New Drugs: Evidence Relating to Their Therapeutic Value After Introduction to the Market

Mariam Ujeyl, Claudia Schlegel, Siegbert Walter, Ursula Gundert-Remy

SUMMARY

Background: Drug approval is based on three criteria: quality, efficacy, and safety. We investigated the types of study design and statistical methods employed to demonstrate safety and efficacy of proprietary medicinal products (PMPs) that were approved for use in the European Union through the centralized procedure.

Methods: We retrospectively analyzed the European Public Assessment Reports of PMPs that the European Medicinal Agency approved, either initially or for extended indications, in 2009 and 2010.

Results: Data were analyzed for 39 PMPs: 64% of these were new active substances, and 36% were approved for extended indications. 46% of the PMPs had been studied in an active-control trial. In only 28%, superiority of the new PMPs compared to active control had been tested. 46% of the approvals included testing of a patient-relevant primary endpoint. The median size of population used to demonstrate safety was 1700 persons.

Conclusion: The centralized procedure does not require comparative information from active-control trials. Accordingly, as our descriptive analysis revealed, this information is often not available at the time of market introduction. Pivotal studies only rarely clearly demonstrate an added therapeutic value of a new PMP compared to existing alternatives.

Cite this as:
TABLE 1

Centralized EU approvals (2009 to 2010) and indications/treated illnesses of PMPs in trials

<table>
<thead>
<tr>
<th></th>
<th>Number, n</th>
<th>Approval of new active substance</th>
<th>Extension of indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Launched on the German market*1</td>
<td>76</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>Included in analyses</td>
<td>39</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>of which:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Coagulation</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Skin</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Lungs</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Cancer</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Psychiatry/neurology</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Other*2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>


*1 excluding generics, biosimilars, vaccines, orphan drugs, new dosage forms, PMPs with the same active substances, and combinations of active substances already known;
*2 gynecology, urology, metabolic disorders, pain sufficient data to assess the clinical value of a new drug at the time of market introduction. Suggestions include the requirement to submit information on comparative benefit that is based on superiority testing over an active control (and not solely on demonstration of non-inferiority or equivalence) (4, e2). However, approval may only be refused if a drug does not demonstrate the necessary pharmaceutical quality, there is insufficient or no evidence of its efficacy, or it does not show adequate safety, i.e. a positive risk/benefit ratio (5).

In addition it has been discussed whether demonstrating efficacy and safety always constitutes evidence of a benefit to patients (6). The efficacy of a drug is demonstrated statistically significant effects on outcomes, e.g. measured values or symptoms. However, improvement in an outcome such as a laboratory value (“surrogate parameter”) cannot always reliably be interpreted as evidence of improved health.

In addition to demonstrating efficacy, demonstrating benefit requires evidence that a drug improves an outcome that is relevant to patients (a “patient-relevant endpoint”) when compared to placebo or other treatment. This can be demonstrated by improved health, shorter disease duration or longer life, improved quality of life, or a more favorable side-effect profile. If a drug affects a surrogate variable that has been demonstrated to affect a patient-relevant endpoint with a high degree of certainty (if the surrogate variable has been validated), the surrogate can be used instead, in order to assess benefit. However, in many cases surrogates do not allow reliable conclusions to be drawn, or their value is controversial (7).

Determining that a new drug provides added benefit requires evidence, preferably based on direct comparisons, that it offers greater benefit than a current standard treatment (known as the “appropriate comparator treatment”; cf. sections 35a and 130b of the German Social Code [SGB], Part V). The probability and extent of an added benefit are also important (cf. German Law on Benefits of Medicines, AM-Nutzen V). Differences that are demonstrated statistically can be very small and must also be clinically significant (7).

Comparison of the risk/benefit ratio of a new drug with that of an established one must take into account the fact that risks that may only arise after long-term exposure, be very rare, or occur more frequently in patient populations other than the population in which research was conducted may be insufficiently known when approval is granted (8, 9). Both the number of patients treated with the drug before approval and the treatment duration are important for a meaningful assessment of a drug’s risk.

Until now, benefit has been relevant to reimbursement decisions by statutory health insurance funds (section 92, SGB Part V). If Germany’s Federal Joint Committee (G-BA, Gemeinsamer Bundesausschuss) determined that a drug provided no benefit, it could be excluded from statutory coverage (e.g. reboxetine [e3]). Since the beginning of the year 2011, benefit—particularly added benefit—has been determining the pricing of drugs with new active substances (section 35a, SGB Part V). If the Federal Joint Committee determines that there is no added benefit, a new drug will be allocated to a reference price group. If it is determined that there is an added benefit, the National Association of Statutory Health Insurance Funds (GKV-Spitzenverband) and the manufacturer will negotiate its reimbursement price (10).

This descriptive study aims to show the following:

- the study designs and statistical analyses used to demonstrate efficacy and safety for selected proprietary medicinal products (PMPs) that were approved in 2009 and 2010 via the Centralized Procedure (comparator treatment, efficacy testing, endpoints, safety population);
- how often all the information best suited to show the therapeutic value of a new PMP in comparison to alternatives was available.

Methods

PMPs included in the study

The study investigated PMPs that were approved by the EMA in 2009 and 2010 via the EU’s Centralized Procedure, either as new active substances or for extended indications (e4), and launched on the German market by April 2011 (eBox). The following were not included in the study: generics, biosimilars, vaccines, orphan
Drugs, new dosage forms, PMPs with the same active substances, new combinations of active substances already known.

Another requirement for inclusion of a PMP in the evaluation was that information on it had been published in the series *Neue Arzneimittel* (“New Drugs”) issued by the Drug Commission of the German Medical Association (11).

**Analysis of available data**

Our assessment is based on information extracted from European Public Assessment Reports (EPARs) (eBox). The following details were extracted:

- number of pivotal studies
- comparator treatment (e.g. placebo, active control), study design/statistical assessment, primary endpoints
- number of treated patients (safety population).

Testing for superiority allows more far-reaching conclusions (e.g. greater efficacy than an active control) to be drawn than testing for non-inferiority or equivalence to an active control. Assessment therefore took into account how often at least one pivotal study investigated superiority over (any) active control, and in terms of a patient-relevant primary endpoint.

It was ascertained how often a “patient-relevant” primary endpoint was selected, rather than an endpoint with questionable clinical relevance (e.g. response to treatment determined on the basis of only a small reduction in disease activity) or a surrogate endpoint. Classification was performed according to the criteria of the Institute for Quality and Efficiency in Health Care (IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen).

**Results**

**PMPs included in the study**

Table 1 shows the number of PMPs launched on the German market following approval according to the EU’s Centralized Procedure, and the number included in analyses by indication. The frequency of extensions of indications versus initial assessments is also shown.

**Analysis of available data**

Table 2 states the number, design, and statistical assessment of the 81 pivotal studies presented for 39 PMPs included in the evaluation. 51% of approvals are based on only one pivotal study. 64% of the included PMPs (n = 25) are new active substances.
Comparison with an active control was performed in at least one trial for 46% (n = 18) of the approvals (33% of the pivotal studies; n = 27). In 28% (n = 11) of approvals (19% of the trials; n = 15) it was investigated whether the PMP was superior to an active control. For a further 18% (n = 7) of approvals (15% of trials; n = 12), only non-inferiority to an active control was tested.

46% (n = 18) of the approvals (58% of the trials; n = 47) involved comparison to a placebo with no active control arm.

At least one study measured a patient-relevant primary endpoint for 46% (n = 18) of the approvals (44% of trials; n = 36).

Table 3 shows how many patients were treated with the PMP in question in the pivotal studies. The safety of new PMPs or PMPs approved for a new indication was observed in a median of approximately 1700 patients for a period of one year.

For the 11 approvals (28%) for which superiority to an active control was investigated, seven (18%) (7% of the licensing studies) involved a primary endpoint that was patient-relevant (Table 4). Superior efficacy was demonstrated (and cited in the EPAR) in only two of the seven trials.

**Discussion**

The legal requirements of the licensing procedure do not require data from head-to-head studies (1, 12, 13, e5). Active-control trials are only required for approval when the use of a placebo is deemed unethical (e6). Thus this evaluation shows that for approximately half of approvals (2009 to 2010) the pivotal studies conducted only compared the drug with a placebo, but not with an active control. As a result it cannot be ruled out that these PMPs may be inferior to alternatives already on the market. A Dutch study had also shown that data from trials with an active control were available for only 48% of the new drugs approved in the EU between 1999 and 2005 (14).

The EMA has only recently addressed this subject (15) and formulated recommendations for conducting three-arm trials (investigational product, placebo, active control). In the opinion of some authors, these recommendations ought to be more extensive (16, 17). The question remains under which circumstances demonstration of superiority should be a requirement. In the present analysis, we observed that testing for non-inferiority was not always followed by testing for superiority.

Naturally, even without superior efficacy a new drug broadens the range of treatments available if it demonstrates better tolerability or a different risk profile than alternatives (18), for example because it offers advantages for specific patient groups such as those with contraindications to existing approved drugs. However, analysis (in a median of 1700 patients in the safety population) shows that when approval is granted there are insufficient data available to draw a firm conclusion as to better tolerability (8, 9). If an adverse drug reaction (ADR) occurs merely more frequently during treatment (i.e. there is a “background incidence”), larger patient numbers are required to detect the ADR (19). For example, in a comparative study a doubling of the frequency of cerebral hemorrhage from 1% to 2% cannot be ruled out until 3200 patients have been observed in each study arm (20). This is why rare ADRs are often discovered only after market introduction, for example via postmarketing surveillance studies or spontaneous reporting systems.

**Table 3**

<table>
<thead>
<tr>
<th>Safety population</th>
<th>Number of patients treated with the PMP</th>
<th>Number of patients treated with the PMP for at least 1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of approvals included</td>
<td>39</td>
<td>33*1,*2</td>
</tr>
<tr>
<td>Mean</td>
<td>2995</td>
<td>1933</td>
</tr>
<tr>
<td>Median</td>
<td>1696</td>
<td>1696</td>
</tr>
<tr>
<td>Range</td>
<td>85 to 17 157</td>
<td>0 to 17 157</td>
</tr>
<tr>
<td>5th to 95th percentile</td>
<td>213 to 10 331</td>
<td>0 to 9982</td>
</tr>
</tbody>
</table>


1* no information in EPAR for n = 2;
2* only short-term medication envisaged for n = 4
endpoints of different weight such as inpatient admission or death (2).

The aim of this study was to show that available data are limited when approval is granted. It is possible that the EPARs do not represent all particulars of the study data. As EPARs form the basis for approval decisions it is unlikely that data missing in the EPARs would significantly change our conclusions. We also concede that comparisons with active controls may not be possible if there are no therapeutic alternatives, such as in the case of orphan drugs. However, orphan drugs have not been included in this evaluation.

The evaluation shows that a patient-relevant primary endpoint was measured in only approximately half of the trials. The authors admit that the significance of an endpoint may be seen differently, for example regarding the question whether they represent changes in health sufficiently large to be clinically significant and therefore patient-relevant (e.g. ACR 20 [e7, e8]). In principle, surrogate endpoints can also serve as substitutes for patient-relevant endpoints. However, as a rule, conclusions based on them are less reliable than those based on patient-relevant endpoints themselves (e.g. measuring HbA1c levels versus evidence of microvascular complications [e9, e10]).

What should be criticized is the selection of a surrogate such as bone density, for example, when a patient-relevant endpoint could have been measured instead, such as the incidence of fractures (e11). This is particularly true when the surrogate has not been reliably validated. In some cases, however, such as early-stage cancers (21), there may be no good alternatives to selecting a surrogate such as disease-free survival (DFS) (22), as survival time does not allow reliable conclusions to be drawn about the effect of a drug (frequent crossover treatment in patients with long survival times). However, even in these cases it seems justified to indicate that only limited conclusions can be drawn from the available data on new drugs, and that surrogate endpoints must be validated.

The use of new drugs shortly after approval is usually based on information compiled and communicated to physicians by manufacturers, for example via satellite symposia, medical sales representatives, patient brochures, and publications involving opinion leaders. High prices are a particular incentive to promote new drugs intensively; the limited availability of data when a drug is launched on the market provides room for interpretation of its benefits. Timely, unbiased information is therefore important.

When new, and usually more expensive, drugs are prescribed, health insurers incur additional costs. In total, the prescription of more expensive PMPs in 2009 led to an increase in sales of €874 million on the previous year, a cost borne by statutory health insurance funds (known as the “structural component”) (23). Up to the end of 2010, pharmaceutical companies were free to set ex-factory prices in Germany, as there were no legal stipulations ensuring that higher prices of new drugs required evidence of added benefit compared to

### TABLE 4

<table>
<thead>
<tr>
<th>Primary endpoint</th>
<th>Number of trials (approvals)</th>
<th>Patient-relevant endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c</td>
<td>5 (1)</td>
<td>No</td>
</tr>
<tr>
<td>Disease-free survival</td>
<td>1 (1)</td>
<td>No</td>
</tr>
<tr>
<td>Bone density</td>
<td>1 (1)</td>
<td>No</td>
</tr>
<tr>
<td>CE* consisting of cardiovascular mortality, nonfatal myocardial infarction, stroke</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>CE consisting of death, reinfarction, stroke, TVR* in ischemia, severe hemorrhage</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>CE consisting of cardiovascular mortality, cardiac infarction, stroke</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>CE consisting of recurrent atrial fibrillation or trial termination due to intolerability/inefficacy</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>Overall survival</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>IPSS (International Prostate Symptom Score): overall score</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
<tr>
<td>Pregnancy rate</td>
<td>1 (1)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*CE: composite endpoint

TVR = target vessel revascularization
alternatives (cf. [24]). Me-too drugs are an exception to this. Since the Statutory Health Insurance Modernization Act came into force in 2004, so called “pseudo innovations” can be allocated to reference prize groups regardless of their patent duration and therefore are reimbursed only up to maximum reimbursement levels (25). This has removed the incentive for manufacturers to launch me-too drugs on the market at prices above those of pharmacologically and therapeutically comparable drugs.

Official regulations and legal requirements are needed to improve the availability of data when approval is granted and to raise cost-effectiveness on the drug market. Germany’s Pharmaceutical Market Restructuring Act for Statutory Health Insurance (AMNOG) is a first step towards this. As it came into force on January 1, 2011, in the future the prices of all new drugs will for the first time be based on their demonstrated added benefits.

Acknowledgement
The authors would like to thank Henry Pachl of the Drug Commission of the German Medical Association for his helpful research.

Conflict of interest statement
The authors declare that no conflict of interest exists.

Manuscript received on 22 July 2011, revised version accepted on 5 October 2011.

Translated from the original German by Caroline Devitt, MA.

REFERENCES


KEY MESSAGES

- European Public Assessment Reports (EPARs) provide information on study design and statistical analyses used to demonstrate the efficacy and safety of PMPs newly approved in the EU.

- Research into 39 of the PMPs launched on the German market in 2009 and 2010 shows how restricted the availability of data is when approval is granted; for example, how often there is no evidence that a new PMP is not inferior to available alternatives.

- Comparison of new PMPs with those already available must take into account the fact that rare serious adverse reactions may not yet be known when approval is granted due to the limited number of patients treated so far and the short treatment duration.

- Limited trial data provide room for interpretation when new drugs are evaluated in comparison with those already available. Marketing can take advantage of this.

- In Germany, a compulsory early benefit assessment has recently been introduced. For the first time there has been a requirement for higher prices of new drugs to be accompanied by evidence of an added benefit compared to alternatives.


**Corresponding author**

Dr. med. Mariam Ujeyl
Arzneimittelkommission der deutschen Ärzteschaft
Herbert-Lewin-Platz 1, 10623 Berlin, Germany
mariam.ujeyl@akdae.de

@ For eReferences please refer to:
www.aerzteblatt-international.de/ref0712

eBox available at:
www.aerzteblatt-international.de/12m0117
New Drugs: Evidence Relating to Their Therapeutic Value After Introduction to the Market

Mariam Ujeyl, Claudia Schlegel, Siegbert Walter, Ursula Gundert-Remy

eReferences

## Methods

### Proprietary medicinal products (PMPs) included in the study

The PMPs launched on the German market and included in the evaluations in this article were those approved by the European Medicines Agency (EMA) in 2009 and 2010 via the EU’s Centralized Procedure, either as new active substances (initial approval) or for extended indications (e4). Further requirements were the publication of a European Public Assessment Report (EPAR) and the launch of the PMP on the German market by April 2011 (ascertained using WINAPO LAUER-Taxe, the database of marketed PMPs, as of April 1, 2011). The following were not included in the study: generics, biosimilars, vaccines, orphan drugs, new dosage forms, PMPs with the same active substances (e.g. market introduction with different trade names), combinations of active substances already known (for the indication concerned). This is because for these PMPs data were already available, for example, and/or other requirements for approval documents were in place (e.g. evidence of bioequivalence for the approval of generics, evidence of serum protection rates for model vaccines).

Another requirement for inclusion of a PMP in the evaluation was that information on it had been published in the open-access series *Neue Arzneimittel* (“New Drugs”) issued by the Drug Commission of the German Medical Association (11). *Neue Arzneimittel* has existed since 2009 and primarily covers PMPs that are relevant to physicians in private practice. Extensions to indications affecting only small subgroups such as restricted age groups, diagnostic products, and drugs usually prescribed to inpatients, for example, were not included. Any one active substance was included in the evaluations only once, for example in the event of multiple extensions to indications.

### Analysis of available data

Assessments are based on EPAR data. These provide an overview of the trials presented for approval and of scientific discussion of the data by the Committee for Medicinal Products for Human Use (CHMP). The following details were extracted from EPARs where available:

- number of pivotal studies;
- comparator treatment (e.g. placebo, active control);
- study design/statistical assessment (testing for superiority versus testing for non-inferiority/equivalence);
- primary endpoints;
- number of treated patients (safety population).

The term “approval” as used here denotes the approval of a drug for a particular indication, regardless of potency or dosage form.

Although testing for non-inferiority does provide information (e.g. equivalence to an active control), testing for superiority (which may be performed subsequently) allows a more far-reaching conclusion to be drawn regarding the drug’s value (e.g. greater efficacy than an active control). Assessment therefore took into account how often superiority over (any) active control was investigated, and how often a patient-relevant primary endpoint was investigated, in at least one pivotal study.

It was also ascertained how often a “patient-relevant” primary endpoint was selected, rather than an endpoint with questionable clinical relevance (e.g. response to treatment determined on the basis of only a small reduction in disease activity) or a surrogate endpoint. Endpoints were described as “patient-relevant” when they had been classified as such by the Institute for Quality and Efficiency in Health Care (IQWiG, *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen*). Where necessary, first benefit assessments (final reports, preliminary reports, report plans, rapid reports) by the IQWiG, and then EMA guidelines, were consulted for clarification.