University of Heidelberg

Department of Economics

Discussion Paper Series | No. 607

Precision requirements in pesticide risk assessments: Contrasting value-of-information recommendations with the regulatory practice in the EU

Timo Goeschl and Daniel Heyen

February 2016

Precision requirements in pesticide risk assessments: Contrasting value-of-information recommendations with the regulatory practice in the EU*

Timo Goeschl
Department of Economics
Heidelberg University

Daniel Heyen†
Department of Economics
Heidelberg University

January 31, 2016

Abstract

Pesticides, while rendering immense agricultural benefits, potentially entail risks to human health and the environment. To limit these risks, market approval of a pesticide is typically conditional on an extensive risk assessment demonstrating its safety. The associated testing procedures, often involving significant numbers of animals, however are not only costly; as has become apparent from recent discussions about the active substance glyphosate, testing is often incapable of providing definitive answers on concerns like human carcinogenicity. An important regulatory task, whether explicitly acknowledged or not, is hence to decide what level of remaining uncertainty is deemed acceptable in making the final market approval decision. Economic principles suggest a value-of-information (VoI) approach for this informational task. After presenting the basics of the VoI framework, this paper analyzes the actual regulatory practice in the EU's pesticide approval process, pointing out the defaults and substance-specific procedures that shape the precision of the European Food Safety Authority's (EFSA) risk assessment and hence the level of knowledge under which the European Commission decides on the approval of substances. The comparison between theory and practice uncovers substantial deviations, providing valuable insights for restructuring the risk assessment guidelines.

Keywords: risk assessment; pesticide; regulation; value-of-information; animal testing; uncertainty; active learning.
JEL codes: D81; D83; Q58; I18.

1 Introduction

Glyphosate, the active substance in plant protection products (PPP) like Monsanto's Roundup®, is widely used and has been providing immense agricultural benefits for

*The authors gratefully acknowledge the support of the German Research Foundation, grant no. GO 1604/2.

†Corresponding author. Email: heyen@eco.uni-heidelberg.de

decades. At the same time, there have always been concerns whether the innate toxicity of active substances carries on to humans or has adverse effects on biodiversity and the environment (Aktar et al. 2009; Baylis 2000; Duke and Powles 2008). Surrounding the re-approval process in the EU we have recently witnessed intense debate about glyphosate's potential carcinogenicity with conflicting views. After the renewal assessment report (RAR) by member state Germany (Germany 2013) had indicated no evidence for carcinogenicity, the WHO's International Agency for Cancer Research (IARC) classed glyphosate as "probably carcinogenic to humans" (IARC 2015). This induced the European Commission to provide the European Food Safety Authority with an additional mandate to evaluate the IARC report. The EFSA's conclusion on glyphosate (EFSA 2015a), building the final point of the EU's risk assessment, challenges the relevance of IARC's findings and endorses the RAR evaluation. With the current permission expiring in late June 2016, the European Commission's decision on re-approval of glyphosate is expected in early 2016.

The conflicting views in the assessment of glyphosate's carcinogenic potential point to a pervasive regulatory challenge of modern societies. A vast number of regulatory decisions, including market approval decisions of new technologies, transgenic crops, pharmaceuticals and active substances like glyphosate, have to be made under substantial uncertainty. The reason is that the testing procedures typically required for a positive approval decision, despite the substantial costs they entail (Bottini and Hartung 2009), are at best partially predictive of actual effects on human health and the environment. For instance, the toxicity tests with animals, being part of the standard requirements for a risk assessment of active substances and other chemicals, has a very limited sensitivity (Abbott 2005; Hartung 2008). Deciding on type and extent of testing requirements is hence a choice of the level of confidence the regulator deems necessary for making the final market approval decision. Or, tu put it differently, a choice on the level of *remaining uncertainty* he is willing to accept.

This societal choice on precision requirements for risk assessments is the focus of this paper. In particular, we will contrast recommendations based on economic principles with the current regulatory practice. With glyphosate as the motivating example to which we recurrently return, our main focus is the approval process of active substances in the EU. We however expect many findings to carry on to other regulatory challenges and jurisdictions.

The paper is organized as follows. Section 2 presents useful background knowledge on the organization and theory of risk regulation, particulary the ubiquitous regulatory separation into risk assessment and risk management. Section 3 presents a simplified decison-theoretic model that underpins the economic value-of-information rationale for determining the welfare maximizing level of precision. Section 4 summarizes the current regulatory practice of pesticide risk assessment and management in the EU. Section 5 contrasts the regulatory practice with the theory-based recommendations derived in sec-

tion 3. Section 6 concludes and discusses important aspects for reshaping the regulatory design.

2 Regulatory background: Risk assessment and risk management

The regulatory challenge of preventing health and environmental disasters while, at the same time, enabling society access to the vast benefits of novel technologies and products has sparked a vast literature on the appropriate design of risk regulation. Irrespective of the regulatory task and respective regulatory stance, the regulatory process is typically understood to consist of two steps (e.g. NRC 1983; Haimes 2005; Renn 2006). Figure 1 gives a graphical representation. The first step, *risk assessment* (RA), aims at providing information about the problem in question and hence reducing detrimental uncertainty. This improved state of knowledge supports *risk management* (RM), the second step of the process, which then determines the final societal outcomes.[1] In the concrete glyphosate example, uncertainty exists – among other things – about whether the substance is carcinogenic; EFSA's RA provides the basis for the European Commission's approval or non-approval decision, which in turn determines societal outcomes in terms of cancer incidence and agricultural productivity. The details of this process will be presented in section 4.

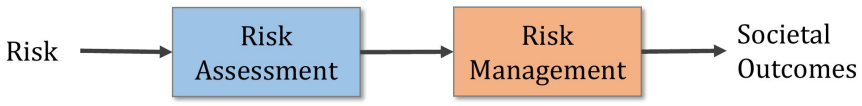

Figure 1: The basic risk regulatory framework, consisting of risk assessment (RA) and risk management (RM).

The concrete design of these two regulatory task and, most importantly, their relationship have been a field of intense controversy. On the one hand, there are proponents of a strong conceptual separation of RA and RM, often invoking scientific principles for RA to be unmarred from political influence.[2] The color coding in Figure 1 (and subsequent figures in this paper) depicts this conceptual separation. The political realm is represented in orange, the scientific realm in blue. The influential "Silver Book" by the National Research Council's Committee on Improving Risk Analysis Approaches

[1] Besides risk assessment and risk management, risk communication is often named as the third task of risk regulation. The challenges of risk communication however are not the focus of the present study which is why we will not further discuss it.

[2] This conceptual separation does not necessarily imply an institutional separation as well. While the EU process institutionally separates both tasks between EFSA and Commission, the U.S. EPA houses both tasks, conceptually separated in the sense that "assessors do not set standards and decision-makers do not conduct risk assessments" (NRC 2009).

Used by the U.S. EPA (NRC 2009), successor of the "Red Book" (NRC 1983), states: "It is imperative that risk assessments used to evaluate risk-management options not be inappropriately influenced by the preferences of risk managers." This call for a strong conceptual separation has been affirmed by a literature turning against biased RAs. In these authors' view, the politically prevalent precautionary stance on regulation has stained the RA process in the form of overly conservative estimates (Nichols and Zeckhauser 1986), with the conservative estimates often cascading in multiple RA steps (Viscusi et al. 1997). As Charnley and Rogers (2011) put it: "Scientists and regulated parties also criticize agency risk assessors for confusing risk assessment and risk management. By making conservative assumptions and precautionary choices as part of the risk assessment process, risk assessors are de facto acting within the realm of risk management."

On the other hand, there are authors challenging a strong separation of RA and RM, possibly best summarized by "separation should not mean divorce" (Wilson and Clark 1991). The main concern raised is that RA needs to be *useful* for RM and should not be a scientific end in itself: "risk assessment exists at the pleasure of risk management" (Charnley and Rogers 2011). Widely mentioned is that the focus on 'scientific principles' of RA is not in accordance with the regulatory task that often requires timely action and hence may have substantially adverse effects.[3] The idea that not abstract scientific principles, but improvements in regulatory outcomes ought to guide the decisions on the precision of RAs is most clearly expressed in

> "Given the imperfect state of the risk assessment art, regulators must decide how much potential but uncertain public protection should be traded for some potential but uncertain improvement in the accuracy of scientific judgments that [...] are far from reliable. The present guidelines assume that every tentative step, however provisional, in the direction of 'good science' is warranted regardless of its possible effect on the scope of protection. The wisdom of this presumption is surely a public policy issue rather than a purely scientific question" (Latin 1988).

This idea of balancing costs and benefits of additional information underpins the value-of-information approach that will be the focus of section 3.

The recent contribution by Heyen et al. (2015) adds a complementary dimension to the appropriate relation of RA and RM. Formulated within a value-of-information setup

[3]Latin (1988) writes: "There is an inherent tension between the disciplinary norms of good science and good regulation. Unlike in pure scientific research, where the proper response to uncertainty is reservation of judgment pending the development of adequate data and testable hypotheses, the risk-assessment process cannot be suspended without significant social consequences [..] Thus, scientists in regulatory proceedings are expected to produce 'answers' in a timely manner even if their predictions are highly speculative. Any reluctance to relax the standards of proof and certainty generally required of valid science may introduce a bias in favor of regulatory inaction." whereas Wagner (1995) isolates various reasons why the "science charade" may be detrimental.

it explicitly acknowledges RA's role in providing useful (and costly) information to risk managers; on the other hand, it demonstrates that if the decision on RA's *precision* – not the assessment itself as in Nichols and Zeckhauser (1986) and Viscusi et al. (1997) – follows a conservative maxmin stance that is often associated with the precautionary principle, this may result in complex and internally inconsistent dampening effect on the precision choice.[4] In this sense, Heyen et al. (2015) does suggests a separation between RM and RA's precision choice and hence an orthogonal dimension to the traditional separation between risk management and risk assessment.

The next section will present a simple decision-theoretic model that helps to appreciate the VoI approach and to isolate basic economic principles of deciding on RA precision.

3 A value-of-information model of precision choice

The following model presents the economic and decision-theoretic framework for determining the precision of RAs based on the value-of-information (VoI). The framework has no new or uncommon features and essentially corresponds with Olson (1990), Gabbert and Weikard (2013) or Heyen et al. (2015). The timeline, a refinement of Figure 1, is depicted in Figure 2.

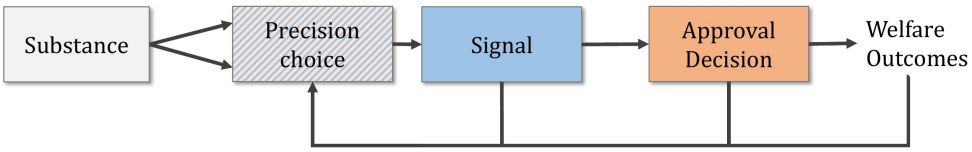

Figure 2: The value-of-information framework.

The first stage comprises all decision-relevant features the regulator has no influence on, often referred to as 'states of the world'. The figure depicts the simplified case of two possible states, e.g. glyphosate being carcinogenic or not. What makes the decision problem non-trivial is that the regulator is uncertain about the true state. The VoI models typically employ the Bayesian learning method; accordingly, the decision-maker initially holds a prior belief, i.e. a probability distribution over the set of possible states of the world.

The second stage is the choice of the precision of a noisy signal that materializes in the third stage. The signal can be thought of as conclusions from (possibly many) animal tests and hence represents a stylized and condensed RA output. In the example, animal test results provide evidence on glyphosate's carcinogenic potential. As is well

[4]The reasons for this *Research Pessimism Effect* is that the value-of-information may be assessed in a conservative, pessimistic way.

known, such a signal is typically noisy in the sense that it provides information without, however, fully resolving uncertainty. In other words, there are errors, often referred to as 'false positives' (an innocent substance is tested carcinogenic, aka type I error) and 'false negatives' (the carcinogenic potential of a substance is not detected, aka type II error) (Lave et al. 1988). Useful measures of the signal's precision are *sensitivity* and *specificity*, indicating the capability to avoid false negatives and false positives, respectively. The observed signal together with the initial prior and knowledge of sensitivity and specificity is then transformed to a *posterior* distribution, reflecting the belief about the substance's carcinogenic potential in light of the test results. The higher sensitivity and specificity, the higher the regulator's ex-post confidence, reflected in a sharper posterior distribution.

The fourth stage is the risk management decision, here thought of as the simplified binary decision of market approval vs. non-approval. The decision is made based on the posterior knowledge. What level of ex-post confidence is required – equivalently, which level of remaining uncertainty is deemed acceptable – is the central risk management decision. The decision on market approval or non-approval together with the true state, carcinogenic vs. non-carcinogenic, fully determines the social welfare outcomes.

The main focus of this model, in line with this paper's focus, is the choice of RA precision, here the precision of the signal. The economic principle underpinning this choice is to balance costs and benefits of a precision increase. The costs are typically assumed to be independent of the payoff structure and positively related to the precision. The benefits of a precision increase stem from improved RM quality and hence typically increase in the precision and, at the same time, depend on the decision problem. A crucial feature of the VoI model is that these benefits of signal precision are determined via *backward induction*. That is, first one determines the optimal decision *given* some posterior belief. The expected values resulting from this optimal behavior at the market approval decision form the basis of calculating the value of an increase in the signal precision. This method of backward induction, directly linking the final payoffs and decisions with the signal precision choice, clearly reflects the way RA is 'useful' for RM. A cornerstone of the VoI model is that the RA fully exists at the pleasure of RM and efficiently uses all available information.

Noteworthy is the ambivalent role of RA and RM in determining the optimal RA precision. While the market approval decision unanimously falls in the political realm, the research precision choice has aspects of both the scientific and the political realm, indicated in the two-color box in Figure 2. On the one hand, it has purely scientific aspects in the sense that the signal generated in the RA can be seen as value-free and politically not influenced; on the other hand, the political realm is present as both the societal outcomes and the anticipated RM decision profile directly impact the benefits of the precision choice. It is in this sense that the Silver Book states "[...] risk managers must see themselves as managing uncertainty and delay as well as managing risk. Managing under uncertainty requires diverse strategies that address different aspects of

the overall decision-making process, including investments to collect, store, and manage information; investments to improve the knowledge base, that is, to generate new knowledge [...]" (NRC 2009).

We conclude this section by summarizing a couple of characteristics that are explicit or implicit in the VoI approach. These characteristics will provide a useful benchmark for contrasting theory and practice in section 5. Starting at the end of the process, the **market approval decision** is *dependent on the testing outcome* and there exist transparently defined and commonly known *approval criteria* that stem from *balancing expected costs and benefits* of the regulatory action.[5] The output of the **risk assessment** is formulated such that its classification *formally fits with the risk managers' needs*, but at the same time the testing process, only depending on sensitivity and specificity, is independent of risk managers preferences and hence *unbiased*. The assessment is *informationally efficient* in the sense that it uses all information available. The **precision choice**, by VoI construction, *balances costs and benefits*, follows the *backward induction* principle and is *substance specific*. In particular, the precision choice depends on expected benefits and damages of market approval or non-approval as well as the approval criteria. The level of *initial knowledge* is exogenous and has substantial implications for the precision choice and market approval decision.

4 The reality of pesticide risk assessment and management in the EU

This section presents the regulatory framework for market approval of active substances in the EU. The restriction to the EU process allows us a detailed and concrete look into the regulatory practice. Many findings can however be expected to carry on to, at least, the structurally similar regulation in the U.S.

4.1 Substances, regulations and actors

The approval of pesticides in the EU is structured as a two tier approach. While the final plant protection products (PPP) are evaluated and authorized at the national level in the member states, active substances (e.g. glyphosate), which are the main determinant of the PPP's properties, are evaluated and authorized at EU level. The main regulation for active substances in the EU is regulation 1107/2009 (European Commission 2009). The renewal process of active substances differs slightly.[6]

There a couple of important actors in the approval process. First, the *applicant*, i.e. the producer of the substance. For glyphosate, the applicant is the "Glyphosate Task Force", comprising 23 companies including Dow AgroSciences LLC., Monsanto Europe

[5] This does not preclude a risk-averse or ambiguity-averse decision rule.

[6] The renewal process of glyphosate and 30 other active substances follows regulation 1141/2010 (European Commission 2010).

S.A./N.V. and Syngenta Limited.[7] The risk assessment mainly lies in the hands of the *European Food Safety Authority* (EFSA). The EFSA is supported by the member states, in particular the *Rapporteur Member State* (RMS). The RMS for the glyphosate application is Germany. The risk management decision is made by the *European Commission* (COM).

4.2 The approval process of active substances in the EU

This section presents the detailed timeline of the approval process of active substances in the EU. A graphical representation is given in Figure 3.

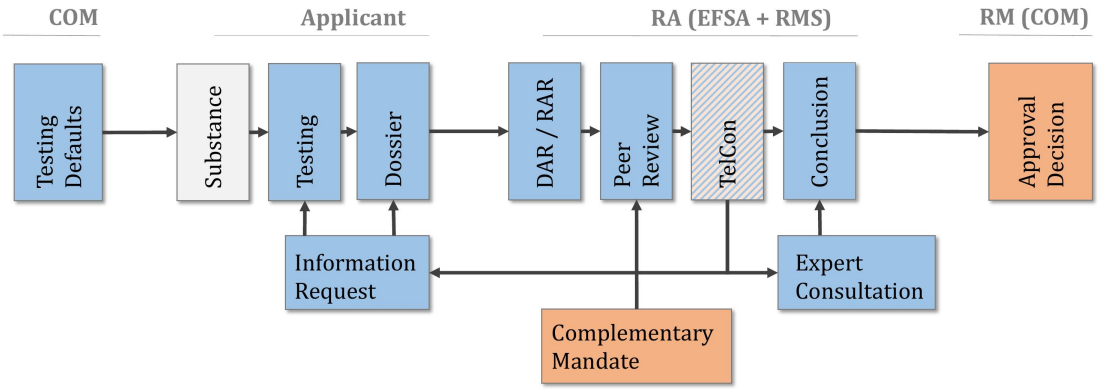

Figure 3: The approval process for active substances in the EU.

The first relevant step depicted in Figure 3 is the European Commission's "setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market" (European Commission 2013). These data requirements apply to all active substances and shape the testing procedures the applicant undertakes or commissions.[8] Noteworthy is that tests with humans and non-human primates shall not be performed (European Commission 2009, Annex 5.3). Studies, data and information ought to be provided, sufficient to evaluate the active substance's potential for, among other issues, acute toxicity, genotoxicity, long-term toxicity and carcinogenicity and reproductive toxicity. The specific requirements for "long-term toxicity and carcinogenicity" test include

> "A long-term oral toxicity study and a long-term carcinogenicity study (two years) of the active substance shall be conducted using rat as test species; where possible these studies shall be combined. A second carcinogenicity study of the active substance shall be conducted using mouse as

[7] http://www.glyphosatetaskforce.org/.
[8] It is common practice that applicants fund the required tests.

test species, unless it can be scientifically justified that this is not necessary" (European Commission 2013, Annex 5.5.).

The applicant assembles the testing results to a *dossier* that is submitted to a Rapporteur Member State of the applicant's choice. The RMS creates a *draft assessment report* (DAR) – *renewal assessment report* (RAR) in the case of a renewal application – assessing whether the active substance can be expected to meet the approval criteria.[9] Then the EFSA takes over, organizing a peer review process in which experts from all member states can comment on the RAR. After this process, a teleconference between EFSA, RMS and COM identifies the need for *requesting further information* from the applicant. Where appropriate, EFSA shall organise a *consultation of experts* to settle open questions. Completing the RA process, the EFSA then provides a *Conclusion on the peer review*. This conclusion also suggests a classification of the active substance according to the regulation on classification, labelling and packaging (CLP) of substances and mixtures (European Commission 2008).

The RA of glyphosate was special insofar as during the peer review process the International Agency for Cancer Research (IARC 2015) classed glyphosate as "probably carcinogenic to humans". This IARC report induced the European Commission to equip the EFSA with a complementary mandate to assess the IARC classification. Accordingly, the EFSA asked RMS Germany to provide an addendum to the RAR (Germany 2013). The results were discussed during a teleconference, including IARC experts, in September 2015. The EFSA conclusion sees the data "overall inconclusive for a causal or clear associative relationship between glyphosate and cancer in human studies". Accordingly, "Glyphosate is not classified or proposed to be classified as carcinogenic or toxic for the reproduction category 2 in accordance with the provisions of Regulation (EC) No. 1272/2008" (EFSA 2015a).

This classification suggestion is directly relevant for COM's RM decision whether or not to include the substance in the EU's list of approved active substances. Regulation 1107/2009 involves classificiation criteria under which an active substance shall not be included in this list. For instance,

> "An active substance [...] shall only be approved, if, on the basis of assessment of carcinogenicity testing carried out in accordance with the data requirements for the active substances [...] and other available data and information, including a review of the scientific literature, reviewed by the Authority, it is not or has not to be classified, in accordance with the provisions of Regulation (EC) No 1272/2008, as carcinogen category 1A or 1B [...]" (European Commission 2009).

In terms of carcinogenicity, substances can either be unclassified (as is the EFSA's suggestion for glyphosate) or be grouped as "known or presumed human carcinogens" (cat-

[9] In the following we will consistently use 'RAR'.

egory 1, with subcategories 1A and 1B) or "suspected human carcinogens" (category 2). The detailed definitions of the classification is found in the appendix.

Why do IARC and EFSA come to different assessments? The present paper does not aim at judging the relative merits of the IARC report and the EFSA RA, but it is valuable to note differences in the processes as they pertain to basic principles of RA. Three differences are apparent. First, IARC used formulations containing glyphosate while the EU process focused on the active substance in isolation. Co-formulants are one reason why studies incorporated by the IARC may suggest carcinogenicity. The second reason is the selection of studies. The IARC did not evaluate some studies the EFSA used (EFSA 2015b), while the EFSA rejected the validity of studies used by the IARC. Thirdly, the two approaches differ in their statistical treatment of crucial studies. While the EFSA restricted to pairwise tests (treatment vs. control), the IARC employed trend tests; the latter potentially uncover significant effects, but the EFSA's critique is that "the planning of a study before the initiation of the experimentation itself as established in the respective protocol – that includes the statistical analysis – is a key element in assessing the quality of a study, therefore deviations from the statistical analysis used by the study authors should be limited and properly justified" (EFSA 2015b).

4.3 Precision requirements: defaults and choices

We now emphasize which steps in the approval process implicitly or explicitly influences the RA precision. We start with the explicit ones, i.e. those steps at which active information acquisition takes place. The first active information step is the teleconference in which the decision is on the additional information request and the initiation of an expert consultation is made. The other active information possibility is, as with glyphosate, a complementary mandate from COM. The latter clearly falls in the political realm, hence the orange color coding. The teleconference, made between EFSA, RMS and COM has features of both, indicated in the orange-blue pattern. In addition to teleconference and complementary mandate by COM, there is in principle a third active information possibility. The EFSA has, in principle, the option to improve the knowledge by funding studies herself.[10] But this option was not used during the glyphosate assessment and seems to be a rarely employed instrument.

More important for the precision level seem the testing defaults that shape the RA process. An important restriction on the testing process is that no active substance should ever be tested with primates (neither human nor non-human). This restriction has substantial implications for the reliability of the RA process in light of the poor predictive power of mice and rat studies. Another default setting is the number and exact type of studies to be undertaken for assessing the risks. For the long-term toxicity

[10] "Although EFSA does not fund research of its own, it will likely be able to access such funding as specific needs are demonstrated. It will also work closely with DG Research and will use its own funds to command short term studies as specific needs arise" (EFSA Homepage).

and carcinogenicity tests, these are the long-term oral toxicity study and a long-term carcinogenicity study (two years) with rats, as well as a second carcinogenicity study of the active substance with mouse as test species mentioned above.

The guidelines for the animal studies are general OECD guidelines, and hence not specific to the approval of active substances, let alone a certain substance. Recommendations for carcinogenicity tests are outlined in OECD (2012), referring to principles of "Good Laboratory Practice" (OECD 1998, 2007). The OECD guidelines present standard statistical approaches, i.e. hypothesis testing, as the preferred tool:

> "the objective [of the test] may be to test a hypothesis that one or more treated group is different from the concurrent group. [..] Much of the work in toxicology has been carried out based upon a traditional frequentist approach particularly around the concept of hypothesis testing. While recognizing that alternative viewpoints exist and this is a controversial area, most of the emphasis in this document will be on the traditional approaches" (OECD 2012).

This clear scientific angle of these guidelines is reflected in the blue color-coding of the leftmost box in Figure 3. Scientific principles would call for very large sample sizes to increase the power of the tests. The OECD guidelines do not fail to give recommendations for the sample size, taking into account "economic practicalities":

> "The OECD Test Guidelines indicate the appropriate sample sizes for each group. In the carcinogenicity study, the sample size is usually at least 50 animals of each sex at each dose level. This group size reflects a trade-off between the statistical power of the design and economic practicalities of the design. In practice, the carcinogenicity study has low power in the sense that treatment effects that might be considered biologically important cannot be detected routinely as statistically significant" (OECD 2012).

5 Congruency and Disparity

This section combines the findings of the previous two sections, contrasting VoI recommendations and the actual regulatory practice. We begin with the market approval decision and then make the way through to decisions that determine the RA precision. Table 1 summarizes our findings.

5.1 Market approval decision

A first conformity of framework and practice is that the risk management decision explicitly makes use of the RA conclusion. In other words, RA is relevant. In general, practice and theory agree on the importance of well-defined and transparent approval criteria.

In the VoI framework these criteria are determined in the backward induction process anyway; in the regulatory practice, they are formulated in Regulation 1107/2009. See the appendix for the full list of criteria.

The precautionary principle is integral part of Regulation 1107/2009.[11] The regulation states on the approval process that

> "On the basis of the review report, other factors legitimate to the matter under consideration and the precautionary principle where the conditions laid down in Article 7(1) of Regulation (EC) No 178/2002 are relevant, a Regulation shall be adopted in accordance with the regulatory procedure referred to in Article 79(3), providing that (a) an active substance is approved [...]; (b) an active substance is not approved; or (c) the conditions of the approval are amended" (European Commission 2009, Article 13(2)),

where Article 7(1) of Regulation 178/2002 reads

> "In specific circumstances where, following an assessment of available information, the possibility of harmful effects on health is identified but scientific uncertainty persists, provisional risk management measures necessary to ensure the high level of health protection chosen in the Community may be adopted, pending further scientific information for a more comprehensive risk assessment" (European Commission 2002, Article 7(1)).

The VoI framework does not preclude precautionary notions since it is flexibility with regard to the decision rule, in particular the utility function, which determines market approval. Specific cautionary notions like ambiguity aversion can also be combined with the VoI framework (e.g. Heyen et al. 2015).

A crucial difference between VoI framework and actual practice is that the latter does not balance (expected) costs and benefits of the RM decision. This is most apparent as COM's approval criteria are not formulated in terms of social benefits and damages of the specific substance; rather, the decision exclusively hinges on the (science based) classification criteria. This lack of transparent cost-benefit criteria is related to findings in the general regulation literature (Viscusi and Zeckhauser 2015; Cropper et al. 1992; Viscusi and Hamilton 1999).

5.2 Risk assessment

Coming to the RA process, a feature on which framework and practice are in accordance is that the RA classification ought to be immediately useful to the RM decision. This is

[11] "The provisions of this Regulation are underpinned by the precautionary principle in order to ensure that active substances or products placed on the market do not adversely affect human or animal health or the environment. In particular, Member States shall not be prevented from applying the precautionary principle where there is scientific uncertainty as to the risks with regard to human or animal health or the environment posed by the plant protection products to be authorised in their territory" (European Commission 2009).

Table 1: Comparison VoI framework (section 3) and actual regulatory practice (section 4).

	VoI framework	Regulatory practice	Conformity
Market approval			
Explicitly dependent on RA	+	+	Yes
Clearly defined approval criteria	+	+	Yes
Precautionary Principle	+/–	+	(Yes)
Balancing costs and benefits	+	–	No
Risk assessment			
Classification useful for RM	+	+	Yes
Informationally efficient	+	–	No
Hypothesis neutrality	+	–	No
Unbiased testing	+	?	?
Precision choice			
Balancing costs and benefits	+	+/–	(Yes)
Backward induction	+	–	No
Substance specific procedures	+	+/–	(Yes)
Initial knowledge unspecified	+	–	No

the case in the regulatory practice as the RM decision is linked to the substance's CLP classification (European Commission 2008), and this is the same scheme the RA process relies upon.

A feature of sharp contrast is the treatment of information. While the VoI framework rests on the Bayesian statistical approach in which updating of beliefs efficiently processes new information, the regulatory practice rests on the traditional statistical approach of hypothesis testing, which deliberately leaves aside information. The scientific rationale behind this is to impose asymmetrically high requirements for an effect to be considered 'significantly' different from the no-effect null hypothesis.

An immediate implication of this traditional statistical method is in conflict with the precautionary principle. The traditional statistical method compares the fate of a control group (not exposed to the substance) to the treatment group (exposed to the substance), setting as the null hypothesis the applicant-friendly statement "the substance is not carcinogenic". Carcinogenic potential needs to be clear enough to be able to reject this null hypothesis. "Overall in the current regulatory assessment, any toxic effect is rst suspected to be a false positive, arising by chance, rather than questioning whether no evidence of effect is a false negative result." (Mesnage et al. 2015). We can only speculate whether the evidence from the animal test studies of the glyphosate dossier might have been interpreted differently under a different null hypothesis or under a fundamentally different statistical method like the Bayesian approach.[12]

[12]This view is also expressed in: "And what about the [...] possibility [...] that a public health

In addition, there is another discrepancy between theory and practice in the treatment of uncertainty and information. Irrespective of the statistical method, the data generated itself ought to be unbiased. We have no specific evidence about the test procedures in the active substance approval process (why we put the question mark in the respective column), but ample evidence exists that studies often cannot be reproduced (McNutt 2014), for instance as a result of wishful thinking on behalf of the experimenters (Holman et al. 2016) or a conservative bias in reporting testing results (Nichols and Zeckhauser 1986; Latin 1988; Viscusi et al. 1997; Charnley and Rogers 2011).

5.3 Precision choices

The last subsection is devoted to the comparison of regulatory practice and framework recommendations regarding the RA precision choice. An aspect innate in the VoI approach is the balancing of research costs and benefits. As shown above, the OECD guidelines actually make explicit reference to the 'economic practicalities', precluding extensive testing procedures. It is in this sense that the actual regulatory practice balances, to some degree, costs and benefits.

Adopting the VoI perspective, this balancing of costs and benefits is, however, insufficient in many respects. The main problem is that the idea of backward induction is mostly missing. Recall that in the VoI framework the precision choice is a function of the possible welfare changes under different states and actions, the decision rule that governs market approval, and characteristics of the signal structure like sensitivity and specificity. This is not the case in the regulatory practice.[13]

Likewise, the default precision requirements are not substance specific. In particular, the OECD guidelines apply very general, exceeding even the realm of active substances. As a result, much of the testing procedure is undertaken without a specific look at the concrete needs of the regulator. The active information acquisition decisions within the approval process, both the complementary mandate and also the teleconference, have some backward induction and substance specific character. These active information decisions however are not explicitly based on the comparison of costs and benefits of information, and typically also do not involve commissioning new studies.

The initial knowledge, encoded in the initial prior, about the substance is a feature not specified in the VoI framework. Much criticism about the VoI method revolves around this initial prior and its substantial impact on the outcomes. The regulatory practice is different. Its take on the initial knowledge is characterized by an intriguing mix of statistical principles, the precautionary principle, and pragmatism. It is based on statistics principles insofar as the null hypotheses "no carcinogenic effect" is the point of departure. At the same time, the regulatory practice follows the precautionary principle

catastrophe is an effect so strong that even an epidemiological study can detect it?" (Fagin 2013).

[13]To be fair: It seems possible that the default OECD testing guidelines are based on experience with past testing experience, in particular with experience about false negatives and false positives.

and assigns the burden of proof for demonstrating the safety of the substance to the applicant. We have already pointed out the obvious conflict between null hypothesis and the precautionary principle above. Finally, the initial stance displays a good deal of pragmatism as the applicant is typically required to fund the research that enters the RA process, even though this has obvious potential for biases.

In summary, despite some level of conformity, we find that the precision requirements for the approval of active substances in the EU deviates from the VoI-based recommendations in many and important aspects. We thus have to expect that the substantial testing costs are not spent efficiently. This inefficiency may occur in the double sense of overspending on questions that need no further clarity and underspending on questions that are pivotal to the regulatory process.

6 Concluding discussion

New technologies and substances confront modern societies with substantial and novel risks. To evaluate nature and extent of those risks, regulatory procedures extensively rely on risk assessment (RA). Over the past decades various debates about the appropriate RA design emerged, in particular about the relationship between RA and risk management (RM). One recurring point of criticism holds that the traditional scientific principles guiding RA constitute a "science charade" (Wagner 1995) that prevents RA to fulfill its central task, namely to provide immediately useful input for RM decisions. Underpinned by influential guidance reports by the National Research Council (NRC 1983, 2009), RA is nowadays moving towards a more utility oriented approach.

One important way to make RA more useful is appropriately setting the RA precision requirements. The inherent imperfection of testing procedures imply that the level of remaining uncertainty present in the final regulatory market approval decision is considerable. Together with the significant costs of testing, this infeasibility of perfect information demonstrate the pivotal role of choosing the level of precision of the RA process. To assist this choice, welfare economic based frameworks emerged to determine the value-of-information (VoI).

This paper has put the RA precision choice center-stage and contrasted VoI recommendations with the actual regulatory practice in a concrete example, the approval of active substances (the effective component in plant protection products) in the EU. The paper has uncovered important deviations between theory and practice, particularly in terms of the application of the backward induction principle to make the precision choice responsive to characteristics of the RM decision and testing process. Rather, scientific principles encoded in OECD guidelines, applicable to a wide array of regulatory problems and hence not substance or process specific, dominate the current precision requirements in the RA practice.

In addition to being a valuable result in its own right, this comparison can also assist the restructuring process of RA that is currently underway. In light of the endorsement of VoI frameworks in the Silver Book, it can be suspected that RA will move into this direction; actually, VoI approaches are already used in medical decision making (Steuten et al. 2013). The present paper's contribution in assisting the RA restructuring is related to the observation that precautionary notions explicitly and implicitly shape large parts of the risk regulation. We have seen above that the European Commission – in the approval of active substances and elsewhere – explicitly refers to the precautionary principle. We have also observed that the European Food Safety Authority's role in influencing the RA precision is quite limited, in contrast to the European Commission that can commission a complementary mandate and directly influence the testing defaults. Combining these findings, it may hence be expected that risk managers, not risk assessors, and the precautionary notions present among risk managers will shape the RA precision requirements if the path to VoI is further taken. This would be in conflict with the results in Heyen et al. (2015) who suggest, based the finding that precautionary mandate can reduce the value-of-information and hence be in conflict with ideas of 'precautionary learning', a separation between risk management preferences and choices on information precision. The interplay of precautionary mandates and informational actions is a fascinating young research field that can favorably stimulate the debate about the adequate RA design.

The present study can be fruitfully extended in two other respects. First, regarding the theoretical background, we have deliberately kept the VoI framework quite general and were hence not able to discuss in detail more sophisticated RA methods like tiered testing approaches (Yokota et al. 2004; Gabbert and Weikard 2013) and impediments to implementing VoI recommendations in practice (NRC 2009; Yokota and Thompson 2004). The other way in which the paper can find a valuable extension is to broaden the current EU Focus. A comparison of U.S. and EU would be particularly interesting in light of the ongoing debate about relative precaution on both sides of the Atlantic (Wiener et al. 2011).

References

Abbott, A. (2005). Animal testing: More than a cosmetic change, *Nature* **438**(7065): 144–146.

Aktar, W., Sengupta, D. and Chowdhury, A. (2009). Impact of pesticides use in agriculture: their benefits and hazards, *Interdisciplinary Toxicology* **2**(1): 1–12.

Baylis, A. D. (2000). Why glyphosate is a global herbicide: strengths, weaknesses and prospects, *Pest Management Science* **56**(4): 299–308.

Bottini, A. A. and Hartung, T. (2009). Food for Thought on the Economics of Animal Testing, *Altex* **26**(1): 3–16.

Charnley, G. and Rogers, M. D. (2011). Frameworks for risk assessment, uncertainty, and precaution, *The Reality of Precaution*, RFF Press, pp. 361–75.

Cropper, M. L., Evans, W. N., Berardi, S. J., Ducla-Soares, M. M. and Portney, P. R. (1992). The Determinants of Pesticide Regulation: A Statistical Analysis of EPA Decision Making, *Journal of Political Economy* **100**(1): 175–197.

Duke, S. O. and Powles, S. B. (2008). Glyphosate: a once-in-a-century herbicide, *Pest Management Science* **64**(4): 319–325.

EFSA (2015a). Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate.

EFSA (2015b). EFSA explains the carcinogenicity assessment of glyphosate.

European Commission (2002). Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food SafetyAuthorityand laying down procedures in matters of food safety.

European Commission (2008). Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures.

European Commission (2009). Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.

European Commission (2010). Commission Regulation (EU) No 1141/2010 of 7 December 2010 laying down the procedure for the renewal of the inclusion of a second group of active substances in Annex I to Council Directive 91/414/EEC and establishing the list of those substancesText with EEA relevance.

European Commission (2013). Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the marketText with EEA relevance.

Fagin, D. (2013). *Toms River: a story of science and salvation*, Bantam.

Gabbert, S. and Weikard, H.-P. (2013). Sequential Testing of Chemicals when Costs Matter: A Value of Information Approach, *Human and Ecological Risk Assessment: An International Journal*.

Germany (2013). Renewal assessment report (R AR) on the active substance glyphosate prepared by the rapporteur Member State Germany in the framework of Regulation (EU) No 1141/2010.

Haimes, Y. Y. (2005). *Risk modeling, assessment, and management*, Vol. 40, John Wiley & Sons.

Hartung, T. (2008). Food for thought on animal tests, *Altex* **25**(1): 3–9.

Heyen, D., Goeschl, T. and Wiesenfarth, B. R. (2015). Risk Assessment under Ambiguity: Precautionary Learning vs. Research Pessimism, *AWI Discussion Paper 605*.

Holman, C., Piper, S. K., Grittner, U., Diamantaras, A. A., Kimmelman, J., Siegerink, B. and Dirnagl, U. (2016). Where Have All the Rodents Gone? The Effects of Attrition in Experimental Research on Cancer and Stroke, *PLoS Biol* **14**(1): e1002331.

IARC (2015). Some organophosphate insecticides and herbicides: tetrachlorvinphos, parathion, malathion, diazinon and glyphosate, *Technical Report IARC Monograph 112*.

Latin, H. (1988). Good science, bad regulation, and toxic risk assessment, *Yale J. on Reg.* **5**: 89.

Lave, L. B., Ennever, F. K., Rosenkranz, H. S. and Omenn, G. S. (1988). Information value of the rodent bioassay, *Nature* **336**(6200): 631–633.

McNutt, M. (2014). Reproducibility, *Science* **343**(6168): 229–229.

Mesnage, R., Defarge, N., Spiroux de Vendmois, J. and Sralini, G. E. (2015). Potential toxic effects of glyphosate and its commercial formulations below regulatory limits, *Food and Chemical Toxicology* **84**: 133–153.

Nichols, A. L. and Zeckhauser, R. J. (1986). Perils of Prudence: How Conservative Risk Assessments Distort Regulation, The, *Regulation* **10**: 13.

NRC (1983). *Risk assessment in the federal government: Managing the process*, National Academy Press.

NRC (2009). *Science and decisions: advancing risk assessment*, National Academies Press, Washington, D.C.

OECD (1998). Principles on Good Laboratory Practice, *Series on Principles of Good Laboratory Practice and Compliance Monitoring*.

OECD (2007). Establishment and control of archives that operate in compliance with the principles of GLP, *Series on Principles of Good Laboratory Practice and Compliance Monitoring*.

OECD (2012). Guidance Document 116 on the conduct and design of chronic toxicity and carcinogenicity studies.

Olson, L. J. (1990). The search for a safe environment: The economics of screening and regulating environmental hazards, *Journal of Environmental Economics and Management* **19**(1): 1–18.

Renn, O. (2006). risk governance – towards an integrative approach.

Steuten, L., Wetering, G. v. d., Groothuis-Oudshoorn, K. and Retl, V. (2013). A Systematic and Critical Review of the Evolving Methods and Applications of Value of Information in Academia and Practice, *PharmacoEconomics* **31**(1): 25–48.

Viscusi, W. K. and Hamilton, J. T. (1999). Are Risk Regulators Rational? Evidence from Hazardous Waste Cleanup Decisions, *The American Economic Review* **89**(4): 1010–1027.

Viscusi, W. K. and Zeckhauser, R. J. (2015). Regulating Ambiguous Risks: The Less than Rational Regulation of Pharmaceuticals, *SSRN Scholarly Paper ID 2392070*, Social Science Research Network, Rochester, NY.

Viscusi, W. K., Hamilton, J. T. and Dockins, P. C. (1997). Conservative versus Mean Risk Assessments: Implications for Superfund Policies, *Journal of Environmental Economics and Management* **34**(3): 187–206.

Wagner, W. E. (1995). The Science Charade in Toxic Risk Regulation, *Columbia Law Review* **95**(7): 1613–1723.

Wiener, J. B., Hammit, J., Rogers, M. and Sand, P. (2011). *The reality of precaution: Comparing risk regulation in the United States and Europe*, RFF Press.

Wilson, R. and Clark, W. (1991). Risk Assessment and Risk Management: Their Separation Should Not Mean Divorce, *in* C. Zervos, K. Knox, L. Abramson and R. Coppock (eds), *Risk Analysis*, number 8 in *Advances in Risk Analysis*, Springer US, pp. 187–196. DOI: 10.1007/978-1-4899-0730-1_20.

Yokota, F. and Thompson, K. M. (2004). Value of Information Analysis in Environmental Health Risk Management Decisions: Past, Present, and Future, *Risk Analysis* **24**(3): 635–650.

Yokota, F., Gray, G., Hammitt, J. K. and Thompson, K. M. (2004). Tiered Chemical Testing: A Value of Information Approach, *Risk Analysis* **24**(6): 1625–1639.

A Classification and Approval Criteria

This appendix presents classification and approval criteria for active substances in the EU. Figure 4 presents the classification criteria for carcinogenicity from Regulation 1272/2008 (European Commission 2008). The definition of the terms "sufficient evidence" and "limited evidence" from animal testing are given in Figure 5.

The list of approval criteria for active substances from Regulation 1107/2009 (European Commission 2009) is presented in Figure 6.

Table 3.6.1

Hazard categories for carcinogens

Categories	Criteria
CATEGORY 1:	Known or presumed human carcinogens A substance is classified in Category 1 for carcinogenicity on the basis of epidemiological and/or animal data. A substance may be further distinguished as:
Category 1A:	Category 1A, known to have carcinogenic potential for humans, classification is largely based on human evidence, or
Category 1B:	Category 1B, presumed to have carcinogenic potential for humans, classification is largely based on animal evidence. The classification in Category 1A and 1B is based on strength of evidence together with additional considerations (see section 3.6.2.2). Such evidence may be derived from: — human studies that establish a causal relationship between human exposure to a substance and the development of cancer (known human carcinogen); or — animal experiments for which there is sufficient (1) evidence to demonstrate animal carcinogenicity (presumed human carcinogen). In addition, on a case-by-case basis, scientific judgement may warrant a decision of presumed human carcinogenicity derived from studies showing limited evidence of carcinogenicity in humans together with limited evidence of carcinogenicity in experimental animals.
CATEGORY 2:	Suspected human carcinogens The placing of a substance in Category 2 is done on the basis of evidence obtained from human and/or animal studies, but which is not sufficiently convincing to place the substance in Category 1A or 1B, based on strength of evidence together with additional considerations (see section 3.6.2.2). Such evidence may be derived either from limited (1) evidence of carcinogenicity in human studies or from limited evidence of carcinogenicity in animal studies.

(1) Note: See 3.6.2.2.4.

Figure 4: Carcinogenicity classification criteria from Regulation 1272/2008 (European Commission 2008).

(b) Carcinogenicity in experimental animals

Carcinogenicity in experimental animals can be evaluated using conventional bioassays, bioassays that employ genetically modified animals, and other in-vivo bioassays that focus on one or more of the critical stages of carcinogenesis. In the absence of data from conventional long-term bioassays or from assays with neoplasia as the end-point, consistently positive results in several models that address several stages in the multistage process of carcinogenesis should be considered in evaluating the degree of evidence of carcinogenicity in experimental animals. The evidence relevant to carcinogenicity in experimental animals is classified into one of the following categories:

— sufficient evidence of carcinogenicity: a causal relationship has been established between the agent and an increased incidence of malignant neoplasms or of an appropriate combination of benign and malignant neoplasms in (a) two or more species of animals or (b) two or more independent studies in one species carried out at different times or in different laboratories or under different protocols. An increased incidence of tumours in both sexes of a single species in a well-conducted study, ideally conducted under Good Laboratory Practices, can also provide sufficient evidence. A single study in one species and sex might be considered to provide sufficient evidence of carcinogenicity when malignant neoplasms occur to an unusual degree with regard to incidence, site, type of tumour or age at onset, or when there are strong findings of tumours at multiple sites;

— limited evidence of carcinogenicity: the data suggest a carcinogenic effect but are limited for making a definitive evaluation because, e.g. (a) the evidence of carcinogenicity is restricted to a single experiment; (b) there are unresolved questions regarding the adequacy of the design, conduct or interpretation of the studies; (c) the agent increases the incidence only of benign neoplasms or lesions of uncertain neoplastic potential; or (d) the evidence of carcinogenicity is restricted to studies that demonstrate only promoting activity in a narrow range of tissues or organs.

Figure 5: Definition of "sufficient evidence" and "limited evidence" from animal testing.

in Annex II, 1107/2009	in Annex I, 1272/2008	Issue	Knock-out criteria
Impact on human health			
3.6.2	3.5.	Mutagenicity	class. 1A or 1B
3.6.3	3.6.	Carcinogenicity	class. 1A or 1B
3.6.4	3.7.	Reproductive Toxicity	class. 1A or 1B
3.6.5		Endocrine disrupting	If classified
Fate and behavior in the environment			
3.7.1		persistant organic pollutant (POP) - if 3.7.1.1 - 3 all fulfilled	If classified
3.7.2		persistent, bioaccumulative and toxic (PBT)	If classified
3.7.3		very persistent and very bioaccumulative substance (vPvB)	If classified
Ecotoxicology			
3.8.2		endocrine discrupting on non-target organisms	If classified
3.8.3		honeybee	
3.10.		groundwater	

Figure 6: List of approval knockout criteria from Regulation 1107/2009 (European Commission 2009).