

RULES VERSUS STANDARDS: WHAT ARE THE COSTS OF EPISTEMIC NORMS IN DRUG REGULATION?

Mattia Andreoletti & David Teira

Forthcoming in Science, Technology and Human Values

Abstract

Over the last decade, philosophers of science have extensively criticized the epistemic superiority of Randomized Controlled Trials (RCTs) for testing safety and effectiveness of new drugs, defending instead various forms of evidential pluralism. We argue that scientific methods in regulatory decision making cannot be assessed in epistemic terms only: there are costs involved. Drawing on the legal distinction between rules and standards, we show that drug regulation based on evidential pluralism has much higher costs than our current RCT-based system. We analyze these costs and advocate for evaluating any scheme for drug regulatory tests in terms of concrete empirical benchmarks, like the error rates of regulatory decisions.

1. Introduction

During this last decade, philosophers of science have addressed the use of scientific evidence for policy-making purposes in many different ways. Such discussions have often adopted a Platonic stance: epistemology goes first. Given a policy-making problem, the philosopher should just identify which sort of evidence will better solve it. But the policy-making process may often be more complex. Our goal in this paper is to show the limitations of this Platonic approach, advocating instead for a more experimental take: rather than just criticizing the evidence actually used on purely a priori grounds and present a principled alternative, philosophers should also care about the empirical benchmarks that would allow the public to see whether their principled alternatives work better than the already established approach.

Our case in point is drug regulation, in which for the last five decades randomized clinical trials (RCTs) have been used to test the safety and efficacy of new compounds. Philosophers of science have extensively criticized the epistemic superiority of clinical trials for such purpose and have defended various forms of evidential pluralism, on purely epistemic grounds.¹ We first restate this epistemic debate in legal terms (section 2): for regulatory purposes, evidential pluralism implies a *standard*-based decision-making process, whereas the current approach to drug testing is based

instead on *rules*. Following the 1962 Food and Drug Administration Act, most regulatory decisions operate on a simple rule: if there are two positive RCTs, approve the drug.² Several philosophers of science contend instead that regulators would do better if they addressed the question of whether a drug is safe and effective searching for the best evidence available, whatever its sources. This is a standard-based approach.

The economic analysis of law has appraised the rule versus standards distinction in terms of their respective costs. We draw on this approach to argue that, in order to avoid Platonism, any normative proposal to reform our regulatory schemes should consider not just the epistemic superiority of an alternative, but also the costs of adjudicating drug approvals on that basis. Not just the costs of purely epistemic deliberation on the basis of a rule or a standard, but also the costs of protecting deliberation from any external conflict of interest. In section 2, we introduce the rules versus standards distinction, highlighting the current consensus in the economic analysis of law: rules have fewer costs than standards when it comes to decision making. In section 3 we restate in these terms a number of recent proposals to change the epistemic yardstick we currently use in pharmaceutical regulation. In section 4, we draw on the experience of the FDA to analyze the costs of regulatory decision making under the external pressure of pharmaceutical interests. We compare decisions achieved through RCT-based rules and decisions attained with a standard-based deliberation. Our conclusion is that, once costs are taken into account, there is no compelling reason to expect that a standard-based system may improve the current RCT-based approach

Yet, we also defend, in the fifth and final section, that a proper clarification of this normative debate requires an empirical benchmark of regulatory success. We propose to use the error rate of regulatory agencies, which is defined as the number of products withdrawn from the market after approval, drawing on the experience of the FDA. Many may object to this index on the basis of the unreliability of the very agency issuing the figures: the FDA, they will argue, is captured by pharmaceutical interests and, as a consequence, it recalls fewer drugs than it should. In our view, this is the most extreme position in the rules versus standards debate in pharmaceutical regulation: no regulatory yardstick can resist the power of lobbies. Yet, as we will conclude, the capture argument also relies on a benchmark of regulatory success that is probably Platonic as well. Overcoming Platonism in the discussion of pharmaceutical regulatory issues is the final recommendation of this paper.

2. RULES VERSUS STANDARDS IN DRUG REGULATION

Let us first introduce the distinction between rules and standards (see also Luppi and Parisi 2011). Both are *legal directives* with a conditional structure of the form “If X, then Y.” The antecedent is

the “trigger”: an event plus some concomitant circumstances. The consequence is the “response.” Rules have a hard-empirical trigger, and a hard-determinate response. Standards have a soft evaluative trigger, and a soft guided response (Kaplow 1992; Schlag 1985).

“Dogs are not allowed in bars” is a rule of the form “If x is a dog, x is not allowed in a bar.” The trigger is an unambiguous fact (either x is a dog or not); the response is not open to interpretation. Rules have justifications: the reason why they were originally established. Preventing dogs from entering bars might have been justified in order to avoid disturbances to customers. According to (Schauer 1993), rules should always be implemented independently of their justification: seeing-eye dogs are surely trained not to disturb anyone in public spaces, but they are dogs nonetheless and, according to the rule, they should not be allowed in bars. A bear will surely be a disturbance in any bar, but the rule does not apply directly to bears. Rules are then based on an *entrenched generalization*: they are at once *under-inclusive* (the bears) and *over-inclusive* (the eye-seeing dogs). Standards are comparatively flexible: “Only good dogs are admitted.” The goodness of a dog should be assessed according to the justification of the standard. Someone should consider what may disturb customers and decide whether each particular dog may behave accordingly (that is the soft evaluative trigger), and only then, may she allow the dog in the bar. These are, of course, ideal types: actual laws mix rules and standards in varying degrees.

Scholars in the field of Law and Economics have discussed the differences between standards and rules in terms of their respective costs, broadly conceived (Kaplow 1992).³ There are, of course, costs to both promulgating and enforcing laws, and these costs may vary, depending on whether legislators choose rules or directives. Hard triggers have a high degree of specificity: they enumerate the precise circumstances under which the rule should be applied, and identifying such circumstances involves clear *ex ante* costs. For example, of all the possible tests that a pharmaceutical regulator might use to establish the effects of a treatment, which one should the legislator choose in order to promulgate the relevant legislation? How should she find out? Standards are typically easier to promulgate, since they have a lower degree of specificity: e.g., “harmful treatments are prohibited in pharmaceutical markets.” Yet, standards are costlier to enforce: the adjudicator should establish in each case how to apply them. For example, pharmaceutical regulators will have to decide for each new treatment on what basis they decide on their potential harms.

There are also costs from the standpoint of the other stakeholders concerned with the laws. The pharmaceutical industry, for instance, should decide in which treatments to invest according to the predictability of regulatory approval. Rules are typically more predictable than standards. In case

there is a potential regulatory mistake (e.g., a harmful treatment entering the market) and patients have to litigate, it is easier to argue about the defective application of rule than a standard. Of course, rules are not preferable to standards by default in every realm (Kaplow 1992; Friedman and Wickelgren 2013), but for the last fifty years rules have played a predominant role in American drug markets.

Following the passage of the 1962 Food and Drug Administration Act, and up until 2016, drug regulation in the United States has been based on a particular combination of rules and standards. The 1962 Act established that drug manufacturers had to apply to the FDA for the approval of new treatments. Such application would only be considered if:

[I]t includes substantial evidence consisting of adequate and well-controlled investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded that the drug will have the effect it purports or is represented to have under the conditions of use prescribed.

The definition of a “well controlled investigation” was further clarified in 1969 (Section 314.126 of Title 21 of the Code of Federal Regulations), when it was formally quantified as two randomized clinical trials plus one previous or posterior confirmatory trial. Clinical trials are comparative experiments in which the effects of an experimental drug are compared to the standard treatment or to a placebo, testing the hypothesis that there is no difference between them according to a given statistical design. Once the manufacturer has collected all the necessary evidence, it is submitted to the FDA for review in the form of a New Drug Application (NDA).

FDA approval is then generally based on a rule: *if there are two positive RCTs, accept*. But there are also occasions to use standards. For instance, the reviewers might disagree on the quality of the evidence submitted. In such cases, the FDA may summon an advisory committee of experts to resolve the dispute. Their conclusion is not mandatory for the FDA, although it is usually accepted (Zuckerman 2006). The FDA directly appoints members of the advisory committee from experts in the relevant scientific fields. For example, the Oncologic Drug Committee consists of 13 voting members including one representative of patients/consumers. The industry is represented with one non-voting member.

Usually the FDA and the company that submitted the NDA present all the relevant data to the members of the committee, who should then proceed to a public collective deliberation. Once this deliberation ends, each member of the committee must vote on some specific yes-or-no questions

formulated by the FDA, such as “Given the current knowledge, does this medicine have a demonstrated benefit?” (Urfalino 2012). The FDA itself stresses the fact that it pays attention not only to the votes, but also to the deliberation process (Urfalino 2012). The entire process is as transparent as possible: the meetings of the committees are open to the public, and the minutes are then published on the FDA website.

In general, the questions for new drugs concern “whether the safety and effectiveness information submitted for a new drug is adequate for marketing approval” (Center for Drug Evaluation and Research, n.d.). The members of these advisory committees should thus cast their votes according to a standard: *if the evidence presented is substantive enough, grant market approval*. This is clearly a soft evaluative trigger, as compared to the outcome of an RCT (e.g., either it reaches statistical significance or not) and each voter may interpret it differently.

The coexistence of rules and standards in drug regulation shows that both decision procedures are possible in pharmaceutical regulation. The primacy of rules seems to illustrate a general point in the economic analysis of law: rules are more efficient than standards when adjudication frequently occurs in settings characterized by common traits (Fon and Parisi 2007). If decisions are rare, it may be easier not to spend much effort in articulating a highly specific rule, but rather promulgate a standard and leave to the adjudicators to make a decision based on the circumstances of the case. For frequent decisions on complex issues by specialized courts, by contrast, Fon and Parisi (2007) shows that, under stylized assumptions, rules lower adjudication costs. Indeed, frequency of decisions is an issue at the FDA: the agency screens dozens of drugs annually following standardized procedures.⁴ Identifying a positive RCT as the trigger has also become a relatively standardized task. Years of methodological debate have yielded checklists and guidelines available to reviewers for the overall assessment of the study design. Although the fine details of the statistical assessment of the trial are trickier, any positive an RCT must show a statistically significant difference between the two compared interventions (the famous $p < 0,05$) in the primary endpoint of the treatment. Agreement on the triggers seems to be usual enough at the FDA, since comparatively few decisions on new drugs are reached through committees (which are appointed to sort out disagreements on evidence).

There is nothing inevitable about the use of RCTs-based rules in pharmaceutical regulation. The consensus formed among pharmacologists in the late 1950s about the superiority of RCTs to test all sorts of treatment may change.⁵ When the FDA adopted RCTs as regulatory yardsticks the assumption was that it was better than any other method of causal inference about treatments. As we discuss in the following section, philosophers of science have argued in different ways against

this assumption. The open question, in this framework, is whether the benefits of adopting a different rule or standard make up for the costs of dispensing with the simplicity of RCTs.

3. EPISTEMOLOGY FIRST: THREE APPROACHES

Historically there have been different ways to assess the safety and efficacy of new treatments. Between 1900 and 1950 expert clinical judgment was the main approach in the assessment of the properties of pharmaceutical compounds, both in Britain and the United States. An experienced clinician would administer the drug to a series of patients he considered more apt to benefit from it. His conclusions would be presented as a case report, with the details of each patient's reaction to the treatment. The alternatives were first laboratory experiments and then controlled clinical trials (from which RCTs would later emerge). The former would proceed either *in vitro* or *in vivo* (on animals and patients): considered superior by clinicians with a scientific background, its scope was usually restricted to safety considerations. It soon gave way to comparative trials, in which two treatments were alternated on the same patient or administered in two groups of patients (simultaneously or not). The arrangements to secure the comparability of the two treatments were the controls and they adopted different forms. The following items counted as controls in these trials: the patients' eligibility criteria, the way treatments were allocated (alternation and randomization), uniformity in administration of treatments and patients' blinding. They were not necessarily used all at once. Statistical reports from controlled trials conveyed their results with different degrees of sophistication. The standardization of what we now call an RCT began with the British Streptomycin trial in 1947 and concluded with the adoption of RCTs as a regulatory yardstick by the FDA in the 1960s. RCTs articulated the controlled experiment with the template statistical design first advocated by Ronald A. Fisher. As mentioned above, they are comparative experiments in which a new treatment is run against either the usual therapy or a placebo on, at least, two groups of patients. Other than the treatment, the patients should be treated exactly alike so that any difference in the outcome can be attributed to the single difference between the two groups (the intervention). A statistical protocol will assist regulators to organize the experiment and interpret the outcome, thereby realizing the superiority of RCTs for causal inference.

The different methods for drug testing were all potential candidate triggers in a regulatory *rule*. As compared to these other methods, RCTs are considered superior at grasping the causal connections between a treatment and its effects. Since regulatory agencies are interested in the safety and efficacy of these effects, it is only natural that RCTs were adopted as triggers in pharmaceutical regulatory rules. If RCTs are *a priori* better than any other form of causal evidence, why look elsewhere? Philosophers of science have challenged this purported superiority of RCTs with

different arguments, advocating either for more complex decision rules or, simply, standard-based decisions.⁶ Let us consider three of these approaches with different versions of what we call the “epistemology first thesis,” in ascending order of decisional complexity.

The EBM+ consortium is a British research team that has been defending for the last decade the necessity to combine mechanistic and statistical evidence about causal interventions in medicine in order to fully grasp their effects.⁷ RCTs on their own are fallible, they argue, and knowledge of mechanisms help in transporting causal claims to new populations beyond the trial (Clarke et al. 2014), a clear concern for any pharmaceutical regulator. In their recently published handbook *Evaluating evidence of mechanisms in medicine* (Parkkinen et al. 2018), the EBM+ consortium presents standardized tools so that any stakeholder in a decision-making process involving medical treatments can assess the quality of the available evidence for or against the interventions. These tools are checklists that assign scores to each intervention according to the variety and quality of the evidence supporting it. Decision processes are complex, the authors acknowledge, and “evidence and its appraisal are but one part of the mix” (Parkkinen et al. 2018, p. 35). EBM+ does not offer any guidance for decision, but, in our view, their evidence assessment checklists could easily replace trials in any regulatory decision rule: “if a treatment scores above this threshold in the checklist, accept.” Except that the assessment of new treatments would now be more complicated given the new sources of evidence, in addition to trials, that regulators should consider. The costs of the adjudication process will rise given the higher (implicit) complexity of the rule. The EBM+ expectation, of course, is that regulatory decisions will improve, reducing the number of type I or II errors.

Barbara Osimani has led a grant from the European Research Council to study alternative ways to use evidence in pharmacology. Landes, Osimani, and Poellinger (2018) have criticized the regulatory focus on RCTs, since they do not provide adequate evidence of potential harms. RCTs assess causal effects of treatments on a sample of patients under controlled conditions. These studies rarely capture the heterogeneity of the actual populations to which the treatment will be administered, and this heterogeneity is a source of potentially harmful effects: not every patient will react to the treatment as those studied in the trial. RCTs may provide a general assessment of safety and efficacy, but they do not capture potential adverse effects on particular groups of patients. There is room for improving regulatory decisions if more of these latter effects are detected before market approval. Therefore, Landes, Osimani, and Poellinger (2018) advocate for a more complex approach in which Bayesian methods will allow decision makers to take into account different forms of evidence that grasp better the potentially harmful effects of treatments as well as their

efficacy. Unlike the EBM+, this approach is willing to consider evidence beyond trials and mechanisms. Moreover, Landes, Osimani, and Poellinger (2018) dare to propose a decision rule about new treatments. In this rule, the evidence about treatments will feed the expected utility function of the decision maker, so that acceptance will happen when the utility reaches a certain threshold. Here, the complexity of the decision process increases both in terms of the amount of evidence to be assessed and the procedure to reach a decision, but we are still within a rule-based framework.

Nancy Cartwright, one of the leading philosophers of causality, has explicitly argued against the “algorithmic” interpretation of causal evidence in policy-making processes, defending instead a deliberative approach. For Cartwright, we live in a world of dappled causal capacities: there is no a priori guarantee that making a medical intervention work in a clinical trial will make it work the same way in a different situation. Causal interventions do not work alone, but in combination with supporting factors that might or not be present in each context of application. In our terms, treatments do not share enough traits to make rule-based decisions efficient: each treatment used on a given population should be considered separately.

For Cartwright, regulatory authorities should find evidence about the causal factors that will contribute to the efficacy of the intervention on the general population of patients. Cartwright suggests collecting all the facts that are relevant to the transition from efficacy in the trial to effectiveness on other populations. For this search there is no algorithmic method, but rather a number of heuristics that regulators should use wisely. For example, they can conduct a pre-mortem analysis, a thought experiment of the form “If the intervention goes wrong, how will it have gone wrong?” (Cartwright and Hardie 2015). This will provide some clues as to the contributing factors that may spoil the effect of a treatment outside the trial. Ultimately these heuristics cannot be synthesized in a simple rule with a hard-empirical trigger. Drug regulation should therefore use *standards* with soft evaluative triggers: *if the evidence presented is good enough, grant market approval*. Deliberation about evidence is necessary to make this decision.

These three approaches exemplify in various ways our epistemology-first thesis: they all share the assumption that going for the best evidence (in quantity and quality) will deliver a better *error rate*, no matter how it complicates the decision process, and whatever the interest or stakeholders at play. Regulators using these modified rules or standards will fail less often in deciding about new treatments, as compared to their current rate of mistakes under the RCT-based rule: fewer dangerous compounds will be erroneously accepted and more effective therapies wrongly rejected.

Yet, both the EBM+ and Osimani's approach are essentially reformist: the evidential pluralism they advocate is compatible with a rule-based system, except that the trigger in the conditional is more complicated to pull, increasing the costs of the decision. Cartwright is the one that more radically changes the current approach to regulatory issues. She is advocating for considering causal interventions on a case-by-case basis, because this is the dappled nature of our world. There are no causal laws governing patients' populations independently of the context in which the treatments are administered and there is no single best method to study treatment effects. This sort of epistemic pluralism requires standard-based decisions reached through deliberation. Here is where the epistemology first thesis reaches its Platonic apex: evidence alone will make those committees agree. Yet, we wonder, what are the costs of this adjudication process? Will the benefits of standard-based decisions offset them? As we will argue in the following section, a pragmatic approach to these questions requires us to take further costs into consideration.

4. THE COSTS OF CONFLICTS OF INTERESTS

In section 2 we showed how the 1962 FDA Act created a regulatory system in which most decisions were made on the basis of an RCT-based rule, resorting to standards-based adjudication in case of disagreement on the evidence. This system can be justified in terms of its efficiency in selecting the safe and effective treatments with the fewer costs. Yet, as we just saw in the previous section, if the epistemic adequacy of the hard trigger is challenged, there is room for alternative decision methods that, on paper at least, might yield better decisions. Yet, for Robert Temple, the legendary FDA officer, adopting the most radical of this proposals (a standard-guided case-by-case approach) was "his idea of a nightmare" (Berry, Goodman, and Louis 2005).

Is Temple's regulatory nightmare worth having? We will argue against Cartwright's standard-based approach on two different accounts. On the one hand, we will show that, as of today, there is no evidence that standard-based deliberation in FDA committees yields better decisions. A quick rejoinder is that at the FDA there is much industry pressure distorting deliberation, so we cannot see regulatory standards operating at their best. In response, we will argue that, in assessing the adequacy of rules and standards for pharmaceutical regulation, we need to consider the costs of protecting rule- or standard-based decisions from conflicting interests.

The first part of our argument is relatively straightforward. If Cartwright is correct, we may expect deliberative committees to perform better than rule-based review boards in detecting unsafe or inefficient drugs. But deliberation per se does not seem to yield very different decisions. A study conducted by the National Research Center for Women & Families provides a quantitative analysis of the decision of FDA advisory committee (Zuckerman 2006b), following their voting patterns.

One of the most interesting findings is that committee members voted unanimously for 66% of the drugs they recommended for the approval, which is particularly surprising given that advisory committees are convened for controversial cases in which the available RCT evidence does not seem particularly persuasive. Moreover, sometimes the FDA has approved drugs for which the advisory committees recommended the opposite. In Zukerman's analysis, between 1998 and 2005 advisory committees voted against 11 drugs, but the FDA subsequently approved 4 (36%) of them. Whereas, according to another report (Smith et al. 2012), the FDA decisions are very consistent with the advisory committee voting: in a set of 63 FDA advisory committee meetings that included votes for or against the approval of a new drug between 2001 and 2010, only 2 times did the FDA approve a drug despite the committee voting.

According to Zuckerman (2006) "many of today's FDA drug and device advisory committees are rubber stamps for approval almost every time they meet." We have no statistics about withdrawals and warnings for committee decisions, but there have been significant scandals on both accounts. For example, there were unanimous votes for the approval of Celebrex in 1998 and Vioxx in 1999, two drugs that subsequently were found to significantly increase the risk of heart attack and stroke. Vioxx was subsequently withdrawn, giving rise to a public scandal, while Celebrex remained on the market, but with strong warnings.

In other words, despite its superior adjudication costs, standard-based deliberation on its own may lead to the same mistakes as rule-based approvals do. Yet, it may be argued that we do not see epistemic deliberation working at its best in today's FDA committees. The pharmaceutical industry imposes too much pressure on regulatory bodies to let them operate properly. Too often committee members have some financial ties with one or more pharmaceutical company. A recent study analyzing the voting pattern from 379 CDER meetings during the 15-year period 1997-2011, (Pham-Kanter 2014) showed a strong pro-sponsor bias among the committee members who have financial relationships with firms. Interestingly, this pro-sponsor bias appears to be larger when the scientific evidence is more ambiguous, and therefore so in all those cases in which a deliberative judgment would be necessary to make a good decision.

For the industry, regulatory decisions are high-stakes bets: developing a drug to the point of submitting a NDA to the FDA is expensive; it costs an average of \$2.6 billion merely to get a drug through the FDA approval process (Mullard 2014). Thereby arises the temptation for the industry to "capture" the regulator and find ways to make the FDA decide according to the industry interest, even if it collides with the interests of the public (Carpenter 2013). By the same token, it is obvious that we are not seeing RCT-based decisions grounded on the best possible evidence (Gøtzsche,

Smith, and Rennie 2013). Bero and Rennie (1996) have discussed how industry-sponsored trials might favor the outcomes of a trials. Two quick examples are the choice of control and outcome measures. As to the former, researchers may, for instance, test an experimental drug against a placebo when an efficacious therapy is already available, or use a dosing in the control group that favors the experimental drug (Rochon et al. 1994). As to the latter, surrogate endpoints are nowadays extensively used to accelerate drug development and save money, but their validity is questionable (Aronson 2005; Pereira, Horwitz, and Ioannidis 2012). Pharmaceutical companies sponsoring trials can make many methodological choices that deviate from the ideal study design for RCTs, and which *a priori* can favor the experimental drug.

In other words, we must consider more than the operation of hard and soft epistemic triggers in a frictionless manner (Holman 2017). Regulatory agencies operate under conflicting pressures from various actors (not just the industry). Therefore, in assessing the costs of epistemic rules or standards for regulatory decision making, we must also factor in how resistant they are to external pressure. If we assume, for the sake of argument, that legislators are capable of resisting such pressures and promulgate epistemically sound rules or standards, how likely is it that regulatory agencies can apply them properly? Rule-based systems are expected to reduce corruption to the extent that they limit discretion and thereby minimize abuses of power (Schafer 2006). Standard-based adjudication, instead, is very sensitive to external pressures, for example, ideological decision making (Fon and Parisi 2007). Defending regulatory deliberation without any caveats about external interferences, *pace* Cartwright, seems too Platonic in this context. Yet, we fear there is no cheap solution to protect committees from conflicts of interests (Intemann and de Melo-Martín 2014).

Justin Biddle, a philosopher of science, has presented what is in our view the most compelling proposal for a standard-based regulatory system. Drawing on Kantrowitz (Kantrowitz 1967, 1976) and Merton (Merton 1973), Biddle has argued for the introduction of an adversarial system in regulatory committees, in which the interests in conflict are explicitly argued for: “Two groups of advocates would present arguments for a specific position, and a panel of judges would adjudicate between these two groups” (Biddle 2013) The industry chooses its own advocates, whereas the other set of advocates can be chosen among industry competitors (insurance companies, public health care agencies, patients and citizen groups). Judges must be scientific experts independent from both parties, having no direct interest or connections to the issue under consideration. The advocates and the judges should be, of course, kept separate. To control for independence, both

groups of advocates should have the possibility to exclude a given number of scientists from the panel of judges. In this way, the decision should be as impartial and objective as possible.

Biddle's proposal captures some of the best arguments for standard-based decisions provided by legal and economics scholars. Invest a limited amount in creating a body of expert judges without conflicts of interest and let the advocates absorb the costs of finding the right evidence for their interpretation of the standard. Under certain idealized assumptions, this may be the most efficient approach (Friedman and Wickelgren 2013). The party that promotes the most efficient outcome will have more to gain by a positive decision, and therefore has more incentives to invest resources in supporting its case. Yet, for this sorting process to work, there should be a certain symmetry between the resources of the conflicting parties, otherwise the quality of legal representation may simply outweigh the merits of the case. Most patients' organizations are poor as compared to pharmaceutical companies, if these latter are not their primary source of funding.⁸ To illustrate this, consider that in the notorious Vioxx scandal, Merck set up a \$4.85 billion settlement fund just to resolve consumer claims, while in 2017 the total revenue of one of the largest patient advocacy organizations, the American Heart Association, was just \$865 million (Forbes June 30, 2017).

Summing up, the operation of rules and standards in pharmaceutical regulation may be distorted by industry pressure. Assuming a good epistemic trigger in both, rules are less discretionary and, therefore, more difficult to game than standards. If our discussion of Biddle's proposal is correct, standard-base deliberation may be also very expensive to protect from conflicting interests.

5. AN EXPERIMENTAL APPROACH TO REGULATORY DECISION-MAKING

Let us take stock of the argument so far. We have examined the proposals of different philosophers of science about how to improve our pharmaceutical regulatory system from an informal law and economics standpoint. Even if regulatory rules and standards hinge on epistemic triggers, having the best epistemic trigger does not necessarily imply that one will have the best regulatory system. Standard-based deliberation is, in itself, costly, as compared to a rule-based system, and, a priori, it is more difficult to protect against external pressures.

That does not imply that our current regulatory system works easily and well. According to Gøtzsche, Smith, and Rennie (2013), the clinical documentation for just three RCT-based NDA can take up 70 linear meters of binders. So maybe the ideally correct decision is out of our reach with either rules or standards. The question, of course, is whether a good enough decision is reachable. So far, we have just discussed costs, but rules and standards should also be judged in terms of the benefits they yield. In this final section we want to explore the possibility that, for the amount of

resources the FDA has, patients are reasonably well protected with RCT-based rules. Or, alternatively, it is not self-evident that with a similar amount of resources standard-based decisions will protect patients better. We need to find an empirical index that allows us to judge which decision procedure is more suitable for protecting patients.

We claim that market withdrawals are a rough indicator of the error rate of RCT-based rules. Market withdrawals in the USA, as in many other countries, follow a standardized procedure: if there is reported evidence suggesting a reasonable risk of death or other serious adverse reactions from use of a medicinal product, the FDA may ask the company authorized to market it to conduct a recall.⁹ The FDA will then study how the drug's benefit and risk balance compares with treatment alternatives. If the risks outweigh the benefits, the drug may be finally withdrawn from the market. Given that the FDA mandate is to test drugs for safety and efficacy, drug withdrawals are *prima facie* explicit regulatory mistakes: drugs that should have been banned have been allowed into pharmaceutical markets. Number of drug withdrawals is admittedly a rough index of regulatory success. For instance, there are no universal guidelines and therefore there is no perfect international agreement about which drugs should be available (Aronson 2017). With every caveat in mind, less than 2% of new drug approvals by the FDA between 1950 and 2011 were withdrawn (Onakpoya, Heneghan, and Aronson 2016). Following the pioneering review of Onakpoya, Heneghan, and Aronson (2016), in absolute figures, 134 drugs were withdrawn in the USA between 1953 and 2013 for serious adverse reactions. About 50 of them were linked to death (Onakpoya, Heneghan, and Aronson 2015), often correctly (Onakpoya, Heneghan, and Aronson 2017).¹⁰

For the sake of the argument, let us now take these figures at face value. Our claim is that market withdrawals provide an *empirical benchmark* of how well the FDA can meet its regulatory mandate on a rule-based system, with its current institutional arrangements. Critics of the RCT-rule should discuss under which circumstances a different rule or standard regulatory system can overperform a 2% error-rate. In our view, it is not enough to claim that current trial system is broken. Perhaps it is, but we need to know how much it would cost to achieve a better result with different sources of evidence and under similar external pressure, before deciding whether it is worth it to invest in making it happen. The point is that, once costs are taken into account, epistemic arguments alone are not enough to defend the necessity of an alternative. If regulatory agencies can be persuaded to run a pilot program for drug assessment with different decision methods, perhaps we will be able to find out empirically which system is better.

An immediate rejoinder would be to challenge the reliability of the submitted 2% FDA error rate. For many, the real number of dangerous compounds in the market would be much higher, because

the FDA is not properly complying with its institutional mission. Technically speaking, the FDA would have been *corrosively captured* (Carpenter 2013): the pharmaceutical industry would have succeeded both at obtaining undeserved RCT-based approvals and keeping those defective treatments in the market. This is a popular vision of the pharmaceutical industry widely advocated by specialized journalists (Angell 2004), epidemiologists (Gøtzsche, Smith, and Rennie 2013) and STS scholars (Abraham and Davis 2013). Their arguments are usually based on case reports about regulatory misdeeds or statistical estimates. For instance, according to Gøtzsche, Smith, and Rennie (2013), by 2007, 20 million consumers had taken Zyprexa, an anti-psychotic drug for which the FDA had issued a warning already in 2003. Meta-analyses would have shown that for every 100 patients treated in a trial, there was a death from the drug. Therefore, taking into account the total number of consumers, 200,000 patients may have died for a drug that, in Gøtzsche's view, should not have been in the market in the first place.

Following Carpenter (2013), we disagree with the adequacy of any of these methods to properly analyze regulatory capture, but, for the sake of the argument, let us assume that the FDA has been indeed captured by pharmaceutical interests. The implication for our own argument is that no matter how good the rule or the standard, the amount of pressure the industry exerts today can bend it. Our entire analysis so far would be pointless. According to critics like Gøtzsche, the problem originates more in the vulnerability of people than on the ideal properties of methods; after all, Gøtzsche is a fierce advocate of Evidence-Based Medicine. The solution is thus to transform the regulatory environment, nationalizing pharmaceutical research, so that no market pressure will interfere with drug assessment.

The costs, and not just the benefits, of this solution are of a different order of magnitude: is it worth it to dispense with pharmaceutical markets just so we can have better drugs? Whereas for any rule- or standard-based regulatory scheme there is at least the possibility of a scale test, we have very little evidence to judge what would be the outcome of restructuring pharmaceutical markets so dramatically. Our only experience of a completely public pharmaceutical industry is the USSR and it was simply a failure in terms of innovation as measured by drug patents (Jack and Mason 1987).

A less radical but still costly alternative would be to nationalize the regulatory tests: conduct trials (or whatever trigger our regulatory system requires) funded by public budgets and designed by public officers (Abraham and Davis 2013). Costly as it may be, we find this a defensible solution, since it would allow us to decide between rules or standards on the basis of the best available evidence.¹¹ Yet, this proposal would probably be objectionable to the most radical critics: if the

FDA, with all its institutional power, has been captured, how could any public testing agency resist pharmaceutical pressure?

Capture arguments are themselves corrosive and should be handled with care. We concur with Carpenter (2013) about the necessity of a clear definition of the pharmaceutical public interest as the starting point of any discussion. In particular, we need to define what levels of risk are acceptable for authorizing new drugs. As we saw in section 3, philosophers of science have argued for increasing safety levels, using methods that would allow regulators to capture adverse events beyond the reach of standard RCTs. Desirable as it is, there might be a limit to this demand. According to Edwards (2005) there are very few drugs in which the overall risk really is greater than the effectiveness. Given that zero risk may be a non-attainable ideal, Edwards suggests focusing more on how patients are actually treated rather than whether a drug is on the market or not. Patients should be asked what risks they are prepared to take and for what benefits. This is a normative debate that goes beyond our rules versus standards dilemma, but any alternative regulatory scheme should address it (Tempini and Teira 2019).

5. CONCLUSION

The distinction between rules and standards allows us to capture the costs at stake in the philosophical controversies on regulatory evidence. There are, indeed, good a priori arguments about the superiority of standards of evidence in the detection of safety and efficacy. And there are equally good a priori proposals on how to organize impartial committees to carry out such deliberation. Framing the decision in terms of the costs of implementing these ideals allows us to grasp the actual trade-offs that regulatory authorities should make: there is only a limited amount of resources for them to use. It is possible that RCT-based rules have offered, so far, a cost-effective way of protecting patients, if their safety is defined in terms of market withdrawals. In any case, our regulatory experience provides an empirical benchmark for other alternatives. We suggest that we need instead to test these alternatives in a pilot committee. As it has often happened in the past, philosophy has been good at setting up ideals, but philosophers of science may find that reality does not live up to them.

ACKNOWLEDGEMENTS

Teira was funded by the Spanish Ministry of Economics research grant FFI2014-57258-P.

P. Luque, J. S. Mora, J. Aronson and B. Clarke provided crucial references. We are grateful for the comments received from audiences at the PSA18. The EBM+ Conference (UCL, May 2018) was equally helpful in revising the paper.

AUTHOR BIOGRAPHIES

David Teira teaches philosophy of science at UNED (Madrid). His personal webpage is: <http://www.uned.es/personal/dteira>

Mattia Andreoletti is a postdoctoral fellow at the University of Turin. He holds a PhD from the FOLSATEC program at the European School of Molecular Medicine (SEMM), affiliated with the European Institute of Oncology (IEO, Milan)

REFERENCES

- Abraham, John, and Courtney Davis. 2013. *Unhealthy Pharmaceutical Regulation : Innovation, Politics and Promissory Science*. Health, Technology and Society. Houndmills, Basingstoke, Hampshire; New York, NY: Palgrave Macmillan.
- Angell, Marcia. 2004. *The Truth about the Drug Companies: How They Deceive Us and What to Do about It*. 1st ed. New York: Random House.
- Aronson, J. K. 2005. "Biomarkers and Surrogate Endpoints." *British Journal of Clinical Pharmacology*, 2005.
- . 2017. "Post-Marketing Drug Withdrawals: Pharmacovigilance Success, Regulatory Problems." *Therapie* 72 (5): 555–61.
- Bero, L A, and D Rennie. 1996. "Influences on the Quality of Published Drug Studies." *International Journal of Technology Assessment in Health Care* 12 (2): 209–37.
- Berry, Donald A., Steven N. Goodman, and Thomas A. Louis. 2005. "Floor Discussion." *Clinical Trials* 2 (4): 301–304.
- Biddle, Justin. 2013. "Institutionalizing Dissent: A Proposal for an Adversarial System of Pharmaceutical Research." *Kennedy Institute of Ethics Journal* 23 (4): 325–53.
- Boldrin, Michele, and David K. Levine. 2008. *Against Intellectual Monopoly*. New York: Cambridge University Press.
- Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2006. "Regulatory Objectivity and the Generation and Management of Evidence in Medicine." *Social Science & Medicine* 63 (1): 189–99.
- . 2009. "Biomedical Conventions and Regulatory Objectivity: A Few Introductory Remarks." *Social Studies of Science* 39 (5): 651–64.
- Carpenter, Daniel. 2013. "Corrosive Capture? The Dueling Forces of Autonomy and Industry Influence in FDA Pharmaceutical Regulation." In *Preventing Regulatory Capture: Special Interest Influence and How to Limit It*, edited by Daniel Carpenter and David A. Moss, 152–72. Cambridge: Cambridge University Press.
- Carpenter, Daniel. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton Studies in American Politics. Princeton: Princeton University Press.
- Cartwright, Nancy, and Jeremy Hardie. 2015. *Evidence-Based Policy. A Practical Guide to Doing It Better. Statewide Agricultural Land Use Baseline 2015*. Vol. 1.
- Center for Drug Evaluation and Research. n.d. "Investigational New Drug (IND) Application - Drug Development and Review Definitions." WebContent. Accessed 19 December 2016. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/InvestigationalNewDrugINDApplication/ucm176522.htm>.
- Clarke, Brendan, Donald Gillies, Phyllis Illari, Federica Russo, and Jon Williamson. 2014. "Mechanisms and the Evidence Hierarchy." *Topoi* 33 (2): 339–60.
- Edwards, I Ralph. 2005. "What Are the Real Lessons from Vioxx®?" *Drug Safety* 28 (8): 651–58.

- Fon, Vincy, and Francesco Parisi. 2007. "On the Optimal Specificity of Legal Rules." *Journal of Institutional Economics* 3 (2): 147–64.
- Friedman, Ezra, and Abraham L. Wickelgren. 2013. "A New Angle on Rules versus Standards." *American Law and Economics Review* 16 (2): 499–549.
- Gøtzsche, Peter C., Richard Smith, and Drummond Rennie. 2013. *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. London: Radcliffe Publishing.
- Holman, Bennett. 2017. "Philosophers on Drugs." *Synthese*, 1–28.
- Intemann, Kristen, and Inmaculada de Melo-Martín. 2014. "Addressing Problems in Profit-Driven Research: How Can Feminist Conceptions of Objectivity Help?" *European Journal for Philosophy of Science* 4 (2): 135–51.
- Jack, D. B., and N. P. Mason. 1987. "THE PHARMACEUTICAL INDUSTRY IN THE U.S.S.R." *Journal of Clinical Pharmacy and Therapeutics* 12 (6): 401–7.
- Kantrowitz, Arthur. 1967. "Proposal for an Institution for Scientific Judgment." *Science* 156 (3776).
 ———. 1976. "The Science Court Experiment: An Interim Report." *Science* 193 (20 August 1976): 653–56.
- Kaplow, Louis. 1992. "Rules versus Standards: An Economic Analysis." *Duke Law Journal* 42: 557–629.
- Landes, Jürgen, Barbara Osimani, and Roland Poellinger. 2018. "Epistemology of Causal Inference in Pharmacology." *European Journal for Philosophy of Science* 8 (1): 3–49.
- Lupkin, S., and S. Findlay. 2017. "Grab Bag Of Goodies In 21st Century Cures Act." Kaiser Health News. 26 August 2017.
- Luppi, Barbara, and Francesco Parisi. 2011. "Rules versus Standards." In *Encyclopedia of Law and Economics*, edited by Gerrit De Geest. Vol. 3. Cheltenham, UK: Edward Elgar.
- Merton, Robert King. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.
- Mullard, Asher. 2014. "New Drugs Cost US\$2.6 Billion to Develop." *Nature Reviews Drug Discovery* 13 (12): 877–877.
- Onakpoya, I. J., C. J. Heneghan, and J. K. Aronson. 2015. "Delays in the Post-Marketing Withdrawal of Drugs to Which Deaths Have Been Attributed: A Systematic Investigation and Analysis." *BMC Med* 13 (February): 26.
- . 2016. "Post-Marketing Withdrawal of 462 Medicinal Products Because of Adverse Drug Reactions: A Systematic Review of the World Literature." *BMC Med* 14 (February): 10.
- . 2017. "Post-Marketing Regulation of Medicines Withdrawn from the Market Because of Drug-Attributed Deaths: An Analysis of Justification." *Drug Saf* 40 (5): 431–41.
- Parkkinen, Veli-Pekka, Christian Wallmann, Michael Wilde, Jon Williamson, Brendan Clarke, Phyllis Illari, Mike Kelly, and Federica Russo. 2018. "Evaluating Evidence of Mechanisms in Medicine: A Handbook for Practitioners." Springer.
- Pereira, Tiago V, Ralph I Horwitz, and John P A Ioannidis. 2012. "Empirical Evaluation of Very Large Treatment Effects of Medical Interventions." *Journal of the American Medical Association* 308 (16): 1676–84.
- Pham-Kanter, Genevieve. 2014. "Revisiting Financial Conflicts of Interest in FDA Advisory Committees." *Milbank Quarterly* 92 (3): 446–70.
- Rochon, Paula A, Jerry H Gurwitz, Robert W Simms, Paul R Fortin, David T Felson, Kenneth L Minaker, and Thomas C Chalmers. 1994. "A Study of Manufacturer-Supported Trials of Nonsteroidal Anti-Inflammatory Drugs in the Treatment of Arthritis." *Archives of Internal Medicine* 154: 157–63.
- Schafer, Hans-Bernd. 2006. "Rules versus Standards in Rich and Poor Countries: Precise Legal Norms as Substitutes for Human Capital in Low-Income Countries." *Sup. Ct. Econ. Rev.* 14: 113.
- Schauer, Frederick. 1993. *Playing by the Rules*. Clarendon Press Publication.
- Schlag, Pierre J. 1985. "Rules and Standards." *UCLA Law Review* 33 (2): 379–430.

- Sismondo, Sergio. 2010. *An Introduction to Science and Technology Studies*. 2nd ed. Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.
- Sismondo, Sergio, and Jeremy A. Greene. 2015. *The Pharmaceutical Studies Reader*. Chichester, UK; Malden, MA: John Wiley & Sons Inc.
- Smith, Jeffrey F, Seth A Townsend, Navjot Singh, and Philip Ma. 2012. “FDA Advisory Committee Meeting Outcomes.” *Nat Rev Drug Discov* 11 (7): 513–14.
- Solomon, Miriam, Jeremy R. Simon, and Harold Kincaid. 2017. *The Routledge Companion to Philosophy of Medicine*. Routledge Philosophy Companions. New York: Routledge, Taylor & Francis Group.
- Tempini N. and Teira D. 2019. Is the genie out of the bottle? Digital platforms and the future of clinical trials. *Economy & Society*, 48 (1)
- Tempini N. and Teira D, forthcoming, The Babel of Drugs: on the challenges of data journeys and evidential pluralism in pharmaceutical regulation, in Leonelli S. and Tempini N., eds., *Varieties of data journeys*, MIT Press
- Urfalino, Philippe. 2012. “Reasons and Preferences in Medicine Evaluation Committees.” *Collective Wisdom: Principles and Mechanisms*, 173–202.
- Zuckerman, Diana M. 2006. “FDA Advisory Committees: Does Approval Mean Safety.” *Washington: National Research Center for Women & Families*.

¹ See Solomon, Simon, and Kincaid (2017) for a survey.

² Our discussion will take the regulatory experience of the FDA in the United States as its reference point, for two reasons: the FDA has set a paradigm for many other regulatory agencies worldwide and it is, by far, the best studied. Our understanding of the FDA is mainly based on Carpenter (2010). For recent regulatory developments in the US, see footnote 4 below.

³ We follow here the informal way in which this literature discusses costs: we are drawing on the mainstream view, without any original contribution.

⁴ For recent statistics on submitted NDAs, see

<https://www.accessdata.fda.gov/scripts/fdatrack/view/track.cfm?program=cder&status=public&id=CDER-RRDS-Number-of-NDA-and-BLAs-submitted-and-approved&fy=All> (accessed on July 7, 2018)

⁵ In December 2016, the US Congress passed the 21st Century Cures Act (21CCA), which invites the FDA to incorporate new evidentiary standards for drug assessment (e.g., Bayesian trials). For the last two years, the FDA has been organizing workshops in which experts debate how to implement this mandate. The 21CCA signals the end of a five decades consensus on how to test treatments. It may be understood as a response to changes in the regulatory environment: e.g., the emergence of drugs targeting particular genetic profiles for which it is difficult to gather enough patients to run a conventional trial. When such changes occur, rules lose efficiency for decision making (Fon and Parisi 2007). Yet, more skeptical voices see it as a success of pharmaceutical interests. For our own (also skeptical) view, see (Tempini and Teira, forthcoming).

⁶ It is interesting to notice, in contrast, how STS-oriented pharmaceutical studies have paid very little attention to regulatory decision processes—see, e.g., the pioneering reader by (Sismondo 2010; Sismondo and Greene 2015; Cambrosio et al. 2006, 2009) for instance. They have highlighted the role of regulatory agencies, among other stakeholders, in the setup of the biomedical conventions about evidence discussed in this section, but they carefully avoid any normative discussion as to how their decision processes could be improved with alternative methods. Abraham and Davis (2013) provide an excellent, but less STS-oriented, sociological

approach to pharmaceutical regulation. They explain how industry pressure has captured regulatory agencies through what they call “neoliberal deregulation.” Following Carpenter (2010), as we will argue in section 5 below, we are not so pessimist as to the power of the post-1962 FDA to resist industry pressure, thanks, in part, to the RCT-based rules. Yet, as we explain in footnote 10 below, even if we disagree on the diagnosis, we agree with the solutions advocated in Abraham and Davis (2013).

⁷ EBM stands for Evidence-Based Medicine, a medical paradigm that affirms the primacy of RCTs in assessing treatments. The + sign in EBM+ indicates that trials should be complemented with evidence about the causal mechanisms underpinning medical treatments. See the EBM+ webpage for further information: <http://ebmplus.org/> (visited on July 5, 2018)

⁸ For instance, the lobbying expenses in the 21st Century Cures Act declared by pharmaceutical, device and biotech companies amounted \$192 million; patient groups reported spending \$6.4 million in disclosures that named the bill as one of their issues (Lupkin and Findlay 2017).

⁹ United States Food and Drug Administration. Regulatory Procedures Manual, May 2018, Chapter 7 Recall Procedures. (Accessed on July 9, 2018): <http://www.fda.gov/downloads/iceci/compliancemanuals/regulatoryProceduresManual/UCM074312.pdf>.

¹⁰ Our success index takes into account only what we take to be the primary mission of the pharmaceutical regulatory agencies, the prevention of harm, rather than the promotion of pharmaceutical innovation. This latter is, in itself, a contentious issue that we cannot discuss here. Our views on the innovation crisis are best captured in (Boldrin and Levine 2008).

¹¹ We should emphasize that we agree with both Gøtzsche, Smith, and Rennie (2013) and Abraham and Davis (2013) on the necessity of strengthening regulatory trials (better comparisons, stronger endpoints, etc.) and protecting them better from industry influence. It is interesting that critics, so to speak, “on the pharmaceutical left” generally defend a regulatory system based on stronger rules. We may reasonably assume that a standard-based approach would easily qualify for them as “neoliberal deregulation.”