PAPER IN PHILOSOPHY OF SCIENCE IN PRACTICE

# Evidence based methodology: a naturalistic analysis of epistemic policies in regulatory science

José Luis Luján [1] (iD) · Oliver Todt [1]

## Abstract

In this paper we argue for a naturalistic solution to some of the methodological controversies in regulatory science, on the basis of two case studies: toxicology (risk assessment) and health claim regulation (benefit assessment). We analyze the debates related to the scientific evidence that is considered necessary for regulatory decision making in each of those two fields, with a particular attention to the interactions between scientific and regulatory aspects. This analysis allows us to identify two general stances in the debate: a) one that argues for more permissive standards of evidence and for methodological pluralism, and b) an opposing one that not only defends strict evidence requirements but also stipulates the use of one particular (or at most a few) scientific methodologies for data generation. We argue that the real-world outcomes produced by alternative regulatory options are a vital piece of information that allows for the empirical assessment of these two stances. In particular, this information on outcomes makes it possible to analyze which standards of evidence and scientific methods generate the most useful knowledge as input for regulatory decision making. Our conclusion is that instead of an a priori selection of methodologies and standards, such decisions ought to be based on empirical evidence related to real-world outcomes.

✉ José Luis Luján
   jl.lujan@uib.es

   Oliver Todt
   oliver.todt@uib.es

1  Department of Philosophy, University of the Balearic Islands, Carretera de Valldemossa Km. 7,5, Palma 07071, Spain

## 1 Introduction

In this paper we argue for a naturalistic solution to some of the methodological controversies in regulatory science, on the basis of two case studies: toxicology (risk assessment) and health claim regulation (benefit assessment). Our main point is that the real-world outcomes produced by alternative regulatory options are a relevant piece of information that allows for the empirical assessment of the methodologies relevant to those same options. The information on outcomes makes it possible to analyze which standards of evidence and scientific methods generate the most useful knowledge as input for regulatory decision making. Our naturalistic conclusion is that instead of an a priori selection of methodologies and standards, such decisions ought to be based on empirical evidence related to real-world outcomes.

Our proposal flows from debates on naturalism in philosophy of science. Naturalism is based on the general idea that there do not exist fundamental differences between philosophical problems and scientific problems, including any philosophical problems related to scientific knowledge. Both types of problems present a conceptual as well as an empirical strand, in varying proportions. A naturalistic analysis implies that information about science generated by scientific research in fields like sociology, history, and psychology becomes relevant for philosophical inquiry.

The different naturalistic proposals in philosophy of science can be categorized according to a) the basis of naturalization,[1] and b) normativity (Giere 1998). The former is related to the scientific disciplines that furnish evidence for philosophical analysis. The latter issue, normativity, touches on the question if a naturalistic philosophy can possess a normative element related to scientific activity.

Giere (1985) argues against normativity in philosophy of science. There are, however, other philosophers of science who have made proposals for a normative naturalism. Laudan (1987, 1990) suggests history of science as a starting point, Kitcher (1993) the cognitive sciences, and Solomon (2001), Fuller (2000) and Longino (1990) social as well as gender studies of science. All of those proposals are clearly normative in nature, even if they generally imply a redefinition of the meaning of normativity (Laudan 1990; Mayo and Miller 2008).

Most of the philosophical analyses related to regulatory science present normative aspects. Their objective frequently is to a) identify the consequences for public health and the environment of alternative scientific or regulatory options, and b) criticize the dominant option and/or argue for any of the alternatives on the basis of ethical, scientific, epistemic, as well as political stances. As an example, Shrader-Frechette (1991, 1994) and Cranor (1993) recur to Rawl's and Scanlon's theses in order to analyze the social distribution of risks. Several of the authors who study regulatory science have taken a stance on naturalism. Shrader-Frechette applies Laudan's historicist naturalism to regulatory science, by incorporating non-epistemic values into Laudan's reticular model of justification (Shrader-Frechette 1989). Mayo and Miller (2008) argue for a general naturalistic approach to regulatory science in which scientific methodologies have to be empirically evaluated as to their ability to allow for valid

---

[1] For example, cognitive psychology, history of science, biology, etc. Different naturalisms use different scientific disciplines as a basis.

inferences on the basis of limited data. Cranor (1995) proposes an empirical assessment of methodologies for identifying risks.

Our proposal in this paper is a naturalistic, normative, as well as consequentialist one. From Giere's (1998) criteria, the basis of naturalization in our proposal is empirical research about the social consequences of different, alternative regulatory options. In other words, in the context of this paper we understand naturalism as the analysis of empirical information about regulatory outcomes, as well as its obtainment, which can then be used to resolve methodological controversies in regulatory science. We argue that in regulatory science it is not possible to determine which is the best or most adequate epistemic policy without empirical information about the real-world outcomes of regulation. Our proposal is normative in the sense that it considers empirical information about social consequences of alternative regulatory options as a fundamental criterion in selecting those scientific methodologies which best contribute to the fulfillment of a regulation's objectives. And the proposal is consequentialist simply because it recurs to empirical information about a regulation's consequences.

For our analysis we will recur to the concept of epistemic policies. The latter are sets of scientific methods, standards and definitions (for instance, of causality) relevant for data generation and decision making in regulatory science. Alternative regulatory options often recur to diverging epistemic policies. In order to assess such epistemic policies, purely epistemic considerations are not sufficient. Rather, it is necessary to take into account the real-world outcomes of regulation. We argue that the study of these outcomes ought to underlie the decision as to which epistemic policies are considered the most adequate in the context of a particular regulation.

In this paper we present one case study each of regulatory science related to risk assessment (toxicology), as well as benefit assessment (health claim regulation). Analyzing cases from both of those fields allows for more robust conclusions, because of the differences in their objectives, either protecting from risk, or determining the benefits derived from a product. In both cases we show that the controversies about standards and methods can be conceptualized as arguments about the relative advantages and disadvantages of alternative epistemic policies. We argue that in order to resolve these arguments it is necessary to recur to empirical information about the outcomes of different regulatory options. As we will see, our proposal implies analyzing the interactions between epistemic and pragmatic values in epistemic policies.

## 2 Regulatory controversy and epistemic policies

The regulation of scientific and technological products, applications and processes has been generating public debate for decades, in fields ranging from chemical products to biotechnology, pesticides and food complements. These controversies center on the aims of regulation, as well as the means necessary for achieving these aims. In other words, they are highly relevant to regulatory science.

Central to these debates are the standards of evidence, as well as the question of methodological monism vs methodological pluralism. The debate on methodological monism turns on the issue of evidentiary hierarchies (Osimani 2014, Osimani 2020, Stegenga 2014, Cartwright and Hardie 2013; Luján and Todt 2020). This is because in risk assessment it is often impossible to generate data from randomized controlled trials

(RCTs, clinical trials). This is relevant because RCTs tend to occupy the highest echelon in most hierarchies of evidence. In risk assessment, RCTs can usually not be used due to ethical reasons, because they would imply the need for exposing individuals to harmful substances and products. To the contrary, RCTs can be applied to benefit assessment. In pharmaceutical testing, RCTs are the standard methodology. When it comes to applying RCTs to the social sciences or to public policy, however, many scientists consider this methodology far from appropriate (Cartwright and Hardie 2013).

Two important arguments in favor of methodological monism are: 1) the establishment of evidentiary hierarchies in such a way that there exist clear-cut characteristics which allow to consider, for example, certain types of observational studies superior to others, and 2) the application of causal analyses in order to determine in which ways changes to the inputs propagate throughout a system, and cause changes to the output (Cox 2013).[2]

As a general rule, those who argue for methodological monism and evidentiary hierarchies are mostly concerned about false positives. That is because they consider accuracy to be the most relevant epistemic value in regulatory science, as would be the case in academic science. There are, however, other proposals that defend monism on the basis of the regulatory outcomes, instead of the primacy of epistemic values. One such proposal is the one by Andreoletti and Teira (2019). The authors argue that decisions in pharmaceuticals testing based on RCT data are preferable to those based on pluralistic methodological options, due to the social consequences of regulation.

As far as methodological pluralism is concerned, there is the proposal by Cartwright and Stegenga (2011) to recur to a weight-of-evidence approach in order to assess evidence in the formulation of public policy.

The authors defend the following three principles: 1) Affirmations regarding the effectiveness of public policy should be conceptualized as counterfactuals; this implies the need for a causal model that identifies any causal factors which operate through the intervention and combined effects of all those causal factors; 2) There is a need for taking into account not only the diverse causal complexes which produce the same effect, but also the different components of each of those causal complexes[3]; 3) There is a need for taking into account any auxiliary factors which are necessary for the policy intervention to produce the desired outcome.

A different pluralistic approach is the one by Vandenbroucke et al. (2016). They criticize the restricted potential outcomes approach (RPOA) in epidemiology, as well as any type of methodological monism that uses as exclusive input evidence from RCTs. They argue against monism because not only does it restrict the evidence considered acceptable, but also the type of questions which epidemiologists are allowed to ask. The first point is related to the uses of evidence, while the second concerns the generation of evidence.

Against monism, the authors argue for triangularization (see Heesen, Bright and Zucker 2019), which is none other than a weight-of-evidence approach. The authors'

---

[2] Some of those methods for causal analysis are: quasi-experimental design, causal network models, conditional independence tests, and marginal structural models.
[3] Causality is understood here as INUS (insufficient but necessary part of an unnecessary but sufficient) conditions.

main point is that there is no single, unique conceptualization of causality. Methodological monism gives preference to one particular conceptualization of causality, while discarding any unrelated evidence, as well as the methods used for generating this evidence.

Another author to defend a pluralistic approach in relation to evidence and causality is Haack (2014). She defends what she calls a weight of combined evidence approach. The basic idea is that in order to fundament any conclusion, a combination of several lines of evidence is considered more effective than each of those lines by itself. Her proposal is aimed at the use of evidence in court, particularly in cases related to undesired side-effects of pharmaceuticals, as well as environmental pollution.

Osimani (2014), in the case of pharmaceuticals, argues for a pluralistic and precautionary point of view, which implies a lowering of the standards of proof. The idea is that the evidentiary requirements have to be different for determining the safety (risk) of pharmaceuticals, as compared to establishing their efficacy. Landes, Osimani, and Poellinger (2018), on the basis of Hill's (1965) criteria, argue for evidence-amalmagation in pharmaceuticals testing.

Beyond the general debate on monism and pluralism, another crucial element of epistemic policies are the standards of proof (Cranor 1995; Douglas 2000; Douglas 2009; Elliott, 2011; Steel 2015; Reiss 2015). These determine the required type and level of evidence, and indicate under which circumstances we can consider that an affirmation of knowledge has been proven. The standards of proof often establish hierarchies among different types of evidence. These hierarchies can be used in risk assessment to show the existence of causal relationships between exposure to a chemical substance and the onset of a particular health problem. In benefit assessment they allow to analyze the relationship between consumption of a food and certain beneficial health effects.[4] In regulatory science such standards of proof are directly relevant to decision making (Douglas 2000; Steel 2015; Luján and Todt, 2015).

Standards of proof in regulatory science imply epistemic, as well as pragmatic aspects, the latter being the societal consequences of regulation. In a restricted sense, the concept of standards of proof makes reference to the type and level of required evidence. This can be used to establish, for instance, the causal relationship between a substance and a disease. In a more general sense, the standards of proof affect the entire process of knowledge generation. That is because in science there are a lot of decisions that are directly related to the standards of proof, like choosing methods for data generation, establishing the level of statistical significance, or the criteria for accepting or rejecting data (Douglas 2000). Any of these decisions implies an increase or decrease in one of the two fundamental statistical errors: false positives and false negatives.

In other words, we can understand the concept of epistemic policies as making reference –apart from other elements like the definition of causality– to the standards of proof (in the restricted sense), plus all the methodological decisions that are related to such standards. Our case studies show how epistemic and pragmatic aspects interact in regulatory science. As we will see, controversies related to regulation can affect the

---

[4] The hierarchies of evidence established on the basis of standards of proof may refer to the scientific evidence itself (data), but also to the methods used for obtaining this evidence (Illari and Russo 2014). Often they are hierarchies of methodologies, and only indirectly of the generated evidence.

standards of proof, which then act upon the selection of scientific methodology, as well as entire epistemic policies.

We argue that it is not only impossible to try to dodge the influence of pragmatic (non-epistemic) aspects in regulatory science, but that doing so might even be pernicious. That is because the real-world outcomes of the alternative regulatory options can help us in assessing those options, and deciding between them.

## 3 Risk assessment

Risk assessment is the obtainment and analysis of scientific data used in the regulation of technological processes and products that may have negative effects on the environment or human health. There have been several important controversies in the last few decades related to technological and scientific risks (Shrader-Frechette 1991; Sunstein 2002; Elliott, 2011; Luján and Todt, 2015; Cranor 2017).

Despite their variety and evolution in time, we can identify two very general stances in many of those controversies. These stances are related to the aims of regulation and to the evidence requirements. The first stance argues for the application of the strictest evidence requirements available or possible with the aim of avoiding arbitrary, excessive or unnecessary regulation. Here the idea is to avoid imposing regulation as a consequence merely of political pressure or economic interests, so as to minimize harm to innovation, as well as the costs that regulation implies for corporations and consumers (Sunstein 2002; Cox 2015). The second stance, to the contrary, defends a reduction in the requirements necessary for establishing that a substance, process or product is considered to entail risks. This would make it easier to authorize regulations that better protect health and the environment (Cranor 1993; Douglas 2000; Elliott, 2011).

The controversies about the aims of regulation and the standards of proof have an important methodological aspect. This methodological debate is in many cases related to the kind of data required in order to be able to proceed with regulation of a substance: are data in humans an absolute requirement? Or is it sufficient to obtain data from animal assays, or even just from in vitro or mechanistic studies? Very strict standards of proof, for instance, would exclude from regulatory consideration any non-human data.

Another very relevant methodological debate concerns the extrapolation models and the rules of inference. The latter are needed for being able to extrapolate from exposure to high doses of a substance (data typically available from animal assays) to exposure to low doses (which in many cases are the doses to which humans would be exposed while in contact with the substance in question, but for which it is very difficult or impossible to obtain data).

The characteristics of toxic substances, as well as their interaction with the human metabolism make it often very difficult to determine the causal relationships involved. But most research about toxicological risk depends on the establishment of such causal relations. The conjunction of both factors, evidentiary needs and behavior of toxic substances, turns most toxicology research into a very slow, as well as time and resource intensive affair. For risk assessment this means that in many instances there exists a (potentially pressing) conflict between, on the one hand, a cognitive value like accuracy or analytic rigor, and on the other, a pragmatic value like the protection of human health and the environment. The principal reason is the time needed (possibly

years) to come to any regulation-relevant conclusions (Cranor 2017). This conflict has given rise to various proposals for minimizing the pragmatic consequences of strict standards of evidence.

Several authors have proposed the use in regulation of short-term tests (STT) to resolve this issue (Cranor 1995). STTs are in vitro assays with biological systems (excluding animals) that can be completed in mere hours or days. These tests are particularly relevant for establishing genotoxicity and mutagenesis.

A similar proposal intended to speed up testing (even at the cost of lower accuracy) is to analyze the relationships between chemical structure and physiological activity. These so-called structure-activity relationships (SAR) imply the classification of chemical substances based on their molecular structure and their known metabolic effects. The idea is that any substance with a molecular structure similar to the one of another substance that is already known to be toxic would be automatically classed as potentially toxic. On the basis of this classification, the tested substance would be subject to regulation, at least provisionally, until a more exhaustive and slow investigation is able to reliably establish its effects, or absence thereof. In other words, SAR and similar methodologies can be understood as a defense of mechanistic information as basis for regulatory decision making.

A third example is the reversal of the burden of proof. This proposal means that those who want to promote a particular technological process or product would be responsible for showing that it does not imply any negative effects. This is contrary to the currently common approach in regulation, in which the responsibility for demonstrating that a product implies risks falls on other stakeholders, particularly the regulatory agencies. Reversing the burden of proof is concomitant to minimizing false negatives (Harremoës et al. 2002). It is yet again a consequence of empirical investigation into the real-world consequences of the establishment of the burden of proof in regulation.

The preceding examples show that controversies in regulation can have an impact on the standards of proof. The latter can determine methodological choices, meaning that regulatory controversy influences those choices.

## 4 Benefit assessment: Health claim regulation

Our case study from the field of benefit assessment is the regulation of health claims in the European Union (EU). Health claims are affirmations of additional health benefits that a food or nutrient may confer upon its consumer, beyond the obvious nutritional aspects of foods. These claims are found on food labels, and are subject to regulation in many countries (due to the added value of foods identified by such claims). We will center our discussion on the European regulatory process for health claims.

The relevant European regulator, the European Food Safety Authority (EFSA), establishes a hierarchy of evidence as a basis for its evaluation of health claims. In this hierarchy the most important source of evidence are intervention studies in humans. Data produced by such studies are considered indispensable for obtaining authorization of a claim. Intervention studies in humans, of which randomized controlled trials are the most relevant type, are clinical trials similar to the ones used for drug testing in the pharmaceutical sector (double-blinded trials, with a control group that is given a

placebo). Other types of evidence are classified according to EFSA's hierarchy as merely complementary evidence. That is, this kind of evidence can never be decisive when authorizing a claim (EFSA 2009; EFSA 2010; EFSA 2011).

RCTs are placed at the top of EFSA's hierarchy of evidence because the regulator considers that this methodology provides scientific evidence of the highest quality, as compared to other methods. In fact, EFSA's hierarchy ranks all possible scientific methods. Within the first (uppermost) tier of the hierarchy, i.e., human intervention studies, there is a subcategorization: a ranking from RCTs (top) to human intervention studies without control groups nor randomization (bottom). The second tier of the hierarchy is composed of observational (epidemiological) studies, again ranked according to quality, from cohort studies to case studies. The third (bottom) tier of the hierarchy comprises all other types of scientific methods, particularly mechanistic studies.[5] The latter are, as in the previous two tiers, ranked, from mechanistic studies in humans (top) to mechanistic studies in animals, and finally in vitro assays (bottom).

EFSA considers RCTs of the highest importance for claim authorization because data from RCTs allow to establish causal relationships between intake of a food or ingredient and the desired outcome (positive health effects) (Heaney 2008). Establishing causality makes it possible to restrict authorization to those health claims of which the efficacy has been proven beyond any reasonable doubt. That means that consumers can be sure that when they purchase a food identified by such a claim the consumption of this food will produce the desired effect (reduction in false positives).

The need for establishing causality is a very demanding evidence requirement. Designing and executing RCTs in order to generate the necessary data is very resource and time intensive, and can be difficult because of the complexities involved. Moreover, very few health claims have been authorized in Europe, precisely because of the difficulties in convincingly establishing causality between intake and outcome.

A number of critics from the nutrition sciences argue that it would be preferable to reduce the evidence requirements so as to make it possible to base decisions for claim authorizations on data from scientific methodologies other than clinical trials. The authors who reject EFSA's current regulatory approach point out that in many cases evidence from observational studies or concerning the mode of action of a nutrient (mechanistic data) could be sufficient to establish that the desired beneficial effect indeed exists, even though it cannot be conclusively proven by a clinical trial (Biesalski et al., 2011; Bast et al. 2013; Todt and Luján 2017).

The critics argue for an approach that they call evidence-based nutrition, emphasizing the differences between research in the field of pharmaceuticals and the one of nutrition. Their core argument is that the scientific methodologies most useful for generating data in nutrition are not necessarily the same as in pharmacology, due to certain characteristics of nutrition that set it apart from pharmacology. Chief among those characteristics are a) low effective doses, b) constant interaction between different ingredients, as well as ingredients with the entire food matrix, and c) long latency periods (Biesalski et al., 2011; Hendrickx 2013). Against EFSA's methodological monism the critics propose the adoption of a methodological pluralism in order to

---

[5] Mechanistic studies allow for the identification of the mechanisms and biological pathways by which a food ingredient produces the desired positive health effects.

generate evidence that is relevant for regulatory decision making in the field of nutrition.

In analyzing this controversy about the standards of evidence in health claim regulation we can identify, as in the preceding case of risk assessment, an evident relationship between regulatory aims, standards of evidence, as well as scientific methods for generating this evidence. The requirement for establishing causal relationships between intake and outcome in EU health claim regulation is the direct consequence of a very demanding standard of proof, which is aimed at minimizing the marketing of foods with ineffective or fraudulent health claims. In contrast, the reduction in the evidence requirements proposed by EFSA's critics would lead to an increase in the authorization of health claims, meaning more choice for consumers. But, the claims would certainly tend to be less reliable than those that currently obtain authorization. That is because authorization on the basis of data produced by observational or mechanistic studies implies an increase in false positives.

In other words, in benefit assessment we can identify a situation that is very similar to the one in risk assessment, at least as far as controversy is concerned. The definition of the evidence requirements leads to controversy, not only related to the methods employed but also to the wider societal outcomes. And again, we are faced with basically two alternative stances: 1) requiring a very demanding standard of evidence that is centered on the reduction in false positives, with the aim of protecting consumers from incorrect information (claims that obtain authorization even though the relevant effect does not exist); and 2) reducing the evidence requirements, implying an increase in false negatives, so that consumers may enjoy a wider (and more timely) selection of foods with authorized claims (even though the reliability of those authorized claims is likely to be somewhat lower than in case 1).

Our argument here is that deciding which of the two regulatory options is preferable will depend on empirical information on the real-world outcomes of regulation (Luján and Todt 2018).

## 5 Two alternative epistemic policies

Our two case studies clearly show that regulatory controversies possess a methodological aspect. Controversy about decision making and its implications (management of risks or benefits) has direct effects on risk or benefit assessment. This shows how in regulatory science the limit between management and assessment is being pierced in both directions. That is why our naturalistic proposal consists in an assessment of the various regulatory options on the basis of the non-epistemic outcomes of regulation.

As we have already seen, in both risk and benefit assessment, the stances which argue for more permissive standards of proof imply an evidentiary, and therefore methodological pluralism (Verhagen et al. 2019). They reject the requirement for classifying any type of evidence (and methodology) as "indispensable". To the contrary, they consider that varying types of evidence (like mechanistic evidence) and methodology (like short-term tests) are acceptable for data generation (Stegenga 2014).

This methodological pluralism, due to its incorporating evidence from very different sources and quality, implies the need for an additional assessment of the evidence, including its sources, quality, and interrelations. This is obviously not

necessary in the case of strict evidence requirements that focus on one single type of method or evidence (Vandenbroucke, Broadbent and Pearce 2016). Thus we can interpret the regulatory controversies in the following manner: on the one hand, there is the stance which considers certain types of experiments and analyses indispensable in order to be able to establish with a high degree of confidence the presence of causal relationships (between intake of a food and a desired, positive health effect, or between exposure to a substance, and various negative effects). On the other, there is the alternative stance which relies on a global, all-encompassing assessment of all the available evidence (generated with various kinds of methods, and from different sources), and which requires the intervention of expert judgment in order to be able to complete a weight-of-evidence[6] (WOE) analysis.

In nutrition, RCTs are supposed to make expert judgment superfluous, by means of an objective experiment. One single type of method and evidence becomes indispensable. The critics of this approach argue for the use of different sources of scientific evidence (due to the possibility of generating data by all kinds of scientific methodologies).

In toxicology, we have a similar situation. There are those who argue for risk assessment to be based almost exclusively on data on humans, usually from epidemiological studies (in toxicology, for ethical reasons it is obviously not possible to use RCTs). These data must allow for a quantification –through statistical analysis– of the effects of any planned regulatory intervention (like limiting emissions into the environment of a particular substance, or prohibiting the sale of a product). If the scientific data at one's disposal do not allow for such a quantitative analysis of the effects of regulatory action, then no action should be taken (because it would not be scientifically justified) until the necessary data become available (Cox 2015).

In this latter proposal, which requires a previous quantification of the effects of any possible regulatory interventions, the relationship between standards of evidence and methodologies is patent. Decision making would have to take account of the effects of possible interventions. In this case the standards of evidence impose the use of quantitative methods to assess causal hypotheses, as well as quantify the effects of regulatory intervention. Causality here gets defined in probabilistic terms. From the vantage point of a proposal like this one, methodological pluralism is rejected with the argument that no standard of proof based on a weight-of-evidence approach would ever be able to avoid bias (due to expectations, preconceptions, etc.).

To the contrary, those who argue for WOE-type standards of proof consider that the latter facilitate the use of various methodologies in order to obtain decision-relevant

---

[6] In general, a weight-of-evidence analysis is based on the assessment of evidence from different sources, meaning different lines of research, methodologies for data generation, and so on. The basic idea is that while the evidence from one single line of research will not be sufficient to be able to accept or reject a particular hypothesis, taking all the evidence together would be sufficient because the different lines of research support each other (Haack 2008, 2014). The concept of weight-of-evidence is used in environmental studies, toxicology, nutrition and other scientific fields to designate the combination of multiple lines of evidence during an analysis in order to reach a conclusion (Weed 2005). The complexity inherent in joining up evidence from very different sources, as well as varying types and qualities has led to multiple ways of interpreting WOE analyses from a philosophical standpoint (evidence amalgamation: Landes, Osimani and Poellinger 2018; Fletcher, Landes, and Poellinger 2019).

evidence. This stance can be found especially among researchers who defend the use of mechanistic evidence.

As our analysis shows, it is possible to conceptualize at least some of the methodological controversies in regulatory science as controversies between two basic epistemic policies. In the following we will refer to those, respectively, as "*sine qua non*" and "weight-of-evidence". The epistemic policies of the sine qua non (SQN) type consider that a certain type of evidence or methodology is essential and/or that the scientific data have to surpass a "threshold of minimum quality" to be taken into consideration. The alternative epistemic policy we will call weight-of-evidence (WOE). That is an epistemic policy which assesses entire ranges of available evidence without privileging any of them.

## 6 Epistemic policies and the distribution of error

In risk assessment there exist several proposals for the adoption of particular methodologies on the basis of the social consequences of alternative methodological options. Most of those proposals, usually applying cost-benefit analyses (Lave and Omenn 1986), argue that in risk assessment false positives are more likely to produce negative social consequences, as compared to false negatives. An example is the above mentioned STT proposal (Cranor 1995, Cranor 2011). These proposals are usually of the pluralistic kind, and imply a lowering of the standards of evidence.

There are other authors, however, who –also on the basis of cost-benefit analyses– argue *against* a lowering of the standards of evidence. Their principal argument is that an increase in false negatives leads to an increase in the costs supported by corporations, implying less wealth and tax receipts, worse public services (including medical attention), and in the end an increase in deaths (Sunstein 2002).

As we can see, those who defend a lowering of the standards of evidence focus their attention on the immediate deaths from pollution, etc., while their critics are concerned with the indirect deaths due to an increase in economic costs. Even though none of these authors uses the expression "naturalism" in relation to methodological choice, all of those proposals can actually be considered naturalist: the methodological choices depend on empirical information related to the social consequences of those choices.

Regulatory practice shows that there exists a direct relationship between epistemic policies and error distribution. This is important because varying error distributions translate into varying social consequences of regulatory decisions. We will show this in the case of the two epistemic policies we have defined above.

The SQN epistemic policy gives preference to, or requires a particular kind or type of evidence for regulatory decision making. All other types of evidence will be taken into account for secondary purposes only, particularly for justifying the need for further research. If, however, the type of evidence required by a SQN epistemic policy is not available, then the hypothesis cannot be confirmed. The implication is that regulation cannot be justified. Thus, SQN is a very cautious approach. To consider an affirmation empirically justified under this epistemic policy requires very demanding evidence. It is reasonable to suppose that some of the hypotheses that get rejected due to the absence of the required type of evidence are in fact true. SQN, as we have already seen, reduces false positives, implying an increase in false negatives.

To put this into the perspective of our two case studies: in risk assessment, some of the substances that under an SQN epistemic policy would not be considered dangerous in reality are. In benefit assessment, a number of the foods that would not be considered beneficial for health actually are. The advantage in both cases is that we can be extremely sure that the substances classified as, respectively, pernicious or beneficial actually have those qualities. In other words, SQN leads to very precise and reliable results with respect to the objects of regulation, be they chemicals, foods, nutrients, or any other kind of substances. A more demanding standard of evidence, however, also implies an increase in false negatives. In risk assessment, for instance, this inexorably leads to the authorization (or non-prohibition) of some substances that are pernicious to health and/or the environment.

The SQN approach is not necessarily methodologically monistic. It could perfectly well be pluralistic. As an example: Leuridan and Weber (2011) argue that in risk management the evidence to be taken into account should be derived from bioassays and/or epidemiological studies, unless there is also mechanistic evidence available. What these authors believe is that in order to be able to affirm a causal connection between exposure to a substance, and a particular outcome, it is always necessary to have at one's disposal both probabilistic and mechanistic evidence. In practice, what this proposal means is a more demanding standard of proof, which automatically increases the number of false negatives. One example for this approach is the controversy about tobacco regulation in the 1970s. Those opposed to regulation argued that despite the available epidemiological evidence, the absence of data on a plausible mechanism meant that regulating tobacco was not considered supported by the evidence (Gillies 2011; Canali 2019).

Other, similar proposals are related to inference guides and models of extrapolation (particularly from high to low doses, and from animals to humans). There are authors who argue that such models necessarily have to be based on mechanistic information (Clewell 2005). In other words, the extrapolation models have to be biologically plausible. These proposals are meant as alternatives to the systematic use of linear extrapolation models (which presuppose toxicity at very low doses) (Krewski et al. 2009).

The alternative epistemic policy, the WOE epistemic policy, calls for the taking into account of the entire spectrum of available information from different sources. The crucial difference to SQN, as we have seen, is that there is no particular preferred or indispensable type of evidence or method. In contrast, the idea is that a combination of individually not necessarily very convincing lines of evidence from multiples sources and types of research will lead to the mutual strengthening of each of them.

As a general rule, WOE produces more false positives than SQN. That is since WOE considers more lines of evidence, the probability of concluding in favor of the validity of the hypothesis are higher than in the case of a SQN epistemic policy. There also is a pragmatic aspect to this distribution of errors between the two epistemic policies: some of the lines of evidence required by SQN are usually difficult to obtain, meaning that in practice WOE implies a reduction in false negatives, and SQN a reduction in false positives.

In risk assessment there have been proposals for applying the precautionary principle. This is equivalent to a lowering of the standards of proof, because mere indications for risk would be sufficient for subjecting to regulation any substance which might

have severe impacts on health or the environment. The precautionary principle is a response to a preoccupation with false negatives (rather than false positives), because false negatives are understood to lead to more severe negative outcomes. Even though as a general rule this rationale might look correct, this is less clear if we go into the details. For instance, substances subject to risk assessment provide benefits, too, particularly economic benefits which in one way or another are beneficial for the entire population. Thus, the preoccupation with false negatives in risk assessment in the case of a substance which carries very few risks but provides important benefits is fairly absurd.

In benefit assessment, however, all of this is even more complex. Looking at benefits only, false positives translate into the marketing of supposedly beneficial products which in reality are not. For pharmaceuticals this would mean that their hoped-for therapeutic effect does not exist, with all the concomitant negative effects for public health. False negatives make it impossible for society to take advantage of certain products, like pharmaceuticals meant to cure serious diseases, again with negative effects for public health. In addition, pharmaceuticals are subject to risk assessment in order to identify any risks they pose to their consumers (Landes, Osimani and Poellinger 2018).

This means that in both risk and benefit assessment we have to clearly define which exactly are the risks and the benefits we are taking into consideration, because only in this way can we determine which errors imply more serious costs for society. In our example on health claims, we have to take into account that the ingredients and foods in question have already passed any necessary food risk assessment. Obtaining a health claim does not affect in any way the safety of consumption of those foods, because there are no risks (and if there were, they would have been detected in a previous risk assessment, with total independence of the subsequent assessment of health claims).

In sum, the non-epistemic consequences of both approaches depend on the regulatory consequences of error distribution.


## 7 Naturalism and epistemic consequentialism

Our proposal for assessing alternative epistemic policies is naturalistic, normative, and consequentialist. It is naturalistic because it recurs to empirical information about the epistemic and regulatory (real-world) consequences of epistemic policies. It is normative because its aim is not only the exploration of factors which might exert an influence on the election of a particular epistemic policy, but also the comparative evaluation of such epistemic policies in order to ascertain which of the available alternatives makes it more likely to fulfill the objectives of a regulation. And it is consequentialist because this evaluation of the epistemic policies flows from the analysis of the consequences on society, human health and the environment of the different regulatory options (Luján and Todt 2018).

The evaluation proceeds as follows:

1. Ascertain the distribution of errors for each of the epistemic policies under consideration;

2.  Determine the social and other consequences (for instance, on public health) of the distribution of errors, as mediated by regulation. The population-level effects of regulation have to be assessed as if the two epistemic policies were general heuristic rules.

The proposal we are arguing for can be considered a generalization of ideas developed by Shrader-Frechette and Cranor in the field of risk assessment, according to which the social consequences of regulation are to be legitimately taken into account in methodological decisions.

As far as benefit assessment is concerned, our proposal differs from other, similar ideas on the use of empirical data in regulation, because we suggest the generation of empirical data on outcomes in each particular case, and then feeding back this data into decision making on methodologies used for regulation in this specific context. There do exist proposals in pharmacology about taking into account data on regulatory outcomes, but they do not argue for explicitly generating such data for each specific context. Rather, they propose using generally available evidence to this end. For example, Osimani (2014) defends different, alternative standards of evidence for assessing the risks and benefits of pharmaceuticals. Hansson (2020), too, considers legitimate to tighten or relax standards of proof depending on each particular case, like, for instance, risks of pharmaceuticals, as opposed to substitution of pharmaceuticals. Both authors recur to epistemic, as well as non-epistemic (related to the social consequences) outcomes of regulation. They do, however, have a rather "static" conception of scientific evidence, in the sense of readily available, off-the-shelf evidence. Their argument is that such evidence can be used in particular cases as basis for regulation. But, importantly, they do not analyze the way in which their proposals about standards of proof might exert influence on ongoing research in regulatory science.

We can illustrate the use of a procedure similar to the one we are proposing with an example of how the generation of empirical data on regulatory outcomes can be used to facilitate methodological decisions. Andreoletti and Teira (2019) argue that in regulatory science costs and impacts of regulation have to be taken into account. Their argument is that empirical information is to be used to decide between two regulatory strategies, one base on standards, and the other on rules. The rule-based strategy is coherent with methodological monism, while the standards-based one is coherent with pluralism. The authors' analysis can thus be considered to be in line with our argument that methodological choices ought to be based on evidence about the consequences of alternative regulatory options. Andreoletti and Teira conclude that in their case study (pharmaceuticals) a rule-base strategy is superior to one based on standards: methodological monism (RCTs) is preferable to pluralism. Teira (2020) also argues for methodological monism (again, RCTs) in pharmacology on the basis of non-epistemic (moral, political, etc.) regulatory outcomes (like consumer protection or impartiality).

To sum up our own point of view: we argue that research about the ultimate outcomes of regulation is a legitimate input for methodological decisions. The legitimacy of the influence of pragmatic values (here, relative to regulation) on scientific research depends on the obtainment of information about the relationship between scientific evidence and different, alternative regulatory options.

The interaction between scientific knowledge and public policy has generally been conceptualized as a one-way relationship. This is expressed, for instance, in the often repeated line that decisions in public policy have to be based on the best available scientific knowledge, or the most accurate evidence, etc. Underlying this particular conceptualization is the belief that there is but a single way in which scientific knowledge can advance, or only one particular model that fits, and so on. In other words, there is only one single way in which regulatory decision making can best be informed by science. This line of thinking considers scientific methodology immune to the context of its application. Our analysis of the controversies in regulatory science reveals that this way of thinking about science and regulation is an ideal one, in the sense that it is impossible to be applied in practice.

As we have already seen, in risk assessment many authors argue that the aims of regulatory science (to inform decisions) exert an influence on the inevitable methodological choices. These authors consider that, as a general rule, less accurate data about risks will lead to better regulation (in spite of the increase in false positives). This argument is based on the population-level effects of public policy. Considering the substantial number of chemical substances currently on the market (several millions), and the material impossibility of thoroughly testing all of these for possible negative effects, a regulation that is more tolerant towards false positives will tend to subject to study and regulation a larger number of substances, even though some of the regulated substances in reality are not harmful. In other words, given its wider effects, a regulatory strategy based on scientific knowledge of "lower quality" (less accuracy, etc.) may be advantageous as compared to one that strives for the "best possible" knowledge about a substance's health or environmental effects before taking a regulatory decision.

The same argument applies to health claim regulation. We want to illustrate this argument with an example. Let us suppose that we have 100 ingredients which are part of a certain number of food stuffs. If we had complete knowledge as to these 100 ingredients' characteristics and their interaction with the human metabolism, then we would know that (for example) 50 of those ingredients are beneficial for human health, while the other 50 are neutral (no particular benefits). If our standard of evidence was very strict, then (for instance) only 30 of those ingredients would get classified as beneficial. If to the contrary our standard was a more relaxed one, then (again, for instance) 70 of those ingredients could be classed as beneficial. The fundamental question now is which of those two results (classifying respectively 30 or 70 ingredients as "beneficial") is better for promoting public health. There is no obvious answer to this question. And it is clear that the answer cannot be an a priori answer, in which we simply would go with the "best possible" or "most accurate" science. Taking into account any aggregate (population-level) effects, it is perfectly possible that the option based on more relaxed standards (less precise methods) is better in promoting public health. Our argument is that which of the two options is the preferable one will depend on empirical information about their outcomes for public health. The regulatory option determines the standards, and these in turn influence our methodological choices.

A demanding standard of proof will very effectively protect consumers from incorrect information and fraudulent claims, so they can be confident that the positive health effects advertised on food labels are actually for real. The downside of this regulatory strategy is that many products which possess beneficial characteristics will

not be labeled as such (due to a lack of adequate evidence), meaning that these properties (and the concomitant products) will in many cases go unnoticed by consumers.

A more permissive standard of proof, however, would likely lead to a larger number of foods receiving an (officially approved) health claim, meaning more choice and information for consumers. In this case the downside is that a part of the products with claims in reality do not possess the advertised beneficial effects. But even so, those consumers who chose to improve their eating habits on the basis of health claims will, overall, have at their disposal more possibilities and options.

In both risk and benefit assessment, the "best" standard of proof, in the sense of producing the best overall outcomes, does not depend on a priori considerations. Rather, it depends on empirical information about the characteristics of chemicals or foods, consumer habits, etc., as well as the aims of public policy. Choosing the most adequate standards of proof will therefore depend on empirical data about regulatory outcomes (Luján and Todt 2018).

## 8 Conclusions

In this paper we have shown that the aims of regulation can legitimately influence methodological choices in regulatory science. In other words, the relationship between assessment and management is bidirectional. That is because the standards of evidence determine the distribution of errors, which in turn affect the real-world outcomes of regulation (on health, the environment, innovation, etc.). A naturalistic analysis of the interrelations between regulatory aims, standards of proof and scientific methodology might therefore facilitate a solution to some of the current controversies in regulatory science.

In an ideal environment, i.e., without resource and time limits, we could certainly, case-by-case, analyze the exact relationship between the aims of regulation, the standards of proof and the methods employed. This relationship depends on a number of factors: if we are concerned with risk or benefit assessment, the magnitude and type of the risks or benefits, the concrete characteristics of the substances under study, the opportunity costs related to regulation, as well as the characteristics of the consumers who might purchase or use these substances. Given that in practice such a detailed analysis will be impossible, we have to rely on heuristic rules: the epistemic policies which define more or less general regulatory strategies.

Our analysis of the two case studies has allowed us to identify two general epistemic policies, which we have termed SQN and WOE. Recurring to a SQN strategy usually implies that some of the substances that are considered innocuous in reality pose risks; and that some of the substances that are considered not to provide any benefits in reality do. We can, however, be very sure that the substances declared to entail, respectively, risks or benefits actually do. The alternative WOE strategy means that a higher percentage of substances will be classified, respectively, as dangerous or beneficial, even though a part of them do not possess those characteristics.

Our argument is that only empirical study of the outcomes of the alternative regulatory options will provide the data that allows us to choose between the options, including standards and methods. In this way we can challenge the fundamental dogma

of the unidirectional relationship between scientific knowledge and public policy, according to which information can and must only flow from science towards policy.

## Compliance with ethical standards

**Conflict of interest**   The authors declare that they have no conflict of interest.

## References

Andreoletti, M., & Teira, D. (2019). Rules versus standards: What are the costs of epistemic norms in drug regulation? *Science, Technology and Human Values, 44*(6), 1093–1115.

Bast, A., Briggs, W., Calabrese, E., Fenech, M., Hanecamp, J., Heaney, R., Rijkers, G., Schwitters, B., & Verhoeven, P. (2013). Scientism, legalism and precaution—Contending with regulation nutrition and health claims in Europe. *EFFL, 6*, 401–409.

Biesalski, H. K., et al. (2011). 26th Hohenheim consensus conference, September 11, 2010 scientific substantiation of health claims: Evidence-based nutrition. *Nutrition, 27*, S1–S20.

Canali, S. (2019). Evaluating evidential pluralism in epidemiology: Mechanistic evidence in exposome research. *HPLS., 41*, 4. https://doi.org/10.1007/s40656-019-0241-6.

Cartwright, N., & Hardie, J. (2013). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.

Cartwright, N., & Stegenga, J. (2011). A theory of evidence for evidence-based policy. In W. Twining, P. Dawid, & D. Vasilaki (Eds.), *Evidence, inference and enquiry* (pp. 291–322). Oxford: Oxford University Press.

Clewell, H. (2005). Use of mode of action in risk assessment: Past, present, and future. *Regulatory Toxicology and Pharmacology, 42*, 3–14.

Cox, L. A. (2013). Improving causal inferences in risk analysis. *Risk Analysis, 33*(10), 1762–1771.

Cox, L. A. (2015). *Breakthroughs in decision science and risk analysis*. Hoboken: John Wiley & Sons, Inc..

Cranor, C. (1993). *Regulating toxic substances. A philosophy of science and the law*. New York: Oxford University Press.

Cranor, C. (1995). The social benefits of expedited risk assessment. *Risk Analysis, 15*(4), 353–358.

Cranor, C. (2011). *Legally poisoned*. Cambridge: Harvard University Press.

Cranor, C. (2017). *Tragic failures: How and why we are harmed by toxic chemicals*. New York: Oxford University Press.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science, 67*, 559–579.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.

EFSA. (2009). Technical report: Briefing document for member states and European Commission on the evaluation of article 13.1 health claims. *EFSA Journal, 7*(1386), 1–10.

EFSA. (2010). *Guidance on human health risk benefit assessment of foods*. Tabiano: EFSA.

EFSA. (2011). Scientific and technical guidance for the preparation and presentation of an application for authorisation of a health claim (revision 1). *EFSA Journal, 9*(2170), 1–36.

Elliott, K. (2011). *Is a little pollution good for you? Incorporating societal values in environmental research*. New York: Oxford University Press.

Fletcher, S. C., Landes, J., & Poellinger, R. (2019). Evidence amalgamation in the sciences: An introduction. *Synthese, 196*, 3163–3188.

Fuller, S. (2000). *The governance of science*. Buckingham: Open University Press.

Giere, R. N. (1985). Philosophy of science naturalized. *Philosophy of Science, 52*(3), 331–356.

Giere, R. N. (1998). *Naturalized philosophy of science. In Routledge Encyclopedia of Philosophy*. New York: Routledge.

Gillies, D. (2011). The Russo–Williamson thesis and the question of whether smoking causes heart disease. In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 110–125). Oxford: Oxford University Press.

Haack, S. (2008). Proving causation: The holism of warrant and the atomism of Daubert. *Journal of Health & Biomedical Law, 4*, 253–289.

Haack, S. (2014). Evidence matters. Science, proof, and truth in the law. New York: Cambridge University Press.

Hansson, S. O. (2020). Values in pharmacology. In A. LaCaze & B. Osimani (Eds.), *Uncertainty in pharmacology. Epistemology, methods, and decisions* (pp. 375–396). Cham: Springer.

Harremoës, P., et al. (Eds.). (2002). *The precautionary principle in the twentieth century: Late lessons from early warnings*. London: Earthscan.

Heaney, R. (2008). Nutrients, endpoints, and the problem of proof. *Journal of Nutrition, 8*(138), 1591–1595.

Heesen, R., Bright, L. K., & Zucker, A. (2019). Vindicating methodological triangulation. *Synthese, 196*, 3067–3081.

Hendrickx, K. (2013). Rivaling evidence-bases and politics in regulatory science. *Food, Science & Law*, vol. 4, Http://hdl.handle.net/2268/162196

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*(5), 295–300.

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. Oxford University Press.

Kitcher, P. (1993). *The advancement of science*. New York: Oxford University Press.

Krewski, D., Andersen, M. E., Mantus, E., & Zeise, L. (2009). Toxicity testing in the 21st century: Implications for human health risk assessment. *Risk Analysis, 29*, 474–479.

Landes, J., Osimani, B., & Poellinger, R. (2018). Epistemology of causal inference in pharmacology. *Euro Jnl Phil Sci, 8*, 3–49.

Laudan, L. (1987). Progress or rationality? The prospects for normative naturalism. *American Philosophical Quarterly, 24*, 19–31.

Laudan, L. (1990). Normative naturalism. *Philosophy of Science, 57*, 44–59.

Lave, L. B., & Omenn, G. S. (1986). Cost-effectiveness of short-term test for carcinogenicity. *Nature, 324*(6092), 29–34.

Leuridan, B., & Weber, E. (2011). The IARC and mechanistic evidence. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 91–109). Oxford: Oxford University Press.

Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press.

Luján, J. L., & Todt, O. (2015). The role of values in methodological controversies: The case of risk assessment. *Philosophia Scientiae, 19*(1), 45–56.

Luján, J. L., & Todt, O. (2018). The dilemmas of science for policy. *EMBO Reports, 19*(2), 194–196.

Luján, J. L., & Todt, O. (2020). Standards of evidence and causality in regulatory science: Risk and benefit assessment. *Studies in History and Philosophy of Science Part A, 80*(April), 82–89. https://doi.org/10.1016/j.shpsa.2019.05.005.

Mayo, D., & Miller, J. (2008). The error statistical philosopher as normative naturalist. *Synthese, 163*, 305–314.

Osimani, B. (2014). Safety vs. efficacy assessment of pharmaceuticals: Epistemological rationales and methods. *Preventive Medicine Reports, 1*, 9–13.

Osimani, B. (2020). Epistemic gains and epistemic games: Reliability and higher order evidence in medicine and pharmacology. In A. LaCaze & B. Osimani (Eds.), *Uncertainty in pharmacology. Epistemology, methods, and decisions* (pp. 345–372). Cham: Springer.

Reiss, J. (2015). A pragmatist theory of evidence. *Philosophy of Science, 82*(3), 341–362.

Shrader-Frechette, K. (1989). Scientific progress and models of justification. In Goldman (Ed.), *Science, technology, and social progress* (pp. 196–226). Bethlehem: Lehigh University Press.

Shrader-Frechette, K. (1991). *Risk and rationality: Philosophical foundations for populist reforms*. Berkeley: University of California Press.

Shrader-Frechette, K. (1994). *Ethics of scientific research*. Lanham: Rowman & Littlefield.

Solomon, M. (2001). *Social empiricism*. Cambridge: MIT Press.

Steel, D. (2015). Philosophy and the precautionary principle: Science, evidence, and environmental policy, Cambridge University Press.

Stegenga, J. (2014). Down with the hierarquies. *Topoi, 33*, 313–322.

Sunstein, C. (2002). Risk and reason: Safety, law, and the environment, Cambridge university press.

Teira, D. (2020). On the normative foundations of pharmaceutical regulation. In A. LaCaze & B. Osimani (Eds.), *Uncertainty in pharmacology. Epistemology, methods, and decisions* (pp. 417–437). Cham: Springer.

Todt, O., & Luján, J. L. (2017). Health claims and methodological controversy in nutrition science. *Risk Analysis, 37*(5), 958–968.

Vandenbroucke, J. P., Broadbent, A., & Pearce, N. (2016). Causality and causal inference in epidemiology: The need for a pluralistic approach. *International Journal of Epidemiology, 2016*, 1776–1786.

Verhagen, H., Robinson, T., Gallani, B., Hugas, M., Kleiner., J., Hardy, A., & Devos, Y. (2019). EFSA's third scientific conference 'science, food, Society': concluding remarks. *EFSA Journal, 17*. https://doi.org/10.2903/j.efsa.2019.e170723.

Weed, D. (2005). Weight of evidence. *Risk Analysis, 25*, 1545–1155.