Drug Saf (2013) 36:1105–1115
DOI 10.1007/s40264-013-0094-y

ORIGINAL RESEARCH ARTICLE

Post-Approval Safety Issues with Innovative Drugs: A European Cohort Study

Peter G. M. Mol · Arna H. Arnardottir · Domenico Motola · Patrick J. Vrijlandt · Ruben G. Duijnhoven · Flora M. Haaijer-Ruskamp · Pieter A. de Graeff · Petra Denig · Sabine M. J. M. Straus

Published online: 19 September 2013

Abstract

Background At time of approval, knowledge of the full benefit risk of any drug is limited, in particular with regards to safety. Post-approval surveillance of potential drug safety concerns is recognized as an important task of regulatory agencies. For innovative, often first-in-class drugs, safety knowledge at time of approval is often even less extensive and these may require tighter scrutiny post approval.

Objective We evaluated whether more post-approval serious safety issues were identified for drugs with a higher level of innovation.

Methods A cohort study was performed that included all new active substances approved under the European Centralized Procedure and for which serious safety issues were identified post-approval from 1 January 1999 to 1 January 2012. Serious safety issues were defined as issues requiring a Direct Healthcare Professional Communication to alert individual healthcare professionals of a new serious safety issue, or a safety-related drug withdrawal. Data were retrieved from publicly available websites of the Dutch Medicines Evaluation Board and the European Medicines Agency. The level of innovation was scored using a validated algorithm, grading drugs as important (A), moderate (B) or modest (C) innovations or as pharmacological or technological (pharm/tech) innovations. The data were analyzed using appropriate descriptive statistics and Kaplan–Meier analysis, with a Mantel–Cox log-rank test, and Cox-regression models correcting for follow-up duration, to identify a possible trend in serious safety issues with an increasing level of innovation.

Results In Europe, 279 new drugs were approved between 1999 and 2011. Fifty-nine (21 %) were graded as important, 63 (23 %) moderate, or 34 (12 %) modest innovations and 123 (44 %) as non-innovative (pharm/tech), while 15 (25 %), 13 (21 %), 8 (24 %) and 17 (14 %) had post-approval safety issues, respectively ($p = 0.06$, linear-by-linear test). Five drugs were withdrawn from the market. The Kaplan–Meier-derived probability for having a first serious safety issue was statistically significant, log-rank (Mantel–Cox) $p = 0.036$. In the final adjusted Cox proportional hazard model there was no statistically significant difference in occurrence of a first serious safety issue for important, moderate and modest innovations versus non-innovative drugs; hazard ratios 1.76 (95 % CI 0.82–3.77), 1.61 (95 % CI 0.76–3.41)], and 1.25 (95 % CI 0.51–3.06), respectively.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-013-0094-y) contains supplementary material, which is available to authorized users.

P. G. M. Mol (✉) · A. H. Arnardottir · F. M. Haaijer-Ruskamp · P. A. de Graeff · P. Denig
Department of Clinical Pharmacology, University Medical Center Groningen, University of Groningen, FB20, P.O. Box 196, 9700 AD Groningen, The Netherlands
e-mail: p.g.m.mol@umcg.nl

P. G. M. Mol · P. J. Vrijlandt · R. G. Duijnhoven · P. A. de Graeff · S. M. J. M. Straus
Dutch Medicines Evaluation Board (CBG-MEB), Utrecht, The Netherlands

D. Motola
Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

R. G. Duijnhoven
Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, The Netherlands

S. M. J. M. Straus
Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

Conclusion A higher level of innovation was not clearly related to an increased risk of serious safety issues identified after approval.

1 Background

Monitoring risks and benefits of drugs post-approval is an important task of regulatory authorities. At time of approval, knowledge of the full benefit-risk profile of any new drug is incomplete due to well-known limitations of pre-approval research. Most pre-approval clinical research is limited in size, targets narrow well-defined populations, and is primarily focused on establishing efficacy [1]. For innovative drugs, which are often first-in-class with a new mechanism of action, this knowledge will be even less. For these drugs, class-related information on rare safety issues is not available from previously approved comparable substances as would be the case with less innovative, later-in-class drugs [2, 3]. It is estimated that a third of all drugs approved between 1995 and 2003 by the European Medicines Agency (EMA) could be considered important innovations [4]. Especially for the innovative drugs, the post-approval period will be important to establish the full benefit-risk profile.

Although, (serious) safety issues can be identified at any time throughout the lifecycle of any drug [5], there are factors which might contribute to the identification of more safety issues for innovative drugs after their approval. First of all, innovative drugs are eagerly awaited by clinicians and patients, resulting in high pressure to speed up their development, rapid approval and uptake in clinical practice. Pressure put on the regulator to speed up approval was reported to result in more safety issues identified post-approval in Northern America [6–8], but not Europe [9, 10]. As innovative drugs offer a potential therapeutic advantage to existing alternative treatments, they may be rapidly taken up in clinical practice early after approval [11], and they may be channeled to patient groups that are more difficult to treat or who cannot tolerate available treatment [12]. This could result in identification of a greater number of safety issues soon after approval. On the other hand, when the post-approval exposure is small, as with innovative drugs targeting orphan diseases, the power to observe serious adverse drug reactions will be limited [13].

Previously, Tavassoli and Montastruc [11] studied the relation between level of drug innovation and safety alerts as issued by the French Drug Agency. They found an association between these alerts and innovative drugs and recommended to monitor these drugs more actively after approval. They included drugs with at least one report of a serious adverse drug reaction (ADR) to the French pharmacovigilance centers. Due to the cross-sectional design, and the coverage of an arbitrary period in the lifecycle of the drugs, it is unclear whether this observation applies to the early stages post-approval only. Moreover, their results may have been confounded by drug class or drug-use characteristics.

To better understand the relationship between innovation level of a drug and post-approval safety issues, one needs to focus not only on whether a safety issue was identified but also consider the time to such an event and take potential confounders into account. We addressed these issues in our study using an algorithm developed by Motola et al. [4, 14] to assess the level of innovation. This algorithm is based on the availability of alternative treatments for the targeted disease and the level of demonstrated effect on clinically relevant endpoints. The drugs classified as important innovations target diseases where treatment is not available for at least important subgroups of patients and these drugs have demonstrated major benefits on clinical endpoints or established surrogate parameters. The aim of our study was to compare the frequency and timing of serious safety issues identified post-approval in relation to the level of drug innovation in a cohort design, adjusting for potential confounding factors.

2 Methods

2.1 Design

A cohort study was performed that included all new active substances approved under the European Centralized Procedure (CP) and associated serious safety issues that were identified post-approval from 1 January 1999 to 1 January 2012. New active substances were defined according the definition of Eichler et al. [15], excluding biosimilars. We used the algorithm of Motola et al. to assess the level of therapeutic innovation of drugs and therefore we also excluded diagnostics and vaccines since these were not included in this classification. The algorithm graded drugs as (A) important, (B) moderate, or (C) modest innovations or as 'pharmacological/technological' innovations (Fig. 1) [4, 14]. For grading, drugs were first divided into three groups, depending on the availability of alternative treatment possibilities: group (A), drugs for diseases without available therapy; group (B), drugs for subsets of patients unresponsive to available therapy; and group (C), drugs with effective alternative therapy available. In the latter group, class C1 drugs are more effective and/or safe than available alternatives, class C2 drugs are only pharmacologically different (pharmacokinetics or mechanism of action), and

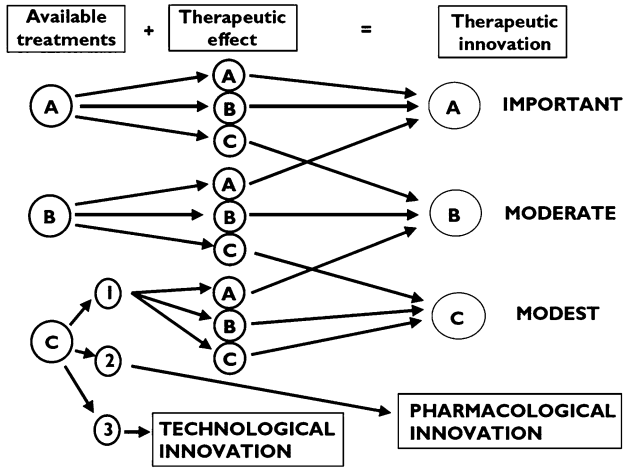


Fig. 1 Classification of drugs. Algorithm used to assign the overall score for innovation. Available treatments: A = drugs for diseases without recognized standard treatment; B = drugs for diseases where subsets of patients are less responsive to marketed drugs and/or other medical interventions, C = drugs for diseases responsive to marketed drugs or other medical interventions (C1 = more effective or safer than existing drugs; C2 = mere pharmacological innovation, i.e. drugs with better kinetics or new mechanism of action; C3 = mere technological innovation, i.e. a new chemical or biotechnological product with therapeutic role similar to already existing ones). Therapeutic effect: A = major benefit on clinical end-points (e.g., increased survival rate and/or quality of life) or validated surrogate end-points; B = partial benefit on the disease (on clinical or validated surrogate end-points) or limited evidence of a major benefit (inconsistent results); C = minor or temporary benefit on some aspects of the disease (e.g., only partial symptomatic relief of a serious disease) (reproduced with permission from Motola et al. [14])

class C3 drugs are only technologically different. Secondly, drugs were graded according to their demonstrated treatment effect; group (A) with a major benefit on clinical endpoint, group (B) with a partial benefit on disease or less robust demonstration of major benefit, and group (C) with minor or temporary benefit on some aspects of the disease.

Motola et al. [4] have previously classified all drugs approved up to 2003. Drugs approved after 2004 were classified independently by two assessors of the Dutch Medicines Evaluation Board and/or co-authors of this paper (PM, PV, PdG, AHA, DM). The team consisted of clinical pharmacologists, clinicians and pharmacists. The level of agreement in classifying the level of innovation was fair only (kappa = 0.31). Therefore, in cases with disagreement, all classifications were resolved in a consensus meeting with a minimum of three assessors.

Scientific and regulatory information on drugs was obtained from the European Public Assessment Report (EPAR) (http://www.emea.europa.eu). The EPAR contains a summary report of the marketing application dossier and scientific assessment.

2.2 Outcome

The primary endpoint was defined as the first issue requiring a Direct Healthcare Professional Communication (DHPC) or a safety-related drug withdrawal. DHPCs were retrieved from the Dutch Medicines Evaluation Board website (http://www.cbg-meb.nl) and only safety-related DHPCs were included. DHPCs related to administration, pharmaceutical quality (including shortage due to manufacturing issues) and/or the malfunctioning of an administration device were excluded. DHPCs issued to inform healthcare professionals on the lack of efficacy of a drug were also excluded. Whether a drug withdrawal was safety-related was determined from the related press release and EPAR 'Procedural steps taken and scientific information after approval' as retrieved from the EMA website.

Time to the primary endpoint, defined as the time in months from the date of market approval to the date of a first DHPC or safety-related withdrawal, whichever came first, was assessed.

2.3 Key Characteristics

Drug and procedural characteristics that could be possible confounders in our analysis were retrieved. The characteristics were chosen based on literature data that suggested an increased risk of drug safety issues post-approval (e.g., drug class [5] and biological drugs [16]). The extent of the patient population exposed to the drug is another possible confounder. The larger the population exposed to a drug, the larger the power to detect adverse effects [7, 17]. The 'channeling' phenomenon—preferential prescribing of new drugs to high risk patients—may be different for drugs that are used by highly specialized prescribers for rare diseases (orphan drugs), for drugs approved under 'exceptional circumstances' or with 'conditional approval' [13]. As of November 2005, Risk Management Plans (RMPs) have become a requirement for new drug applications in Europe. This more pro-active approach in risk management could potentially result in identifying new safety issues earlier or more frequently.

Drug class was classified using the anatomical main group of the Anatomical Therapeutic and Chemical code (ATC-1 level). Three classes (A, J and L) comprise more than 50 % of all drugs and are separately presented, the remaining drugs are grouped in the 'other' category. Second, the type of molecule was categorized as either a biological or a small molecule, identified from the EPARs. Third, the number of drug users was determined from the Drug Information System of the Health Care Insurance Board. This database comprises drug dispensing data for reimbursed drugs in the ambulatory care setting of 26 Dutch healthcare insurance companies, covering nearly all

of the Dutch population (16 million) (http://www.gipdatabank.nl). The number of drug users was split into tertiles, according to the median number of users per year during the period 2007–2011. A separate group was created for the group of drugs not covered by the database as such as drugs not reimbursed, not marketed, or used solely in the hospital setting in the Netherlands. The size of the study population pre-approval was the total number of subjects exposed to the drug for any duration in the clinical development program before approval [17]. Three procedural issues were identified: orphan drug status (y/n); registration type (exceptional circumstances or receiving conditional approval); and whether the drug had been approved before or after Risk Management Plans (RMPs) had become a requirement (November 2005).

2.4 Analysis

Descriptive statistics, Chi-squares and Kruskal–Wallis were used to describe differences in baseline characteristics, across the four different levels of innovation and for drugs with and without a first DHPC/withdrawal. The probability of drugs at different levels of innovation reaching the primary endpoint (first DHPC/safety withdrawal) is evaluated using Kaplan–Meier analysis correcting for follow-up duration; p-value was determined using the log-rank test using a trend analysis. In a multivariable Cox proportional hazard model (HR and 95 % CI), the association between level of innovation and the primary endpoint was determined, correcting for the key characteristics. Characteristics were included in the model (model 1) if they were unevenly distributed across the different levels of innovation or with respect to the occurrence of a serious safety issue ($p < 0.05$). Using backward stepwise regression, only those characteristics were retained in the model that contributed to the model at $p < 0.2$ (model 2).

We performed sensitivity analyses to probe the robustness of our findings in view of the subjective nature of the innovation classification system. First, we dichotomized the classification of innovation grouping drugs graded as level A or B innovations (innovative) and drugs graded as level C or pharm/tech drugs (other). We then performed the same analyses as described above. In the *second* set of sensitivity analyses we compared the heterogeneity of our findings with respect to the classification made by the individual Dutch reviewers. Since the total number of drugs classified by the Dutch team was relatively small (105 in total), we performed these analyses using the dichotomized classification (innovative versus other). Additionally, we evaluated the association for those 71 drugs on which individual assessors agreed on the classification.

We performed an additional post-hoc analysis comparing our classification with other classifications of drug innovation as used by the Canadian Human Drug Advisory Panel (HDAP) and Prescrire International. We used the dichotomized classifications that were used by Lexchin in his paper in Health Policy of 2012 [18] to facilitate comparison with our classification set. The number of drugs that were graded by all three systems was 65. Again considering this limited number of drugs on which we could compare, we used the dichotomized safety outcome and present Chi-square and log-rank test results only. We calculated kappa values to determine the agreement between the three systems in grading the level of drug innovation.

3 Results

A total of 633 drugs obtained a marketing authorization in our study period from 1 January 1999 to 1 January 2012. We excluded 354 drugs that could not be classified by Motola's algorithm of therapeutic innovation; 309 drugs contained no new active substances, 35 were vaccines and 10 were diagnostics (Fig. 2). Of the remaining 279 new drugs, 59 (21 %) were graded as important (grade A), 63 (23 %) as moderate (B), and 34 (12 %) as modest (C), and 123 (44 %) as pharmacological/technological (pharm/tech) innovations (Table 1, individual classifications are provided in the

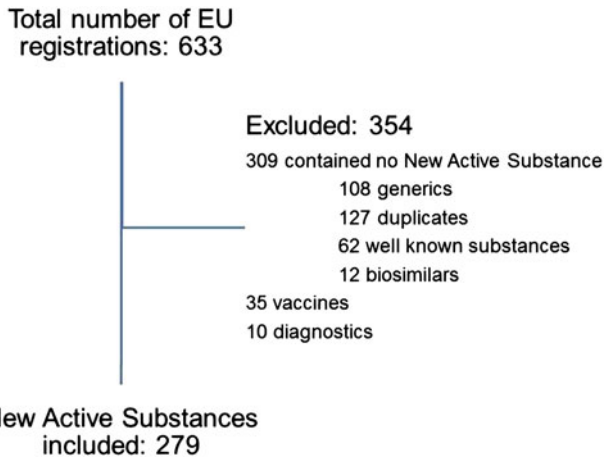


Fig. 2 Flow chart of study drugs. Number of registrations with the European Medicines Agency (EMA) through the centralized procedure from 1 January 1999 to 31 December 2011. Duplicates are drugs registered under different trade names; known substances are drugs approved earlier either through national approval procedures (e.g., Tobi Podhaler containing tobramycin) and/or that were approved earlier for a different indication under a different registration number (e.g., sildenafil as Viagra™ for erectile dysfunction and subsequently as Revatio™ for pulmonary hypertension); combinations with at least one new active substance were included in our study (e.g., Tredaptive™ for dyslipidemia that contained the new active substance laropiprant in addition to the known substance nicotinic acid), but not those containing only known active substances

Table 1 Key characteristics across innovation level of drugs in Europe (1999–2011)

	All drugs *N* (%)[a]	Innovation level of drugs *N* (%)[a]				
		Important (A)	Moderate (B)	Modest (C)	Pharm/tech	*p*[b]
Total	279 (100)	59 (100)	63 (100)	34 (100)	123 (100)	
Drug characteristics						
Drug classes (ATC-1 level)						**<0.001**
Alimentary tract and metabolism (A)	39 (14)	7 (12)	7 (11)	4 (12)	21 (17)	
Anti-infectives for systemic use (J)	38 (14)	21 (36)	7 (11)	1 (3)	9 (7)	
Antineoplastic and immune-modulating agents (L)	75 (27)	23 (39)	19 (30)	16 (47)	17 (14)	
Other drug classes[c]	127 (46)	8 (14)	30 (48)	13 (38)	76 (62)	
Number of drug users per year[d]						**<0.001**
Hospital use only or not reimbursed in NL	121 (43)	31 (53)	33 (53)	18 (53)	39 (32)	
≥228	53 (18)	14 (24)	15 (24)	8 (24)	16 (13)	
<228 and ≥2.117	53 (19)	12 (20)	8 (13)	5 (15)	28 (23)	
<2.117	52 (19)	2 (3)	7 (11)	3 (9)	40 (33)	
Biologicals (y)	83 (30)	20 (34)	17 (27)	10 (29)	36 (29)	0.866
Size of study population, median (IQR)[e]	1,227 (522–2,437)	900 (511–1,712)	1,041 (373–2,150)	1,155 (397–2,949)	1,789 (735–3,410)	**0.001**
Procedural characteristics						
Orphan drugs (y)	59 (21)	15 (25)	22 (35)	13 (38)	9 (7)	**<0.001**
EC and CA registrations (y)	47 (17)	19 (22)	18 (29)	7 (21)	2 (2)	**<0.001**
Post-RMP approval (y)	139 (50)	23 (39)	42 (67)	22 (65)	52 (42)	**0.001**

Bold values are statistically significant at $p < 0.05$

[a] Percentages are expressed within drugs, and level of innovation (column)

[b] *p*-value based on Chi-square (categorical data) and Kruskal–Wallis (continuous data) test

[c] All drugs that are not categorized on ATC-1 level as A (alimentary tract and metabolism), J (anti-infectives for systemic use) or L (antineoplastic and immunomodulating agents)

[d] Median number of users per year in the Netherlands based on 2007–2011 reimbursement data from the Drug Information System of the Health Care Insurance Board

[e] European Public Assessment Reports (EPARs) were not available for six drugs thus the size of the study population could not be established. For the analyses values were imputed based on the mean number of patients studied in the clinical development program for drugs with the same ATC-5 or if not available the same ATC-3 level

ATC therapeutic and chemical code, *CA* conditional approval, *EC* exceptional circumstances, *IQR* interquartile range showing 25th and 75th percentiles, *NL* The Netherlands, *RMP* risk management plan; this variable indicates whether drugs were approved after RMPs became a requirement for new drug applications in the European Union (Nov 2005), *(y)* variable is dichotomous and value represents the 'yes'

Electronic Supplementary Material). Per year, the number of approvals ranged between 12 and 26 (Fig. 3).

In total, 114 DHPCs were issued, of which 25 DHPCs for 15 drugs were issued that were not safety-related and were excluded. Nine of these excluded DHPCs concerned the administration of the drug, ten concerned the pharmaceutical quality of the drug product and two concerned the malfunctioning of an administration device. Four DHPCs for two drugs classified as important innovations (drotrecogin alfa, and three DHPCs for tenofovir) were issued because of lack of efficacy and were also excluded. This resulted in 53 first and 36 repeated DHPCs that were included in our study.

Five drugs (<2 % of all drugs approved) were withdrawn for safety-related reasons; efalizumab, sitaxentan, rimonabant (all grade B), rosiglitazone and valdecoxib (both pharm/tech).

Overall, 53 first serious safety events (primary endpoint) were identified for 15 (25 %) drugs graded A, 13 (21 %) graded B, eight (24 %) graded C, and 17 (14 %) for the drugs graded as pharm/tech ($p = 0.23$, Chi-square test, $p = 0.06$ linear-by-linear association, Table 2).

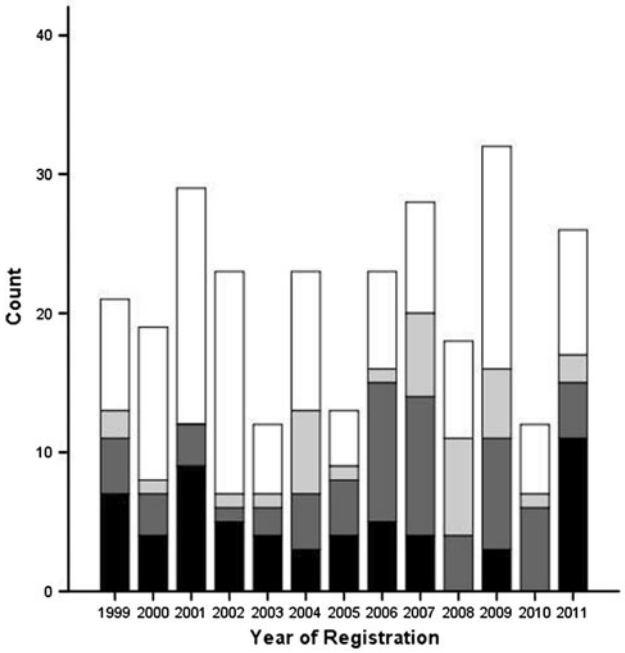


Fig. 3 Level of innovation of drugs approved in Europe (1999–2011). Drugs are new active substances that are approved through the centralized procedure in Europe. Classification of innovation according to Motola et al. [14]; (A) important (*black*), (B) moderate (*dark grey*), (C) modest (*light grey*), (Pharm & Tech) merely pharmacological/technological innovations (*white*)

The median follow-up, from marketing approval to end of study period or drug withdrawal, was 7.8 years (interquartile range [IQR] 4.0–10.4) for drugs graded A, 4.6 years (IQR 2.4–6.9) graded B, 4.2 years (IQR 3.2–7.2) graded C, and 7.3 years (IQR 2.9–10.4) for the drugs graded as pharm/tech ($p = 0.01$, Kruskal–Wallis test). The Kaplan–Meier-derived probability for having a first serious safety issue, correcting for duration of follow-up, suggests a trend in safety issues with increasing level of innovation, log-rank (Mantel–Cox) $p = 0.036$. Subsequently, the association between level of therapeutic innovation of a drug and a first serious safety issue post-approval was determined in a multivariable Cox model. All identified key drug and procedural characteristics, except whether a drug was a biological or small molecule, were unequally distributed across the various levels of innovation ($p < 0.05$) and therefore included in the multivariable Cox model (Table 1). All key characteristics were equally distributed across drugs with or without serious safety issues. For six drugs, we were unable to retrieve one drug characteristic, i.e., the size of the pre-approval study population, as the scientific discussion had been removed from the EMA website after withdrawal of these products from the market. For these drugs we imputed the study population based on drugs approved within the same ATC class (interferon alfacon-1 [1197], valdecoxib [3550], dofetilide [3410], apomorphine [2764] and fomivirsen [1328]). Except valdecoxib, all withdrawals were for commercial reasons and thus not recorded as a serious safety issue.

In the multivariable Cox proportional hazards model (model 1) there was no statistically significant difference in occurrence of a first serious safety issue across the more innovative (A, B, and C) drugs when compared with the pharm/tech (reference) drugs (Table 2). The results remained similar in model 2, which, in addition to the level of innovation, incorporated only potential confounders with a $p < 0.2$. These were drug class, size of the pre-approval study population, and whether the drug had been approved after RMPs had become a requirement. The adjusted hazard ratios were: for grade A drugs 1.76 (95 % CI 0.82–3.77), grade B drugs 1.61 (95 % CI 0.76–3.41), and for grade C drugs 1.25 (95 % CI 0.51–3.06) (Table 2; Fig. 4). Antineoplastic and immunomodulating drugs (ATC-class L) and the size of the study population in the clinical program were the only significant confounders ($p < 0.05$).

The *first* sensitivity analysis showed that a numerically higher proportion of innovative drugs (grade A and B) had a first serious safety event than all other (grade C and pharm/tech combined) drugs, 23 versus 16 %, respectively. Although the difference was not statistically significant ($p = 0.14$; Chi-square test), the Kaplan–Meier curves still separated, $p = 0.07$ (log-rank test, Fig. 5a). The Cox proportional hazards models, HR 1.83 (95 % CI 0.99–3.39) model 1 and HR 1.71 (95 % CI 0.94–3.11) model 2, showed similar not statistically significant results. The *second* sensitivity analysis for 105 drugs for which individual assessments of level of innovation were available showed again that event rates were numerically higher for the innovative drugs and that Kaplan–Meier curves separated (Fig. 5b, c). These findings were not significant. For the 71 drugs where both reviewers agreed in their first assessment, the effects were also similar to our original analysis (Fig. 5d). The post-hoc comparison using two different classification systems showed that agreement in grading of drug innovation between the three systems was poor. Kappa values ranged from 0.03 (HDAP vs our model), 0.06 (Prescrire International vs our model) to 0.20 (HDAP vs Prescrire International). HDAP, Prescrire International and our model classified 10, 9 and 33, respectively, as innovative of the 65 drugs that were classified by all three systems. This subsequently resulted in no observable association with the occurrence of serious safety events (p values from 0.66 to 0.97 in Chi-squares and log-rank tests) for the HDAP (Fig. 5e) and Prescrire International (Fig. 5f) classifications.

Table 2 Are serious safety issues and level of innovation associated (Cox-proportional hazards analyses) for centrally approved drugs in Europe (1999–2011)

	All drugs *N* (%)[a]	Serious safety issue (*N* %)			Cox-proportional hazards analyses	
		Yes	No	*p*[b]	Model 1	Model 2
Total	279 (100)	53 (19)	226 (80)			
Level of innovation				0.232		
Important (A)	59 (100)	15 (25)	44 (75)		2.09 (0.90; 4.87)	1.76 (0.82; 3.77)
Moderate (B)	63 (100)	13 (21)	50 (79)		2.03 (0.92; 4.47)	1.61 (0.76; 3.41)
Modest (C)	34 (100)	8 (24)	26 (77)		1.49 (0.59; 3.72)	1.25 (0.51; 3.06)
Merely pharmacological or technological (pharm/tech)	123 (100)	17 (14)	106 (86)		Reference	Reference
Drug characteristics						
Drug classes (ATC-1 level)				0.137		
Alimentary tract and metabolism (A)	39 (100)	6 (15)	33 (85)		1.24 (0.47; 3.23)	1.06 (0.42; 2.71)
Anti-infectives for systemic use (J)	38 (100)	10 (26)	28 (74)		1.54 (0.61; 3.88)	1.85 (0.78; 4.37)
Antineoplastic and immune-modulating agents (L)	75 (100)	19 (25)	56 (75)		**2.25 (1.08; 4.70); p = 0.031**	**2.35 (1.14; 4.85); p = 0.021**
Other drug classes[c]	127 (100)	18 (14)	109 (86)		Reference	Reference
Number of drug users per year[d]				0.062		
Hospital use only/not reimbursed in NL	121 (100)	19 (16)	102 (84)		1.06 (0.43; 2.61)	–
≤228	53 (100)	9 (17)	44 (83)		1.17 (0.41; 3.32)	–
>228 and ≤2.117	53 (100)	17 (32)	36 (68)		2.00 (0.78; 5.11)	–
>2.117	52 (100)	8 (15)	44 (85)		Reference	–
Biologicals (y)	83 (100)	16 (19)	67 (80)	0.938	–	–
Size of study population, median (IQR)[e]	1,227 (522–2,437)	1,304 (877–2,998)	1,220 (462–2,347)	0.162	**1.00 (1.00; 1.00); p = 0.012**	**1.00 (1.00; 1.00); p = 0.005**
Procedural characteristics						
Orphan drugs (y)	59 (100)	11 (19)	48 (81)	0.938	0.80 (0.35; 1.83)	–
EC and CA registrations (y)	46 (100)	8 (17)	38 (83)	0.761	1.48 (0.65; 3.36)	–
Post-RMP approval (y)	139 (100)	22 (16)	117 (84)	0.179	1.59 (0.82; 3.10)	1.63 (0.87; 3.04)

Bold values are statistically significant at $p < 0.05$

[a] Percentages are expressed within groups/characteristics (rows)

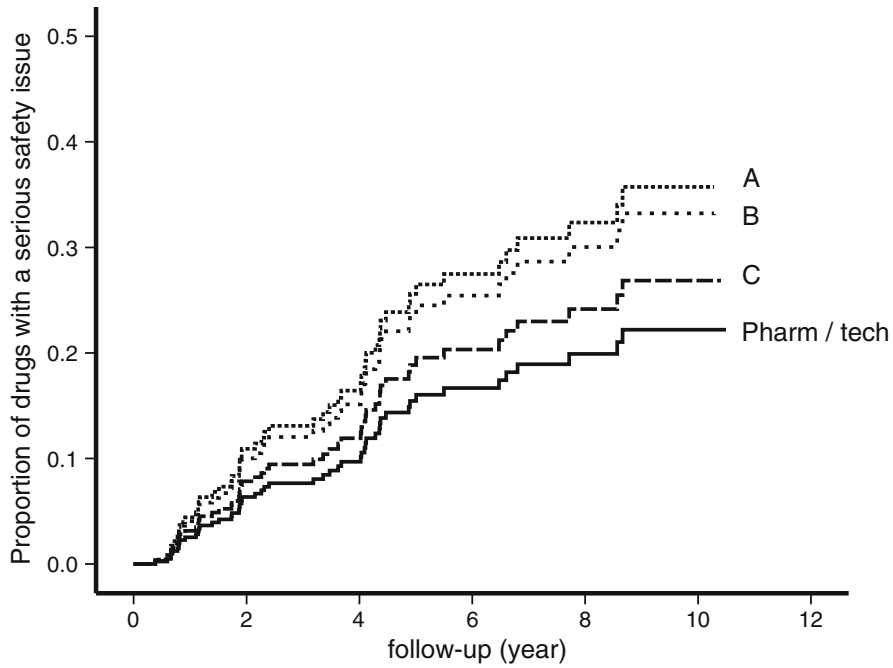
[b] *p*-value based on Chi-square (categorical data) and Kruskal–Wallis (continuous data) test

[c] All drugs that are not categorized on ATC-1 level as A (alimentary tract and metabolism), J (anti-infectives for systemic use) or L (antineoplastic and immunomodulating agents)

[d] Median number of users per year in the Netherlands based on 2007–2011 reimbursement data from the Drug Information System of the Health Care Insurance Board

[e] European Public Assessment Reports (EPARs) were not available for six drugs thus the size of the study population could not be established. For the analyses values were imputed based on the mean number of patients studied in the clinical development program for drugs with the same ATC-5 or if not available the same ATC-3 level

ATC therapeutic and chemical code, *CA* conditional approval, *EC* exceptional circumstances, *IQR* interquartile range showing 25th and 75th percentiles, *NL* The Netherlands, *RMP* risk management plan; this variable indicates whether drugs were approved after RMPs became a requirement for new drug applications in the European Union (Nov 2005), *(y)* variable is dichotomous and value represents the 'yes'



Grade	Drugs remaining in model						
A	59	41	36	26	20	11	4
B	63	47	33	15	11	8	2
C	34	28	17	11	5	3	1
Pharm / tech	123	101	76	60	42	27	5

Fig. 4 Proportion of first serious safety issue according to level of innovation (Model 2*: adjusted Cox-proportional hazards model). Model 2 hazard ratios of level of innovation are corrected for drug class (ATC-1 level), size of study population and post-RMP approval. *A* Important drug innovations, *B* moderate drug innovations, *C* modest drug innovations, *Pharm/tech* pharmacological or technological drug innovations, *ATC* anatomical therapeutic and chemical code, *RMP* risk management plan

4 Discussion

Between 1999 and 2011, 21 % of all new drug approvals can be considered important innovations, while nearly half (44 %) can be considered as 'me-too' (pharmacological or technological) developments. Nineteen percent (53/273) of all centrally approved drugs received one or more DHPCs during the study period and five (<2 %) drugs were withdrawn. Our study showed that the level of innovation was not clearly correlated with frequency and timing of serious safety issues that are identified post-approval.

Serious safety issues have been suggested to be more common with biological agents versus small molecules [16], while fewer issues were identified with orphan drugs [13]. Fast-track approval or 'priority review' may be associated with identification of more safety issues post-approval in Canada and the USA [7, 8]. Both Health Canada and the Food and Drug Administration (FDA) offer 'priority review' for drugs that are intended to treat life-threatening disease or are considered important therapeutic advancements, somewhat similar to drugs classified as important innovations in our study. More post-approval safety issues were identified also for drugs approved by the FDA close to the review deadlines in comparison to drugs approved with more remaining review time [6]. In contrast, drugs approved through exceptional circumstances or conditional approval procedures that are aimed to facilitate drug approval for the same type of drugs in Europe did not show an increased risk [9, 10]. The difference may be that in Europe there is not necessarily a shorter review period for this type of drug, allowing a robust assessment procedure. Our current findings indicate that the level of innovation is not clearly related to identification of serious post-approval safety issues. Our sensitivity analyses support our principal findings, although there are some numerical differences. A trend in separation of curves was observed also when our grading of innovation was dichotomized, and was independent of whether individual assessments or only cases of agreement were considered.

Possibly, as for the more innovative drugs, additional pharmacovigilance activities to closely monitor and minimize risks post-marketing may have been implemented. In recent years, regulators have, under large public pressure and in the wake of several high-profile drug safety issues [19], intensified their pharmacovigilance activities [20, 21]. Grading of level of innovation could be considered as a possible additional tool to guide problem-based risk management for new drugs that is currently done on a case-by-case basis.

Not all classification systems of innovation may be equally valuable for this purpose. The classification systems of the Canadian HDAP and Prescrire International

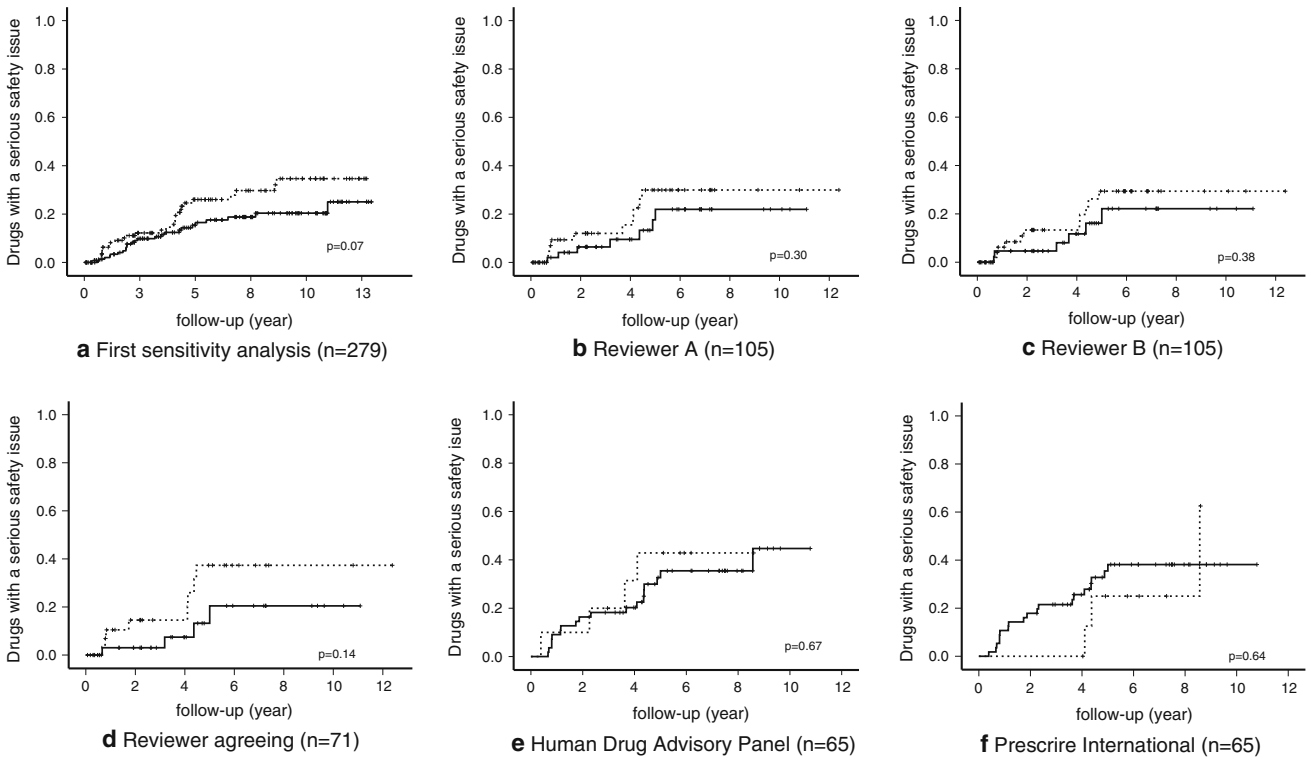


Fig. 5 Survival curves of innovative versus other drugs with serious safety issues post-approval (sensitivity analyses). **a** First sensitivity analyses where grade A and B drugs in our classification system are grouped as 'innovative' drugs (*dotted line*) and grade C and pharmacological and technological drugs as 'others' (*solid line*). The *p* value mentioned here is based on the log-rank test. The *y*-axis presents the proportion of drugs with a first serious safety issue post-approval. The *x*-axis presents the number of years of follow-up post-approval. **b**–**d** Second sensitivity analyses where the 105 drugs classified by the Dutch review team are presented. Reviewer A and B show initial classifications as made by a single independent reviewer. 'Reviewers agreeing' presents the set of drugs on which reviewers agreed in their initial classification of level of innovation ($n = 71$). Finally, **e** and **d** represent the survival curves when the classifications of drug innovation of the Canadian Human Drug Advisory Panel and Prescrire International are used

classified only a few drugs as innovative, with no relationship to the occurrence of serious safety issues [18]. The poor agreement in classification of level of drug innovation calls for some form of harmonization. Which classification system is most suitable, and based on what criteria (e.g., demonstrated efficacy, cost effectiveness, available alternative therapies, etc.), should be the subject of further research, one such European model could possibly also be useful to guide pharmacovigilance activities as discussed above.

Only two of the studied characteristics were true confounders in the association of innovative drugs and a serious safety event. For antineoplastic and immunomodulating drugs, we observed a doubling of the risk of a serious post-approval safety event independent of the level of innovation. These drugs have been shown to be more frequently associated with safety-related regulatory actions [5, 22]. Still, accelerated approval of cancer drugs may be considered safe, as only a few drugs are withdrawn, and these procedures are therefore useful in facilitating earlier access to important new treatment options [22]. Another reason why more safety issues are observed with these drugs is that they are used relatively often outside of their approved indication, where safety experience is even more limited [23].

In our study, a larger size of the pre-approval study population was associated with an increased risk of a serious safety issue post-approval. The size of the clinical development programs was larger for drugs graded as pharmacological or technological innovations. The 'me-too' products are developed to get a share of large drug markets [24, 25], and can therefore easily include more patients in the clinical development program without really generating new drug safety knowledge. Of note, not the level of use in clinical practice, nor orphan drug designation nor the approval procedure affected our model. Orphan drugs and drugs approved through the Exceptional Circumstances or Conditional Approval procedure are often used in highly specialized treatment centers and reimbursed for the very specific patients that are often also included in the pre-approval studies. They are less likely to be channeled to high risk populations after marketing. Our

postulated effects of rapid uptake and channeling of innovative drugs seemed thus not to have played any important role with respect to identification of safety issues for these drugs.

Although innovative drugs have been approved throughout the study period, it is of some concern that the overall proportion of innovative drugs is lower than that reported in 2006 [4]. This finding fits in the debate about the declining efficiency in drug development, drug approval success rates and lack of truly novel drug products [15, 26, 27] and warrants further consideration. Although, for approximately a fifth of all drugs a new serious safety issue was identified post-approval, this only resulted occasionally in the drug being withdrawn from the market.

4.1 Limitations

Our study only evaluated centralized-approved products in Europe, which does not—especially in the earlier period—cover all important new drug approvals across the various European countries. This is, however, unlikely to have affected our primary analysis of whether level of innovation was associated with the probability of identifying serious safety issues post-approval. Innovative and non-innovative drugs were seemingly randomly distributed across the study years and although the total follow-up period (i.e., until the end of the study period or event) had a sort of U-shaped curve across the levels of innovation, this should allow for a proper analysis of time to event.

The power to detect differences may have been limited. Since there is no accepted level of relevant difference in the proportion of safety issues, it is difficult to conduct a formal power calculation. We based our analysis on the total number of approved products in Europe, almost from the inception of the EMA in 1995. We argue that if a difference cannot be observed after a maximal 13 years (or a median of 8 years) of follow-up this difference may be less relevant. The sample size was sufficiently large to find a significant difference in proportion of drugs with a first serious safety issue post-approval in the log-rank test. Although, when we corrected for relevant confounders, that difference did not remain significant. Repeating this study in 5 years, for example, may therefore be worthwhile.

Any assessment has an inherent level of subjectivity to it. The classification used by us is no exception. We performed independent assessments for each drug and resolved any discrepancies in consensus. This was a laborious exercise but strengthened the robustness of the classification. In addition, we conducted several sensitivity analyses. A possible limitation is the retrospective nature of assessing the level of innovation. However, the EMA keeps the EPARs of the original market application at its website, which we used for assessing therapeutic innovation at the time of initial approval. Finally, our data on drug exposure were limited. We only had access to a national drug use database with ambulatory care reimbursement claims for the period 2007–2011 and used this as a proxy for overall drug exposure. This might not be representative for the whole of Europe or for worldwide uptake. In addition, the restricted period made it impossible to incorporate drug use as a time-varying covariate in our model, which would better reflect the true drug exposure in relation to the safety event.

5 Conclusion

Our study confirms that more than 40 % of new approved drugs represent important or moderate therapeutic innovations and showed that, after adjusting for potential confounders, the level of innovation was not clearly related to the frequency and timing of serious safety issues identified after marketing.

Acknowledgments This study was performed in part supported by an unconditional grant in the context of the Escher project (T6-202), a project of the Dutch Top Institute Pharma. The funder did not have any role in the design or execution of the study. The authors who are employed by the MEB (Pieter de Graeff, Sabine Straus, Patrick Vrijlandt, Ruben Duijnhoven and Peter Mol), state that the opinions in this paper are their own, and do not necessarily reflect those of the MEB.

Conflict of interest Peter G.M. Mol, Arna H. Arnardottir, Domenico Motola, Patrick J. Vrijlandt, Ruben G. Duijnhoven, Flora M. Haaijer-Ruskamp, Pieter A. de Graeff, Petra Denig and Sabine M.J.M. Straus report no conflicts of interest that are directly relevant to the content of this study.

References

1. Califf RM. Benefit assessment of therapeutic products: the Centers for Education and Research on Therapeutics. Pharmacoepidemiol Drug Saf. 2007;16(1):5–16.
2. Arnardottir AH, Haaijer-Ruskamp FM, Straus SMJM, de Graeff PA, Mol PG. Safety issues of HIV drugs, regulatory consequences. Drug Saf. 2011;34(11):1101–14.
3. Stefansdottir G, Knol MJ, Arnardottir AH, van der Elst ME, Grobbee DE, Leufkens HG, et al. Safety learning from drugs of the same class: room for improvement. Clin Pharmacol Ther. 2012;91(5):872–80.
4. Motola D, De Ponti F, Poluzzi E, Martini N, Rossi P, Silvani MC, et al. An update on the first decade of the European centralized procedure: how many innovative drugs? Br J Clin Pharmacol. 2006;62(5):610–6.
5. Mol PG, Straus SM, Piening S, de Vries JT, de Graeff PA, Haaijer-Ruskamp FM. A decade of safety-related regulatory action in the Netherlands: a retrospective analysis of direct healthcare professional communications from 1999 to 2009. Drug Saf. 2010;33(6):463–74.
6. Carpenter D, Zucker EJ, Avorn J. Drug-review deadlines and safety problems. N Engl J Med. 2008;358:1354–61.

7. Moore TJ, Furberg CD. The safety risks of innovation: the FDA's Expedited Drug Development Pathway. JAMA. 2012;308(9):869–70.
8. Lexchin J. New drugs and safety: what happened to new active substances approved in Canada between 1995 and 2010? Arch Intern Med. 2012;172(21):1680–1.
9. Arnardottir AH, Haaijer-Ruskamp FM, Straus SM, Eichler HG, de Graeff PA, Mol PG. Additional safety risk to exceptionally approved drugs in Europe? Br J Clin Pharmacol. 2011;72(3):490–9.
10. Boon WP, Moors EH, Meijer A, Schellekens H. Conditional approval and approval under exceptional circumstances as regulatory instruments for stimulating responsible drug innovation in Europe. Clin Pharmacol Ther. 2010;88(6):848–53.
11. Tavassoli N, Montastruc J. Is there any relationship between actual benefit and added value of drugs and pharmacovigilance alerts? Br J Clin Pharmacol. 2009;68(1):124–5.
12. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991;10(4):577–81.
13. Heemstra HE, Giezen TJ, Mantel-Teeuwisse AK, de Vrueh RL, Leufkens HG. Safety-related regulatory actions for orphan drugs in the US and EU: a cohort study. Drug Saf. 2010;33(2):127–37.
14. Motola D, De Ponti F, Rossi P, Martini N, Montanaro N. Therapeutic innovation in the European Union: analysis of the drugs approved by the EMEA between 1995 and 2003. Br J Clin Pharmacol. 2004;59(4):475–8.
15. Eichler HG, Aronsson B, Abadie E, Salmonson T. New drug approval success rate in Europe in 2009. Nat Rev Drug Discov. 2010;9(5):355–6.
16. Giezen TJ, Mantel-Teeuwisse AK, Straus SMJM, Schellekens H, Leufkens HGM, Egberts ACG. Safety-related regulatory actions for biologicals approved in the United States and the European Union. JAMA. 2008;300(16):1887–96.
17. Duijnhoven RG, Raine JM, Straus SMJM, De Boer A, Hoes AW, De Bruin ML. Number of patients studied prior to approval of new medicines: a database analysis. PLoS Med. 2013;10(3):e1001407.
18. Lexchin J. International comparison of assessments of pharmaceutical innovation. Health Policy. 2012;105(2–3):221–5.
19. Morrato EH, Staffa JA. Effectiveness of risk management plans: a case study of pemoline using pharmacy claims data. Pharmacoepidemiol Drug Saf. 2007;16(1):104–12.
20. Morrato EH, Ling SB. The Drug Safety and Risk Management Advisory Committee: a case study of meeting frequency, content, and outcomes before and after FDAAA. Med Care. 2012;50(11):970–86.
21. Raine J, Wise L, Blackburn S, Eichler HG, Breckenridge A. European perspective on risk management and drug safety. Clin Pharmacol Ther. 2011;89(5):650–4.
22. Richey EA, Lyons EA, Nebeker JR, Shankaran V, McKoy JM, Luu TH, et al. Accelerated approval of cancer drugs: improved access to therapeutic breakthroughs or early release of unsafe and ineffective drugs? J Clin Oncol. 2009;27(26):4398–405.
23. Leveque D. Off-label use of anticancer drugs. Lancet Oncol. 2008;9(11):1102–7.
24. Regnier S. What is the value of 'me-too' drugs? Health Care Manage Sci (Epub 26 Feb 2013).
25. Hollis A. Me-too drugs: is there a problem? 2004. http://www.who.int/intellectualproperty/topics/ip/Me-tooDrugs_Hollis1.pdf (Accessed 25 June 2013).
26. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. Clin Pharmacol Ther. 2010;87(3):272–7.
27. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11(3):191–200.