-hoc discussions about the irrelevance of differences between treatments are convincing to a limited degree, only. In consequence, post-hoc it may be seen as the major obstacle to the interpretation of this trial that, despite the fact that noninferiority of clopidogrel as compared to aspirin would have been a plausible basis for licensing clopidogrel, no noninferiority margin had been prespecified for the global assessment, or the assessment of the prestratified disease entities to be combined in this trial. Admittedly, regulatory consultation had been the basis for the decision to combine the three subpopulations into one trial, because a similar mode of action and similar treatment targets were addressed in all three subpopulations by both drugs. One lesson to be learned from this important example is, that, instead of merely aiming at demonstrating superiority, clarifying the minimal requirements (i.e. which degree of inferiority of the comparator drug has to be excluded in each of the subindications) would in retrospect have been the best advice and should guide future planning of trials with a similar objective.

An elaborate discussion about similarities and differences between the licensing and the reimbursement decision making can be found in recent methodological literature (Bender et al., 2010; Hasford et al., 2010). To sum up, it can be said that in licensing and reimbursement decision making a close inspection of a clinical trial regarding consistency of the primary outcome, of secondary endpoints, and, likewise important, in relevant subgroups is a mandatory second step after a positive outcome is concluded for the overall clinical trial.

Regulatory consequences of inconsistent findings in subgroups or in relevant endpoints may be a refusal of the license in case the overall outcome is substantially put into question. A restriction of the indication may be considered in case, e.g., the treatment effect is nonconvincing in a relevant subgroup so that the overall benefit/risk ratio in this subgroup is considered negative (this is sometimes called the precautionary principle). This is expressed in various regulatory guidelines and will be the main topic of the forthcoming guideline on the role of subgroup analyses in phase III clinical trials (Committee for Medicinal Products for Human Use, 2010). Drugs can be licensed as soon as noninferiority in comparison to an established active comparator has been demonstrated. The application of these principles can be demonstrated by the licensing process of clopidogrel. Given that overall superiority had been demonstrated for the primary endpoint, in the regulatory discussion the totality of the information was considered sufficient to grant a license even in the subpopulation of patients after myocardial infarction (where the treatment effect numerically was in favor of the aspirin treatment): despite the fact that no formal margin had been prespecified, the estimated treatment effect and the associated confidence interval did not raise concern that comparable efficacy might not hold true. Thus, as an overall summary evaluation, clopidogrel has been licensed with proven noninferiority across all subindications.

Within the process of reimbursement added benefit needs to be established for the new drug, to justify higher costs on the market. Added benefit is, obviously, best supported by superior efficacy. In the CAPRIE trial this was considered to be given in the subgroup of patients with previous PAOD. Such decision making is clearly not supported for the two other subpopulations as evidenced from the estimates and confidence intervals for the treatment effect. A confirmatory conclusion about superior efficacy in PAOD patients would have required prespecification of a multiple testing strategy clarifying how to assess the subgroup after overall superiority has been established (Committee for Proprietary Medicinal Products, 2002).

Recently, Li et al. (2007) have shown that the likelihood is low for seeing significant interaction and a zero (or negative) differential treatment effect in a subgroup of a clinical trial that has shown a significant positive treatment effect overall under the assumption of a homogeneous treatment effect. Consequently, it can be assumed that the same applies to the equivalence (noninferiority)
situation: the likelihood should be low to see a significant interaction and a positive treatment effect in a subgroup of a clinical trial that has overall established (significant) noninferiority under the assumption of a homogeneous treatment effect. In line with this in the CAPRIE-trial, the significant test for heterogeneity ($p$-value = 0.042) is a clear additional indicator that the treatment effect is not homogeneous and supports, for the decision on reimbursement, the conclusion of superior efficacy in the subgroup of patients with PAOD.

Imagine further that the same outcome would have been observed in a study comparing clopidogrel to placebo (instead of aspirin). The overall $p$-value of 0.028 would allow the conclusion, that superiority has been established globally from a regulatory and a reimbursement point of view. Let us assume further that the same point estimates (and confidence intervals) would have been observed in the three subgroups. Then, the regulatory decision on drug licensing would likely be different, because the overall result of superior efficacy is substantially put into question by two prestratified subgroups demonstrating “no effect”.

This example demonstrates that firm grounds from the formal proof of efficacy and a principle decision about a positive benefit/risk ratio are needed to proceed to the assessment of added benefit if added benefit is determined without prespecification of necessary subgroup analyses (e.g. in subpopulations or different endpoints).

### 4.2 PLATO: Differing focus of the assessment during licensing and reimbursement

PLATO (James et al., 2009; Wallentin et al., 2009) is a randomized, double-blind, multi-center, and multi-regional phase III clinical trial. In this trial ticagrelor (the experimental drug) is compared to clopidogrel both on top of aspirin. A total of 18,624 patients were recruited with acute coronary syndrome (instable angina pectoris (IA), non-ST-elevation myocardial infarction (NSTEMI) or ST-elevation myocardial infarction (STEMI)), including patients planned for invasive management, that is coronary angiography with percutaneous coronary intervention (PCI) or coronary-artery bypass grafting (CABG), as well as patients intended for medical management. The primary objective of the trial was to demonstrate superior efficacy of the ticagrelor-based strategy over the clopidogrel-based strategy regarding a reduced number of events in a composite primary endpoint including vascular death, myocardial infarction, and stroke. With an event rate of 9.8% in the ticagrelor group compared to the event rate of 11.7% in the clopidogrel group for the primary endpoint ($p$-value < 0.001) the study met its primary objective (see Table 2).

Based on these efficacy results and a positive benefit/risk ratio, tigragrelor was successfully licensed for the whole patient population of the PLATO trial. In addition, the study achieved high public interest during the licensure process, because despite the fact that overall superiority of the ticagrelor-based strategy over the clopidogrel-based strategy could be demonstrated, the outcome for the primary endpoint in the United States was qualitatively different. In this region, patients in the clopidogrel + aspirin group (the control) graded substantially better than those treated with the experimental treatment ticagrelor + aspirin. Post-hoc analyses suggested that high dose aspirin in co-medication may be responsible for the poorer outcome in this region and has been contraindicated thereafter. Consequently, the regulatory authority in Europe recommends using ticagrelor with an aspirin maintenance dose of $\leq 150$ mg (results for the licensing population are not presented in this paper).

In consequence, the IQWiG and the G-BA both focus the discussion of the added benefit for ticagrelor on this subpopulation. Additionally, the G-BA determined that added benefit have to be assessed within four subindications (IA/INSTEMI, STEMI, STEMI/PCI, and STEMI/CABG) separately. The G-BA defined the appropriate comparator in certain subindications differently from what had been investigated in PLATO. While clopidogrel + aspirin were acceptable as a control for all subindications within the PLATO trial for the purpose of drug licensing, the G-BA determined that clopidogrel + aspirin is only the appropriate comparator for the subindications IA/INSTEMI and STEMI. For the patients with STEMI that will undergo PCI the combination of prasugrel +
Table 2  PLATO (Wallentin et al., 2009; AstraZeneca GmbH, 2011; IQWiG, 2011): Treatment effects in the full population and in subgroups for the primary and secondary endpoints.

| Population                          | Ticagrelor + aspirin | Clopidogrel + aspirin | Hazard ratio | 95%-confidence interval for hazard ratio | p-value<sup>a</sup> |
|-------------------------------------|-----------------------|-----------------------|--------------|------------------------------------------|---------------------|
| **Primary endpoint: death from vascular causes, myocardial infarction, or stroke** |                       |                       |              |                                          |                     |
| Full<sup>b</sup>                    | 9.8 [864/9333]        | 11.7 [1014/9291]      | 0.84         | [0.77; 0.92]                             | < 0.001             |
| **Secondary endpoint: death from any cause** |                       |                       |              |                                          |                     |
| Full<sup>b</sup>                    | 4.5 [399/9333]        | 5.9 [506/9291]        | 0.78         | [0.69; 0.89]                             | < 0.001             |
| Instable angina pectoris/non-ST-elevation myocardial infarction and aspirin ≤ 150 mg<sup>c</sup> | 3.8 [165/4725]        | 5.3 [226/4751]        | 0.73         | [0.60; 0.89]                             | 0.0022              |
| **Secondary endpoint: death from vascular causes** |                       |                       |              |                                          |                     |
| Full<sup>b</sup>                    | 4.0 [353/9333]        | 5.1 [442/9291]        | 0.79         | [0.69; 0.91]                             | 0.001               |
| Instable angina pectoris/non-ST-elevation myocardial infarction and aspirin ≤ 150 mg<sup>c</sup> | 3.1 [137/4725]        | 4.6 [197/4751]        | 0.70         | [0.56; 0.87]                             | 0.0012              |

<sup>a</sup>Two-sided p-values are derived from tests for superiority.
<sup>b</sup>Population PLATO.
<sup>c</sup>Population reimbursement.

aspirin was determined as the appropriate comparator. For patients with STEMI that will undergo CABG aspirin monotherapy was determined as the appropriate comparator. The IQWiG assessment respected this decision but according to a lack of data for the other subgroups, the added benefit could only be determined for patients with IA/INSTEMI and patients with STEMI that will undergo PCI, who are treated with an aspirin maintenance dose of ≤ 150 mg. The assessment of added benefit for patients with IA/INSTEMI was solely based on the clinical data of the PLATO trial, but this deemed to be acceptable, because of the high quality design and the huge number of patients in the subgroup of patients with IA or INSTEMI (4725 patients in the ticagrelor + aspirin group and 4751 patients in the clopidogrel + aspirin group, see also Table 2). An attempt was made to assess added benefit for patients with STEMI that will undergo PCI with an indirect comparison based on data from the TRITON study. However, in this indirect comparison “no added benefit” was ascertained by IQWIG, because no relevant differences could be detected between ticagrelor and the appropriate comparator (prasugrel).

For the assessment of added benefit, study results for patient-relevant outcomes (e.g. overall mortality) are the basis for decision making. For this reason, the focus of the determination of added benefit for ticagrelor was not on the primary efficacy endpoint, but on two secondary endpoints “death from any cause” and “death from vascular causes” within the subgroup IA/INSTEMI and aspirin ≤ 150 mg. For both endpoints superiority of the ticagrelor based strategy above the clopidogrel-based strategy could be demonstrated: the event rate for the endpoint “death from any cause” was 3.8% in the ticagrelor + aspirin group compared to 5.3% in the clopidogrel + aspirin group (p-value: 0.0022) and the event rate for the endpoint “death from vascular causes” was 3.1% in the ticagrelor + aspirin group compared to 4.6% in the clopidogrel + aspirin group (p-value = 0.0012). For the determination of the
overall added benefit, additional endpoints, for example myocardial infarction and stroke, number of unexpected events were evaluated, as well. Based on the joint assessment of different outcomes, a proof of a major additional benefit could be determined for the tigragrelor-based strategy. Consequently, a higher price for tigragrelor could be negotiated for the subgroup IA/INSTEMI, despite the fact that tigragrelor was licensed for all patients with acute coronary syndrome. 

Both examples clarify that discussions about added benefit are not merely secondary, exploratory investigations of clinical trial data, which are considered hypothesis-generating, or would be confirmed in additional clinical trials. Reimbursement decisions are used in a confirmatory way and the availability of a new drug in the market may depend on a positive decision about added benefit. They thus directly impact on the treatment of the patient population under investigation. Both, false-positive and false-negative decisions matter. Obviously, the price to be paid for having the option of a full confirmatory assessment within the process of reimbursement would be high when patient-relevant outcomes (or decision making in subgroups) are directly incorporated into the confirmatory assessment strategy: prespecification and appropriate control of multiplicity issues would be required and demand for larger sample sizes to be recruited for such trials.

Until now a more epidemiology-style decision making seems to be acceptable that benefits from the well-structured context of randomized clinical trials and the fact that formal proof of efficacy did precede the assessment of (added) benefit as a gate-keeper.

Nevertheless, decision making based on subgroup-findings from large randomized clinical trials is one of the most challenging tasks in clinical epidemiology. In the given example, the subgroup IA/INSTEMI with an aspirin maintenance dose of $\leq 150$ mg was not predefined and not respected in the randomization scheme. In consequence, there is no principle reassurance that the two treatment groups are balanced in the subgroup under investigation regarding important patients’ characteristics (e.g. demographic variables and risk factors prognostic for the outcome). Additional investigations are required. Some reassurance needs to be provided that observed differences in outcome are not caused by an imbalance of such factors. This problem can be ameliorated by conducting a detailed investigation of the baseline characteristics for the two treatment groups in the IA/INSTEMI subpopulation. Such baseline comparisons can be found only for three basic variables presented in the IQWiG dossier: results for gender, age and Caucasian race. According to this narrowed analysis of the baseline characteristics, the two groups were balanced, a thorough discussion is however required, which other parameters should be investigated as well.

5 Discussion

During drug licensing, a new drug has to pass the classical three hurdles: appropriate pharmaceutical quality, preclinical safety, and clinical trials that allow the evaluation of efficacy, safety, and a positive benefit/risk ratio (Paul and Trueman, 2001; Taylor et al., 2004). In Germany, market access and a basic nonnegotiable price is guaranteed with the licensing of a product, but the assessment of added benefit is relevant for negotiations of a higher price and thus additional requirements need to be fulfilled for full commercialization of new medicines.

In this paper, we pointed out methodological differences and similarities between the process of drug licensing and reimbursement in Germany regarding principle assessment strategies. We argue that both processes apply the same methods but that the strategies within the assessment of added benefit are more liberal from a methodological point of view compared to the formal proof of efficacy within the process of drug licensing. Furthermore we hypothesize that from a methodological point of view the assessment of added benefit compares well to the assessment of a positive benefit/risk ratio following formal proof of efficacy within the licensing process. In order to support these hypotheses, we discussed two examples from the field of cardiovascular disease as an area where it is obvious that requirements for trials should be in accordance regarding principal design, comparators, and, particularly, endpoints for decision making in both worlds.
5.1 Comments on the similarities and differences of drug licensing and reimbursement

By referring to these examples, similarities of the two processes become clear. A first similarity is that the assessment in both instances is usually based on data from clinical trials. It is current standard in both worlds that best evidence comes from randomized and well-controlled clinical trials. According to the Rules of Procedure of the G-BA (Der Deutsche Bundestag, 2010b), it is even mandatory that the data used for the licensing procedure has to be included into the assessment of added benefit. In addition, it has to be noted that according to the German Social Code Book, the assessment of added benefit has to be performed within the approved label. This may be criticized as a deviation from what has been the basis of a planned clinical trial, but could also be seen as an attempt to improve consistent decision making in drug licensing and reimbursement.

Unfortunately, attempts to be consistent do not include the choice of the comparator in a certain situation: in drug licensing the best current standard has to be chosen and is usually selected from those products that are licensed for a certain indication. This process includes some negotiations between the pharmaceutical company and regulatory bodies, in case different standards are preferred in different European countries. The scientific advice procedure of the Committee for Medicinal Products for Human Use (CHMP) is a measure to arrive at a mutually recognized decision between member states, regarding all aspects of study design for studies in the package to be submitted for licensing. This also includes the discussion of the comparator. In this process, the scientific basis regarding the assessment of efficacy and of the benefit/risk profile is jointly discussed and evaluated to come up with a binding decision that thereafter applies to the European region and would only be drawn into consideration, if at a later point in time new information was made available from other trials.

In contrast, the G-BA defines a therapy as an “appropriate comparator” according to its Rules of Procedure (Der Deutsche Bundestag, 2010b). These rules include additional criteria (e.g. the evidence base of the comparator in the indication under investigation is evaluated and comparators are preferred if studies for patient-relevant endpoints have already been conducted) and this may not always lead to the same recommendation regarding the choice of the comparator. Probably, further attempts are mandated to arrive at a mutual recognition of important aspects for the selection of one comparator that may suffice for both purposes, the principal assessment of efficacy and positive benefit/risk ratio on the one hand and the assessment of added benefit on the other.

5.2 Implications of hierarchical decision making

In principle separate studies could be considered necessary or be requested to provide the basis for evaluation in the context of drug licensing on the one hand and in the context of reimbursement on the other hand. Nevertheless, it is current practice (and an ethical demand) that wherever possible the same clinical data are used to substantiate both decision making processes.

It has been criticized that evaluations of clinical trial data during reimbursement would involve endpoints, or subgroups, for which the original trial has not been powered (Hasford et al., 2010; Sattelmeier et al., 2013; Witte and Greiner, 2013). Obviously, there would be a price to be paid if one wished to have fully powered trials for the assessment of added benefit. Either separate trials would need to be conducted postlicensing to provide the respective confirmatory evidence, or phase III trials need to be planned and powered to substantiate principal efficacy to assess a prespecified added benefit. The assessment of added benefit according to current practice and classical safety assessment share that (unless appropriately powered) interpretation needs to be more open and has to be less formalized than classic assessment of principal efficacy in order to avoid that some relevant effects are not detected.

The legal background determines the hierarchy of regulatory decision making. First, the formal proof of efficacy and a principal decision about a positive benefit/risk ratio is established during the licensure process. Positive outcome of this step is needed to go on with more detailed analyses (e.g. analyses in subgroups and of patient-relevant endpoints) to explore and finally decide whether there is
added benefit of a new drug. Examples of the application of hierarchical regulatory decision making are the licensing process and the reimbursement decision for clopidogrel and tigagrelor based on the data of the CAPRIE trial and the PLATO trial. In both examples efficacy and a positive benefit/risk ratio has been concluded for the full population (beyond the dosing issue of aspirin in the PLATO trial) within the licensing process. Thereafter, added benefit has been concluded based on subgroups and secondary endpoints without a prespecification of a formal multiple testing procedure within the reimbursement process. This strategy is in line with the assessment of a generally positive benefit during drug licensing, which also integrates other endpoints and subgroup assessments beyond the preplanned primary analyses.

It is the main misunderstanding in some of the discussions about assessment strategies that the rules for formal proof of efficacy should be compared to the methods of benefit assessment in the field of reimbursement. In fact, the assessment of a principally positive benefit/risk ratio is reflected in the summary of product characteristics (SmPC) and the summary of the European Public Assessment Report (EPAR) as outcome of the licensing process. Rules governing benefit assessment during licensure are, if at all, less clearly described in current regulatory guidance documents as compared to the IQWiG methods. Here, the discussion regarding benefit assessment in the IQWiG methods paper is well advanced beyond what is available in the regulatory context. Formal methodology for the assessment of the benefit/risk ratio is an ongoing project at the European Medicines Agency (Abadie et al., 2009).

In instances drugs were licensed despite the fact that no formal proof of efficacy had been established (e.g. in rare disease under the restriction of limited sample-sizes). This is, however, no contradiction to the general hierarchical regulatory strategy for decision making, namely that a formal proof of efficacy has to be established, before an assessment of benefit (or the evaluation of a positive benefit/risk ratio) can be reasonably interpreted. Decision making about added benefit follows thereafter in the same hierarchy: the assessment of added benefit is only interpretable once a formal proof of efficacy is available and positive benefit/risk ratio is conceded. Decision making in clinical trials for very rare diseases within the licensing procedure signifies the opposite situation: in instances no formal proof of efficacy is possible and benefit/risk assessment needs to integrate findings from different endpoints or different subgroups without some sort of shelter from a preceding formal proof of efficacy. According to the German Social Code Book the pharmaceutical company do not have to show the added benefit for any licensed orphan drug, so that the AMNOG procedure do not have to deal with this problem directly.

So in principle, a multiple testing procedure with control of a type-I error could be outlined, that mimics the described hierarchy (1. formal proof of efficacy, 2. assessment of benefit or the evaluation of a positive benefit/risk ratio). Unfortunately, the assessment of drug-risks needed for benefit/risk assessment can barely ever be preplanned before the phase III clinical trials. This is an argument against multiplicity adjustment for safety endpoints. As it is not possible to preplan, all adjustment would complicate the appropriate evaluation of safety signals. This also clarifies that the prespecification of variables to be considered in the benefit/risk assessment is hardly ever possible.

5.3 The role of surrogate endpoints and patient-relevant outcomes

Based on definition in the German Social Code book, the IQWiG specifies the type of variables in its method paper that should be considered for the discussion of what would constitute an added benefit to the patient. Usually, a large number of variables are and should be included into the assessment of added benefit. Nevertheless, clinical studies and their assessment could be planned for both, formal proof of efficacy and the assessment of added benefit as soon as reimbursement bodies would specify per indication, which variables should precisely be used in the discussion of added benefit. Just consider for a moment that all reimbursement strategies would be based on an assessment of overall mortality. As then, there was no choice regarding endpoints for the reimbursement decision, there would be no
problem with a control of the experimentwise type-I error as long as decision making is based on one study.

Obviously, the attempt to preplan would come at huge costs. These costs would increase with increasing numbers of variables to be included into a “confirmatory” strategy for the evaluation of benefit. Complaints about an additional hurdle and an increase in the burden of what has to be demonstrated before a drug can be successfully established in the market would then be justified. It could be seen as an attempt to compromise with the current approaches in drug development that on one the hand, the process is specific about concentrating mainly on patient-relevant outcomes, but it is less stringent regarding methodological expectations compared to the formal proof of efficacy, on the other hand.

The regulatory system is more used to take outcome of surrogate endpoints into consideration, but the discussion about the safety of anti-diabetic drugs has clarified that the surrogacy has to be reestablished from time to time and particularly if new therapeutic principles are investigated. Conclusions about added benefit from surrogate endpoints are even more difficult and again an example from diabetes treatment is illustrative:

Strategies to lower blood glucose levels further (which could be understood as investigating new drugs with better efficacy on the level of the surrogate endpoint) turned out to increase mortality. This surprising outcome of the ACCORD-trial (Gerstein et al., 2008) may have been caused by various reasons among which polypharmacy in the population under investigation has been discussed. However, this may also be seen as another argument that isolated investigations of a short-term outcome may not suffice to address the complexity of drug development (and treatment) for a multimorbid population.

Both examples (cardiovascular safety of anti-diabetic drugs as quoted above, and the outcome of the ACCORD-trial) justify the principle approach in reimbursement to decide by default based on clinical endpoints. In the regulatory system, requirements for drug licensing need to be justified by evidentiary needs regarding required knowledge about efficacy and safety of drugs in a certain disease area. In most instances in clinical practice decision making is successfully based on blood glucose levels. In this situation, regulatory standards can only change if new evidence becomes available that disproves, for example, the universal surrogacy of HbA1c lowering for clinical benefit. In these instances increases in the regulatory hurdle for certain classes of drugs can be justified and have to be incorporated into the assessment strategies for drug licensing based on surrogate endpoints if eventually for a new drug better outcome is observed based on a surrogate endpoint.

It is a basic principle in statistics that more/better information comes at a price, which is in clinical research to be paid by the number of patients to be included into clinical trials, or the duration of clinical trials. Currently, a compromise is being taken that benefit assessment during licensing and reimbursement does not have to be done with the same rigor than formal proof of efficacy. In many instances the same clinical trials have been used to support the decision making in both worlds. Prudent planning of phase III clinical trials should include endpoints of relevance for the reimbursement discussion and considerable work on what can be expected for these. In the end, it may be more efficient to invest more into the phase III clinical trials and preparing a proper dossier for both, licensing and reimbursement decision making strategies based on the same clinical trials. A compromise position might be to plan phase III clinical trials that utilize surrogate endpoints such that at least a trend for patient-relevant outcome can be demonstrated.

6 Conclusion

Further investments in planning of clinical trials may be the only way forward to improve the overall regulatory decision making process while maintaining development costs at a reasonable level. Formal proof of efficacy, as the first step in drug licensing and the gate-keeper for further discussions is building the solid scientific ground. The relevance of the primary endpoint(s) needs careful discussion
and likewise it is important that the relevance of eventually observed significant treatment effects is discussed and agreed upon upfront, not only in the process of drug licensing. Thereafter, a more epidemiology-style decision making about the benefit/risk ratio or added benefit is justified if the minimal amount of information (e.g. endpoints per indication) is agreed upon between the different stakeholders (e.g. regulatory bodies, pharmaceutical companies, patients), that needs to be provided.

**Acknowledgment**

The authors wish to express their appreciation of the comments received by the editors and reviewers that helped to clarify the messages in this paper. This work has been funded partly by the FP7-HEALTH-2013-INNOVATION-1 project Advances in Small Trials Design for Regulatory Innovation and Excellence (ASTERIX) Grant Agreement No. 603160.

**Conflict of interest**

The authors have declared no conflict of interest.

**References**

Abadie, E., Calvo Rojas, G., Demolis, P., DeGraeff, P., Flamion, B., Hudson, I. and Brassart, L. (2009). *Benefit-Risk Methodology Project*. London, UK, p. 28.

Bender, R., Koch, A., Skipka, G., Kaiser, T. and Lange, S. (2010). No inconsistent trial assessments by NICE and IQWiG: different assessment goals may lead to different assessment results regarding subgroup analyses. *Journal of Clinical Epidemiology* 63, 1305–1307.

CAPRIE Steering Committee. (1996). A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. *Lancet* 348, 1329–1339. Retrieved from: http://www.ncbi.nlm.nih.gov/pubmed/8918275.

Committee for Medicinal Products for Human Use. (1998). ICH Topic E9: statistical principles for clinical trials. *CPMP/ICH/363/96*. Retrieved July 12, 2013, from http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf.

Committee for Medicinal Products for Human Use. (2010). Concept paper on the need for a guideline on the use of subgroup analyses in randomised controlled trials. *EMA/CHMP/EWP/117211/2010*.

Committee for Proprietary Medicinal Products. (1998). ICH Topic E8: Note for guidance on general considerations for clinical trials. *CPMP/ICH/291/95*.

Committee for Proprietary Medicinal Products. (2000). ICH TOPIC E 10: Note for guidance on choice of control group in clinical trials. *CPMP/ICH/364/96*.

Committee for Proprietary Medicinal Products. (2002). Points to consider on multiplicity issues in clinical trials. *CPMP/EWP/908/99*.

Der Deutsche Bundestag. (2010a). Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz—AMNOG). *Bundesgesetzblatt* 2010, 2262–2277.

Der Deutsche Bundestag. (2010). Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung—AM-NutzenV). *Bundesgesetzblatt* 2010, 2324–2328.

Der Deutsche Bundestag. (2013). Sechzehntes Gesetz zur Änderung des Arzneimittelgesetzes. *Bundesgesetzblatt*, 2013, 3813–3819.

Gerstein, H. C., Miller, M. E., Byington, R. P., Goff, D. C., Bigger, J. T., Buse, J. B. and Friedewald, W. T. (2008). Effects of intensive glucose lowering in type 2 diabetes. *The New England Journal of Medicine* 358, 2545–2559.

Hasford, J., Bramlage, P., Koch, G., Lehmacher, W., Einhäupl, K. and Rothwell, P. M. (2010). Inconsistent trial assessments by the National Institute for Health and Clinical Excellence and IQWiG: standards for the performance and interpretation of subgroup analyses are needed. *Journal of Clinical Epidemiology* 63, 1298–1304.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2013). Allgemeine Methoden 4.1. James, S., Akerblom, A., Cannon, C. P., Emanuelsson, H., Husted, S., Katus, H. and Wallentin, L. (2009). Comparison of ticagrelor, the first reversible oral P2Y(12) receptor antagonist, with clopidogrel in patients with acute coronary syndromes: rationale, design, and baseline characteristics of the PLATElet inhibition and patient Outcomes (PLATO) trial. *American Heart Journal* 157, 599–605.
Koch, A. (2011). [Benefits and harms—two sides of the same medal?]. Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen 105, 163–170.

Li, Z., Chuang-Stein, C. and Hoseyni, C. (2007). The probability of observing negative subgroup results when the treatment effect is positive and homogeneous across all subgroups. Drug Information Journal 41, 47–56.

Nissen, S. E. and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. The New England Journal of Medicine 356, 2457–2471.

Paul, J. E. and Trueman, P. (2001). “Fourth hurdle reviews”, NICE, and database applications. Pharmacoepidemiology and Drug Safety 10, 429–438.

Sattelmeier, J., Prenzler, A. and Frank, M. (2013). Das Arzneimittelmarktneuordnungsgesetz (AMNOG) in der Praxis—erste Erfahrungen, Kritikpunkte und Weiterentwicklungsmöglichkeiten. Gesundheitsökonomie & Qualitätsmanagement 18, 213–220.

Taylor, R. S., Drummond, M. F., Salkeld, G. and Sullivan, S. D. (2004). Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. BMJ (Clinical Research Ed.) 329, 972–975.

U.S. Department of Health and Human Services, Food and Drug Administration, C. for D. E. and R. (CDER). (2008). Guidance for Industry Diabetes Mellitus—Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Guidance for Industry (p. 8). Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf

Victor, N., Schäfer, H. and Nowak, H. (1991). Arzneimittelforschung nach der Zulassung. Springer, Berlin, Heidelberg, DE.

Wallentin, L., Becker, R. C., Budaj, A., Cannon, C. P., Emanuelsson, H., Held, C., Thorsén, M. (2009). Ticagrelor versus clopidogrel in patients with acute coronary syndromes. The New England Journal of Medicine 361, 1045–1057.

Witte, J. and Greiner, W. (2013). Problembefunde der Quantifizierung des Zusatznutzens im Rahmen der frühen Arzneimittelnutzenbewertung. Gesundheitsökonomie & Qualitätsmanagement 18, 226–234.