Synthese (2024) 204:33
https://doi.org/10.1007/s11229-024-04677-x

ORIGINAL RESEARCH

The reliability of evidential pluralism in drug regulation

Mattia Andreoletti[1] · **Saúl Pérez-González**[2]

Received: 7 September 2023 / Accepted: 12 June 2024 / Published online: 9 July 2024

Abstract

The aim of this paper is to shed light on and critically evaluate the potential impact of evidential pluralism in the realm of pharmaceutical regulation. In the complex landscape of drug evaluation and approval, the role of evidence is pivotal. Firstly, we delve into the role of evidence of mechanisms within drug regulation, with a particular focus on the Accelerated Approval pathway administered by the US Food and Drug Administration. Our analysis reveals that this program, in practice, closely resembles a pluralistic approach to evidence. Secondly, we undertake a thorough examination of the reliability of evidential pluralism within the context of drug regulation. We dissect and discuss the open-cases approach recently put forth by Sung and Holman. While acknowledging its potential merits, we also identify and articulate significant limitations inherent in this approach. Finally, we propose an alternative empirical approach that centres on the real-world outcomes of regulatory programs.

Keywords Drug regulation · Accelerated approval · Evidential pluralism · Evidence · Mechanisms

✉ Mattia Andreoletti
mattia.andreoletti@hest.ethz.ch

Saúl Pérez-González
saul.perez@uv.es

[1] Department of Health Sciences and Technolgy, ETH Zurich, Zurich, Switzerland

[2] Department of Philosophy, University of Valencia, Valencia, Spain

1 Evidential pluralism in medicine

In the last decades, philosophers of science have engaged in extensive discussions regarding the use of scientific evidence in the process of making policy decisions. These debates have often taken an epistemological perspective, deliberating on the types of evidence that can more effectively bolster and validate policy choices. Within the domain of the philosophy of medicine literature, the emphasis has been on the Evidence-Based Medicine (EBM) paradigm.

EBM, as a distinct epistemic framework, was introduced in the early 1990s. Its central tenet is the conscientious, explicit, and judicious use of current best evidence in clinical practice and decision-making (Sackett et al., 1996). As such, EBM de-emphasises expert intuition and mechanism-based reasoning. In this framework, available evidence is assessed on the basis of hierarchies of evidence that rank various methods of evidence gathering. These hierarchies privilege evidence from comparative clinical studies. Among those studies, randomised controlled trials (RCTs) are the best ranked. RCTs are experiments in which members of the study population are randomly assigned to the experimental or the control group. In this kind of trial, an experimental treatment is tested against either the standard of care or a placebo. Subsequently, statistical tools, such as p-values and confidence intervals, are used to make inferences about the observed differences. Roughly speaking, EBM tends to privilege evidence of correlations between the effect and the postulated cause (Howick, 2011b). This approach focuses on observing and measuring the (putative) effects of the considered interventions. While EBM has become very popular, some concerns have been raised about evidence of correlations provided by RCTs and other comparative studies (e.g., Deaton & Cartwright, 2018; Worrall, 2007).

In response to concerns about EBM, some authors have stressed the relevance of evidence of mechanisms and advocated a pluralist approach to evidence in medicine (Anjum et al., 2020; Clarke et al., 2014; Gillies, 2018; Parkkinen et al., 2018). According to this approach, both evidence of correlations and evidence of mechanisms are crucial for establishing causal claims in the biomedical sciences. The pluralist approach to evidence in medicine is summarised in the well-known Russo-Williamson Thesis (Russo & Williamson, 2007). This epistemological thesis claims that "[t]o establish causal claims, scientists need the mutual support of mechanisms and dependencies" (Russo & Williamson, 2007, p. 159). In the pluralist framework, evidence of mechanisms is usually understood as evidence about the existence or the properties of mechanisms in the domain of inquiry (Illari, 2011) —biochemical pathways are paradigmatic examples of this kind of evidence. There is no specific method through which evidence of mechanisms must be gathered. It can be obtained by means of various methods such as in vitro studies, in vivo studies in animals, case studies, and even agent-based models.

In support of pluralism, it is generally argued that both statistical evidence and evidence of mechanisms have strengths and weaknesses, and that they complement each other (see Clarke et al., 2014; Gillies, 2018; Williamson, 2019). For example, Gillies (2018) refers to this complementarity as "strength through combining". Statistical evidence faces important difficulties related to confounding factors and non-causal correlations. When a correlation between the variables A and B is established,

it may still be the case that A does not cause B. Evidence of mechanisms can help to overcome those difficulties. Evidence of the mechanisms through which A causes B supports the causal interpretation of the correlation. Nonetheless, evidence of mechanisms also comes with caveats. Even when evidence of one or more mechanisms linking A and B is obtained, it may still be the case that there are other unknown mechanisms through which A influences B. Statistical evidence is crucial for addressing such a problem: evidence of the correlation between A and B provides information about the overall effect of the operating mechanisms and reveals whether the identified mechanisms are counteracted.

Furthermore, in support of evidential pluralism, several cases in which evidence of mechanisms played (or could have played) an important role have been analysed (Abdin et al., 2019; Auker-Howlett & Wilde, 2020; Gillies, 2011). For example, Gillies (2011) shows that both evidence of correlation and evidence of an underlying mechanism were crucial for establishing a causal relationship between smoking and heart disease.

However, there have also been several criticisms raised against evidential pluralism. Critics argue, for example, that biomedical mechanisms are difficult to identify and our knowledge about them is often fragmentary (Howick, 2011a; Howick et al., 2013). Moreover, it has been shown that mechanisms may not behave regularly, and their behavior may change outside the tightly controlled laboratory circumstances, making it difficult to anticipate their behavior under intervention (Howick, 2011a; Howick et al., 2013). Furthermore, in certain scenarios, complementary evidence of mechanisms may be redundant, and demanding it may result in unnecessary delays (Andersen & Kjær, 2019). Regarding pluralist regulatory frameworks, it has also been argued that they have higher associated costs and are more exploitable by pharmaceutical companies than the current RCT-based system (Andreoletti & Teira, 2019). Overall, while evidential pluralism has its defenders, its critics raise important questions about how to use multiple types of evidence in a rigorous and transparent way.

In recent years, the application of evidential pluralism in drug development, evaluation, and approval has been a topic of significant discussion (Aronson et al., 2018; Parkkinen & Williamson, 2020; Rocca, 2016). In a seminal paper, Rocca (2016) has explored the potential contribution of evidence of mechanisms in post-marketing risk assessment of drugs. In pharmacovigilance, available evidence usually consists of case series in peer-reviewed literature, single-case reports, and spontaneous reports from healthcare professionals and patients. Rocca argues that to establish a causal relation between an untargeted symptom and a drug it is important to formulate (and provide evidence in support of) a mechanism hypothesis about *how* the drug could have provoked the symptom. Aronson et al. (2018) have analysed the diverse roles that evidence of mechanisms plays (or could play) in drug approval. It is argued to be crucial for tasks such as testing the stability of drug formulations, establishing the efficacy of treatments, and assessing the external validity of trials. In fact, the authors claim that "evidence of mechanisms is relevant for all the tasks that are important in drug approval" (Aronson et al., 2018, p. 1167). Parkkinen and Williamson (2020) have discussed the relevance of evidence of mechanisms in the extrapolation of causal claims from model organisms in pharmacology. They argue that evidence of

mechanisms is necessary for establishing both causality in the model organism and similarity of the model organism to humans.

The aim of this paper is to illuminate and assess the potential contribution of evidential pluralism in drug regulation. First, we want to clarify the role that evidence of mechanisms can play in regulatory decision-making. For that purpose, we focus on the Accelerated Approval (AA) pathway implemented by the US Food and Drug Administration (FDA). We argue that the AA already adopts a pluralist approach towards evidence, taking into account both statistical evidence and evidence of mechanisms. Second, we aim to address the reliability of evidential pluralism in drug regulation. We analyse and discuss a recent proposal developed by Sung and Holman (2023) to assess evidential pluralism in such a context. Then, after identifying important limitations of Sung and Holman's proposal, we introduce an alternative empirical framework that focuses on the outcomes of regulatory programs.

The paper is structured as follows. In Sect. 2, we introduce drug regulation and examine the AA program. We contend that AA embraces a pluralistic approach towards evidence and analyse the roles played by evidence of mechanisms. In Sect. 3, we turn our attention to the reliability of evidential pluralism in drug regulation and introduce an empirical (or naturalistic) approach. Within this context, we explore and discuss the open-cases framework proposed by Sung and Holman. We also extensively examine the case study presented in support of the proposal, i.e., the accelerated approval of Aducanumab. Subsequently, in Sect. 4, we propose an alternative empirical approach that focuses on the outcomes of regulatory programs to assess evidential pluralism. We consider and address diverse potential objections that may arise. Finally, in Sect. 5, we make some concluding remarks.

2 Accelerated drug approval

Regulatory agencies such as the FDA play a pivotal role in safeguarding patients against potentially harmful and ineffective treatments. The widely accepted societal perspective on the FDA's gatekeeping authority is rooted in a paternalistic rationale centred on collective risk aversion. The idea is to establish a regulatory body with gatekeeping powers to shield the market from hazardous compounds, thereby preventing pharmaceutical disasters arising from faulty drugs (Andreoletti, 2021; Teira, 2020). Notably, the FDA stands out as the largest and most influential among these agencies, overseeing a market of products valued at over 1 trillion dollars—equivalent to nearly a quarter of consumer spending. Its influence extends globally, setting the precedent for drug regulation in Western countries.

The origins of the FDA date back to the early 20th century (see Marks, 1997; Temple, 1995), but it assumed its current structure in October 1962 with the passage of the Kefauver-Harris Drug Amendments to the Federal FD&C Act by US Congress. This legislation mandated that pharmaceutical companies, prior to marketing a drug, were obligated to demonstrate not just its safety but also furnish substantial proof of its efficacy for the intended purpose. That evidence had to consist of adequate and well-controlled studies. By refraining from explicitly delineating the terms "adequate" and "well-controlled", the legislation allowed experts in the field to set

the standards that would characterize both concepts within the new statutory framework. While the law itself did not provide a specific definition for a well-controlled study, testimonies presented to the US Congress clarified that it entailed, at the very least, the incorporation of control groups, random assignment of patients to control and therapeutic groups, and the implementation of methods to mitigate bias, such as standardized criteria for assessing effectiveness (Junod, 2008).

Equally crucially, the 1962 Drug Amendments mandated that the FDA grant explicit approval for the marketing application prior to the drug's commercialisation, representing another significant change. The approval process underwent systematic structuring across various phases. Phase 1 studies are uncontrolled investigations in humans, typically involving few healthy volunteers. Their purpose is to collect data on pharmacokinetics and pharmacodynamics, assess the initial safety of the drug at various doses, and, in certain cases such as chemotherapy, gather preliminary evidence of efficacy. Phase 2 trials involve the evaluation of adverse effects and efficacy in several hundred participants with the targeted medical condition. These trials usually lack a comparison group, are frequently not blinded or randomized, and commonly utilize surrogate measures, such as laboratory tests or imaging studies, to assess results. Phase 3 trials typically encompass several hundred to several thousand patients. Ideally, these studies incorporate concurrent controls, randomization, and clinical endpoints. The primary objective of phase 3 trials is to provide substantial evidence regarding the benefit-risk relationship of the drug to facilitate FDA approval.

After the enactment of the Drug Amendments in 1962, manufacturers and advocacy groups expressed concerns, contending that the heightened evidentiary requirements imposed by the FDA were prolonging the time and costs involved in obtaining approval (Lasagna & Wardell, 1975). These concerns resurfaced in the late 1980s and early 1990s, notably during the AIDS crisis (Epstein, 1996). In response to these criticisms, the FDA established the Accelerated Approval Program in 1992 (Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses, 1992). This program introduced an alternative regulatory pathway explicitly designed to expedite the translational process by applying looser regulatory rules in exceptional circumstances (Teira, 2020).

The AA program requires the drug to be formulated to address serious or life-threatening diseases, which are defined as conditions affecting daily functioning or assumed to pose a threat to survival if left untreated. Examples encompassed within this definition include HIV/AIDS, heart failure, and cancer. The second requirement entails that the drug must offer a "meaningful therapeutic benefit over existing treatments" (Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses, 1992, § 314.500), delineated as addressing a situation where a serious medical need is not met by currently available therapies. The third condition is related to surrogate endpoints. The FDA defined a surrogate endpoint as "a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and that is expected to predict the effect of the therapy" (New Drug, Antibiotic and Biological Drug Product Regulations: Accelerated Approval, 1992, p. 13,235). These measurements were not firmly established at the time but were defined as "rea-

sonably likely […] to predict clinical benefit" (Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses, 1992, § 314.500). For example, myocardial infarction or survival may be taken as a reference, instead of requiring a demonstration of improvement in the actual clinical endpoints.

However, not all surrogate endpoints are equal, as the causal relationship between surrogate endpoints and clinical outcomes varies across a spectrum. On one end, there are validated surrogate endpoints that the FDA frequently accepts for regular approval. For instance, high cholesterol or high blood pressure (hypertension) are significant risk factors for heart disease and stroke. Consequently, reducing these variables has long been considered acceptable data for approval. Nevertheless, the FDA has acknowledged that some surrogate endpoints may indicate spurious correlations. For instance, people with pneumonia often exhibit fever, but a fever cannot serve as a surrogate endpoint because it can be reduced while the pneumonia persists. The FDA expressed its readiness to grant accelerated approval when there is a "basis of adequate and well-controlled trials establishing that the drug has an effect […] that is reasonably likely (based on epidemiologic, therapeutic, or other evidence) to predict clinical benefit." (Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses, 1992, § 314.510).

Finally, in return for accepting such a lower bar of evidence, the FDA mandates post-marketing studies to confirm the predicted benefits. Pharmaceutical manufacturers are required to conduct studies to confirm the clinical benefit anticipated from the surrogate markers. Typically, these studies are of the kind submitted alongside the traditional approval process (phase III trials, i.e., RCTs). The approval provided by the AA pathway is conditional to such studies, and if the manufacturers fail to confirm the benefit in a full-fledged RCT, the FDA may withdraw approval.

From a philosophical perspective, the FDA's approach to drug regulation is grounded on the EBM epistemology. On this account, marketing authorization decisions must be based only on our best form of evidence, namely (statistical) evidence provided by RCTs. They serve in fact as the gold standard for investigating causal relationships between interventions and outcomes. This is because the use of randomization and other methodological devices (e.g., blind, allocation concealment, etc.) helps mitigate many of the inherent biases associated with alternative study designs.

Notably, the AA program moves slightly away from the EBM logic. In fact, under this program, drugs can be approved on the basis of evidentiary standards which are considered inferior to full-fledged RCTs. Approval can indeed be granted based on experimental designs other than phase III RCTs (e.g., single-arm trials, phase II trials, etc.) and adopting a surrogate outcome that can be measured earlier than the relevant clinical endpoint. Adopting surrogate outcomes (and/or alternative experimental designs) allows manufacturers to conduct shorter, smaller, and therefore more cost-effective clinical trials. This enables regulators to bring drugs to the market years before confirmatory trials that use a clinical endpoint are typically completed. In this sense, blood pressure, one of the most common examples of surrogate outcomes, is easier, faster, and cheaper to measure than long-term mortality resulting from stroke.

Nonetheless, manufacturers must prove that the considered surrogate outcome is a "reasonable" predictor of the relevant clinical outcome. Interestingly, in order to convince the regulators, manufacturers should provide evidence of the mechanism(s)

that link the surrogate outcome and the effects of interest. Ideally, a comprehensive understanding of the disease's causal pathways and the mechanisms underlying intervention would be provided (Fleming, 2005). However, it is a very complex task, and it is not always achievable (see, e.g., Aronson, 2005). In fact, when possible, validation often appeals to common grounds already accepted.[1]

The FDA aims to determine the acceptance of a surrogate endpoint by evaluating the scientific evidence for that particular endpoint. For instance, research studies illustrating a drug's impact on such endpoints must adhere to the standards of being "adequate and well-controlled", as mandated by the FD&C Act. In practice, however, the assessment of this evidence is ultimately a subjective judgment based on consensus, due to the challenging, if not impossible, standardization across diverse contexts (Lathia et al., 2009). The FDA maintains a list of surrogate endpoints that have been successfully used to approve new drugs, nonetheless "the acceptability of these surrogate endpoints for use in a particular drug or biologic development program will be determined on a case-by-case basis" (FDA, 2022). Their appropriateness is context-dependent, hinging on factors such as the specific disease, the characteristics of the studied patient population, the therapeutic mechanism of action, and the availability of existing treatments. Furthermore, in some cases, evidence of mechanisms linking the intervention and the surrogate outcome is also required. As we have noted, the correlation between the tested drug and the surrogate outcome can be established based on experimental designs different from RCTs. In such cases, given that statistical evidence provided by those alternative methods is considered less reliable, evidence of mechanisms (which complements and strengthens it) is also required.

From an evidential perspective, AA adopts a pluralist framework.[2] Both statistical evidence and evidence of mechanisms are required to establish a causal relation between the considered treatment and the relevant effects (see, e.g., FDA, 2019, 2023). Evidence of mechanisms is in fact aimed to "compensate" for the limitations of available statistical evidence.

Based on our analysis of AA, we consider that there are two ways in which evidence of mechanisms can contribute to establishing a causal relation in drug regulation. First, evidence of mechanisms can play an *extending* role. It can contribute to the extension (to further outcomes) of previously established causal relations. In AA, manufacturers should provide evidence of the mechanisms that link the surrogate outcome and the effects of interest. This evidence supports the extension (to the

[1] As pointed out by an anonymous reviewer, regarding surrogate biomarkers, the concept of "validation" is weak and fuzzy. There are no standardised evidentiary requirements to establish that a surrogate biomarker is a reliable predictor of a clinical outcome. This validation process usually relies on factors such as biological plausibility, correlations, and consensus within the scientific community. Given its significance, the validation of surrogate biomarkers is a widely discussed topic in the scientific literature (see, e.g., Ciani et al., 2017).

[2] Before the introduction of the AA framework, the FDA already occasionally considered evidence of mechanisms. For example, in 1961, FDA's non-approval of thalidomide was informed by evidence of mechanisms (Gillies, 2017). In this case, evidence of the mechanism of action of the drug was crucial to anticipate potential side effects (e.g., birth defects) (Stephens & Brynner, 2001). Given its molecular weight, thalidomide may cross the placenta and enter the foetal blood. Nonetheless, previous cases in which the FDA considered evidence of mechanisms are rare and non-systematic. The gathering, evaluation, and use of evidence of mechanisms differ from case to case.

effects of interest) of the causal relation established between the treatment and the surrogate outcome. In principle, a surrogate endpoint should fully capture the effect of the intervention on the mechanisms that influence the clinical outcome. Therefore, it is crucial to account for the causal pathway of the disease process, as well as the mechanisms of action of the intervention.

Second, evidence of mechanisms can play a *supporting* role. It can support the causal interpretation of a detected correlation. In AA, the supporting role is crucial when suboptimal evidence-gathering methods are employed to establish a correlation between the treatment and the surrogate outcome. In those cases, regulators require evidence of mechanisms accounting for the correlation. When statistical evidence is limited, even if a correlation between two variables is detected, it may still be the case that there is no causal relation between them. The identified correlation may be the result of a common cause, a semantic connection, or even mere chance. Providing evidence of the mechanism(s) through which the treatment influences the surrogate outcome supports the causal interpretation of the correlation.

Nonetheless, the evaluation of evidence of mechanisms is another challenging task. Like statistical evidence, evidence of mechanisms may diverge in quality. However, unlike with statistical evidence, there are not yet well-established and widely accepted criteria. Nonetheless, a tentative framework has recently been proposed. Parkkinnen et al. (2018) have developed a three-step procedure for assessing evidence of mechanisms in medicine. They argue that, when assessing evidence of mechanisms, we should consider the method, the implementation, and the results. For each of those aspects, quality indicators are introduced. First, based on our understanding of how methods work, we have to assess methods' reliability. Moreover, when model systems are employed, the similarity of models to humans should also be considered. Second, we have to evaluate the methods' implementation in the particular mechanistic studies. Special attention should be paid to the typical sources of error. Finally, we have to evaluate the stability of the obtained results. For this purpose, we should consider the independent detectability, consistency, and robustness of the identified mechanisms features. While this three-step approach is certainly an important advance in the evaluation of evidence of mechanisms, further work is required to translate it into a systematic evaluation procedure for drug approval.

2.1 Illustration of AA: the approval of crizotinib

Let us illustrate how the AA program works in practice and clarify the role that evidence of mechanisms plays into it by discussing the accelerated approval of Crizotinib. Crizotinib is a small molecule tyrosine kinase inhibitor, developed to block the activity of anaplastic lymphoma kinase (ALK) fusion proteins (an oncogenic driver) in patients with advanced or metastatic non-small cell lung cancer (NSCLC) who are ALK-positive.

Crizotinib received AA in 2011, based on the findings of two studies. Study 1001 represented a phase I clinical trial of Crizotinib conducted at multiple centres. This single-arm study included patients with various tumour types, excluding leukaemia (Camidge et al., 2012). During the dose-escalation phase, two individuals with ALK-positive NSCLC exhibited stable disease, prompting an amendment to introduce an

expansion cohort specifically for patients with ALK-positive NSCLC. This cohort received the recommended phase II regimen of 250 mg orally twice daily. Conversely, Study 1005, a multicentre and single-arm trial, was designated as a phase II investigation of Crizotinib. This study involved administering 250 mg of Crizotinib orally twice daily to patients with advanced NSCLC who had experienced tumour progression after at least one round of chemotherapy (Kim et al., 2012). In both trials, the endpoint for efficacy was the durable objective overall response rate (ORR), as determined by investigator assessment through imaging evaluations at baseline and every other cycle.[3] Both trials yielded positive results, showcasing ORRs of 50% and 61%, with median response durations of 42 and 48 weeks, respectively (Kazandjian et al., 2014). Further considerations, however, warranted the accelerated approval of Crizotinib.

The causal pathway of the disease is well-understood, ALK has been shown to be essential for the proliferation of tumour cells in different forms of cancer, including NSCLC (Kwak et al., 2010; Soda et al., 2007, 2008). The therapeutic hypothesis is that by inhibiting its activity we can slow down the tumour growth and therefore increase patients' survival. Evidence of mechanisms provided support for the hypothesis linking the surrogate outcome (tumour response) to the clinical outcome (survival). Indeed, it is well-established that when we effectively eliminate the tumour, whether through surgical excision or targeted destruction of cancer cells within the body, patients exhibit a significantly increased likelihood of survival. The eradication or substantial reduction of the tumour directly correlates with extended survival periods for patients. In fact, this correlation has been observed in different types of cancer (see, e.g., Ellingson et al., 2023). For this reason, the ORR is a well-established and largely adopted surrogate outcome for overall survival (the clinically relevant outcome) in solid tumours and it appears in the FDA list of validated surrogate endpoints (FDA, 2022).

Nonetheless, because the drug was tested in a single-arm study without a control group, the statistical evidence was deemed incomplete, necessitating further investigation to elucidate the mechanisms behind the observed correlation between drug administration and ORR. And that was provided by the drug mechanism of action. As mentioned above, Crizotinib works by blocking certain proteins that tell cells to grow, a mechanism well confirmed in vitro and in vivo, and consistent with the trial results (Shaw et al., 2011).

As a post-marketing requirement for AA, the manufacturer (Pfizer) was required to conduct a confirmatory study (i.e., an RCT) to verify the drug's clinical benefit. The results confirmed the efficacy of Crizotonib in improving the progression-free

[3] The ORR is calculated as the proportion of patients who achieve either complete response (CR) or partial response (PR) according to the predefined criteria (Park et al., 2003). The tumour response is categorized using standardized criteria for assessing tumour size changes, e.g., RECIST (Response Evaluation Criteria in Solid Tumours) criteria. These criteria define various response categories, including complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD), based on the extent of tumour shrinkage or growth compared to baseline measurements. This assessment is performed by measuring the dimensions of the lesions in images such as computed tomography scans.

survival in NLSCS (ALK positive) patients (Shaw et al., 2013), and the FDA granted the drug regular approval in 2013.[4]

To summarize, Crizotinib was approved based on low-quality statistical evidence (i.e., phase II trials), which was compensated for by well-established evidence of mechanisms. Evidence of mechanisms supported the causal interpretation of the detected correlation and linked the surrogate endpoint to clinically relevant outcomes.

3 The reliability of evidential pluralism

Up to this point, everything seems to be on track. AA operates within a framework of evidential pluralism, which carries inherent epistemic strengths and has demonstrated effectiveness in diverse cases. However, it is essential to recognize that relying on theoretical arguments or a few cases is insufficient to establish the reliability of evidential pluralism as a regulatory guideline and its potential as a superior alternative to standard EBM.

As some philosophers of science have argued, the reliability of regulatory strategies (and the choice between alternatives) does not indeed depend on theoretical considerations alone, but also on real-world outcomes. For example, López-Mas and Luján (2023) suggest defining regulatory success by considering regulatory objectives, opportunity costs, and pragmatic costs associated with different epistemic policies. What consequences count as costs, however, is open to interpretation and depends on the perspectives of involved actors and on the context of regulatory decisions.

In light of this, we claim that a complementary empirical (or naturalistic) approach to regulatory decision-making seems to be crucial to settle the issue. In this sense, Andreoletti and Teira say: "[e]ven 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" (Andreoletti & Teira, 2019, p. 1106). That is, a regulatory policy might be well-supported by our best casual epistemology (e.g., evidential pluralism), but this does not necessarily mean that the policy will deliver the best regulatory outcomes. Following the same argument, more recently, Luján and Todt have argued 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 method" (Luján & Todt, 2021, p. 3). Controversies about standards and methods should be conceptualized as empirically testable arguments about the relative advantages and disadvantages of alternative epistemologies.

[4] Progression-free survival (PFS) refers to the length of time, during and after a treatment, in which the patient lives without the disease worsening. In the context of cancer, it specifically measures the duration from the start of treatment until the tumor shows signs of progression or until the patient experiences a defined event (such as death). PFS is contested by many (see e.g., Booth et al., 2023), but it is widely considered a very well-established primary endpoint that can be used to get traditional approval for cancer drugs.

3.1 Sung and Holman's open-cases approach

Following the abovementioned considerations, Sung and Holman (2023) have recently challenged the reliability of evidential pluralism in the context of pharmaceutical regulation. They claim that there is often a difference in focus between proponents and critics of evidential pluralism. While proponents have considered medical evidence in general, most critics have focused on a narrow role. Sung and Holman suggest that to overcome this mismatch, arguments should be confined to specific domains. They also argue that when we consider how evidential pluralism would perform in actual contexts of medical research its epistemic virtues seem to fade away. This is because, for example, evidential pluralism is permeable to non-evidential social forces (e.g., industry influence) (see Andreoletti & Teira, 2019; Holman, 2019).[5]

To evaluate the performance of evidential pluralism, Sung and Holman introduce an approach based on open case studies. On this account, we should focus on cases that suit one of the competing proposals—standard EBM or evidential pluralism—and whose outcome is still unknown. Depending on the resulting outcome (positive or negative), the adopted approach would be reinforced or undermined.

Sung and Holman argue that historical case studies raise important difficulties. First, reasoning about cases whose outcome is already known involves complicated counterfactual arguments. We have to consider the potential consequences of having adopted the alternative approach. Second, there are always diverse interpretations of each case, which can lead to strategic adoption by advocates of competing proposals. For example, when a case's outcome is positive, advocates of evidential pluralism could adopt the interpretation that emphasizes the contribution of evidence of mechanisms. Sung and Holman hold that open case studies, where the outcome is still unknown, avoid such difficulties.

However, the proposal advanced by Sung and Holman carries some limitations that necessitate careful consideration. First, although the value of previous cases is not denied, nothing is said about how to assess them and integrate them into evaluations of regulatory frameworks. Their suggestion exclusively focuses on scenarios where outcomes remain uncertain. It does not address the integration of the invaluable insights and information that historical cases can offer. Consequently, their proposal risks disregarding crucial evidence, potentially leading to inconclusive or partial conclusions.

Furthermore, a second limitation of their proposal emerges: the assessment of both EBM and evidential pluralism is deferred until the results of a series of ongoing cases crystallize. Additionally, the extent of the track record needed for a compelling argument remains ambiguous. This delay effectively stalls the discourse surrounding medical evidence until conclusive outcomes materialize.

Finally, as we are going to illustrate in the following section, the focus on open cases does not avoid the emergence of conflicting interpretations. As Sung and Holman explain, regarding a particular open case, diverse predictions are possible. Advo-

[5] Although Sung and Holman (2023) consider both the reliability and the exploitability of evidential pluralism, in this section, we will focus on the reliability of evidential pluralism.

cates of different approaches may make different predictions about the outcomes. Nonetheless, regarding a specific open case, diverse interpretations are also possible. Interpretations of a case or regulatory decision may differ concerning which epistemic framework it matches. It should be noted that diverse interpretations are possible even when similar predictions are made. For instance, advocates of standard EBM and evidential pluralism may agree that a certain decision will have a negative outcome but differ in their interpretation of what epistemic framework the decision followed. Advocates of standard EBM may argue that the decision matches evidential pluralism, while advocates of evidential pluralism argue that the decision matches standard EBM. In that case, when the outcome was settled, it would be difficult to establish which framework is reinforced or undermined.

3.2 The accelerated approval of Aducanumab

In order to illustrate their approach, Sung and Holman (2023) discuss the accelerated approval of Aducanumab by the FDA. They claim that Aducanumab was approved on both statistical and mechanistic grounds, thus representing an example of evidential pluralism. In contrast, applying the traditional EBM approach would have led to the rejection of the approval. Eventually, the fate of the drug will depend on confirmatory trials, and evidential pluralists should be prepared to face the consequences of their stance, which may include the possibility of a negative outcome.

In what follows, we present a pluralist case for rejecting Aducanumab's approval, drawing on our previous analysis of the evidential approach adopted in AA. This discussion illustrates the relevance of our analysis of the roles of evidence of mechanisms. Nevertheless, before exploring the details, it is worth noting that the approval of Aducanumab was highly unusual and controversial.[6] The FDA made the decision to approve the drug against the Advisory Committee's recommendation. This was a rare exception, as the FDA usually follows the committee's recommendations, and in a few cases of disagreement, the FDA tends to be less likely than its advisory committees to approve new products (Zhang et al., 2019).

As explained by Sung and Holman, the Peripheral and Central Nervous System Drugs (PCNS) Advisory Committee and the FDA evaluated both the statistical evidence and the evidence of mechanisms presented by the manufacturer (Biogen) regarding Aducanumab. It is important to note that the statistical evidence alone was not sufficient to justify its approval. While there are several reasons for this, here, we can just highlight a few points that are particularly relevant to our argument.

Two RCTs of Aducanumab were conducted (Study 301 and Study 302), with the objective of assessing the efficacy of Aducanumab in reducing cognitive decline measured on the CDR-SB. The Study 301 yielded negative results, while the Study 302 detected a significant effect in high-dose patients. All in all, what the studies could demonstrate was just a correlation between the drug and a surrogate biomarker, namely the reduction of amyloid plaques as measured by a PET scan (see Tampi et al., 2021 for more details on these two studies). Yet, Biogen managed to downplay

[6] In a recent paper, Carome (2022) has analyzed in detail the anomalous process that led to the approval of Aducanumab.

the discrepancy between the unfavorable outcomes of Study 301 and the somewhat favorable outcomes of Study 302. Basically, they presented a post hoc analyses of Study 302 data, supplemented by data from Study 103 (phase I study, originally not designed to gauge efficacy but focused on safety and the impact of Aducanumab on brain amyloid-beta), as an accurate reflection of Aducanumab's efficacy in addressing Alzheimer's disease. However, this selective approach, often referred to as "cherry-picking", lacks both statistical and scientific appropriateness.

At this point, the evidence of mechanisms became crucial. The association between the drug and the reduction of amyloid plaques was supported by robust evidence of mechanisms not only from study 103 (as mentioned by Sung and Holman), but also from numerous other studies investigating the mechanism of action of Aducanumab-like antibodies and the reduction of plaques in the brain (Sevigny et al., 2016). These studies range from in vitro experiments and in vivo animal studies to clinical trials. Therefore, the association between the drug and the surrogate biomarker was considered genuine, as the evidence of mechanisms supported and complemented the weakness of the statistical evidence.

As we explained above, the existence of a reliable association between the drug and a surrogate biomarker is not sufficient for AA because it does not prove any benefit for patients. The regulatory guidelines demand that the surrogate outcome must be a "reasonable predictor" of a clinical outcome (FDA, 2020b). Therefore, the key question in the case of Aducanumab is whether there is evidence linking the surrogate outcome to a clinical benefit. In other words, the hypothesis to be proven is that we can slow down the cognitive decline of Alzheimer's patients by reducing the amyloid plaques in their brains. One should keep in mind that the main issues reported by dementia patients have not been related to excessive amyloid or tau in the brain. Instead, these issues have centered around problems such as cognitive impairment, depression, sleep disturbances, agitation, confusion, and so on (Khachaturian, 2022).

However, the statistical evidence alone was not enough to establish a causal relationship. There is a need to explain the precise mechanism of action of this intervention, i.e., the molecular signaling pathways for maintaining or restoring neuronal functioning that underlie the behavior-cognitive impairment. In other words, we need evidence of mechanisms that extend the causal relationship between the drug and a surrogate outcome to the clinical effect of interest.

Unfortunately, this sort of evidence is missing in the case of Aducanumab. Certainly, a mechanism hypothesis has been formulated to account for the link between the surrogate outcome and the clinical endpoint. The amyloid cascade hypothesis, in principle, could explain the link between plaque reduction and cognitive improvements. However, no evidence has been obtained in support of that mechanism hypothesis (see Herrup, 2015). Over the years, many studies have failed to support the amyloid cascade hypothesis. So, as of today, there is no evidence of mechanisms warranting the extension from the surrogate outcome to the clinical effect.

The PCNS committee clearly recognized such an evidential gap in their evaluation report.[7] For instance, in the "Final Summary Minutes of the Peripheral and

[7] Meeting materials are available on the following link: https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-6-2020-meeting-peripheral-and-central-nervous-system-drugs-

Central Nervous System Drugs Advisory Committee Meeting", we find the following information:

> The Committee agreed that the data shown provided clear evidence that aducanumab cleared amyloid plaques and reduced tau deposits in the brain. However, the Committee was not persuaded that the clearing of these neurodegenerative aggregated proteins correlated with cognitive improvement. The Committee recommended that there should be substantial evidence relating specific biomarkers to disease progression before there is a determination that the clearance/reduction of these biomarkers are truly related to clinical benefit (cognitive improvement). (FDA, 2020a, p. 5)
>
> A marginal majority of the committee members were uncertain as to whether the Applicant presented strong evidence of a pharmacodynamic effect on Alzheimer's disease pathophysiology. These members indicated that the Applicant failed to show the correlation between the biomarkers and clinical effects. Although the members were convinced of aducanumab's pharmacodynamic effects on amyloid plaques, they noted that the data associated with tau deposits were "murky". In addition, they expressed uncertainties as to whether the biomarkers shown are true indicators for Alzheimer's disease and urged the Applicant to study other biomarkers (proteopathies) that may have played a role in this setting. (FDA, 2020a, p. 5)

In sum, the Aducanumab approval does not suit the pluralist framework at the basis of the AA program. In AA, statistical evidence, which is often scarce and suboptimal, is insufficient to warrant approval. Having evidence of mechanisms is essential to support the detected correlation, mitigate any biases, and extend the observed effect on a surrogate outcome to a clinical effect. However, in the case of Aducanumab, there is no evidence of mechanisms that can justify its extension to the relevant clinical outcome. As such, its approval contravenes the core principle of evidential pluralism: the limitations of statistical evidence are overlooked, and no valid complementary evidence of mechanisms is provided.

The conditional approval of Aducanumab by the FDA, consequently, constitutes an anomaly or exception in the AA framework. Although the AA demands evidence of mechanisms that links the surrogate and the clinical outcomes, this kind of evidence was not provided in the Aducanumab evaluation. In fact, this lack of evidence has driven most of the criticisms of Aducanumab's approval (see, e.g., Mullard, 2021). Nonetheless, while the advisory committee correctly reasoned in pluralistic terms and its recommendation was negative, the FDA overlooked the absence of crucial evidence of mechanisms and approved the drug.

advisory-committee-meeting. Accessed 4th September 2023.

4 The reliability of regulatory programs

As we have argued, Sung and Holman's proposal falls short in assessing the reliability of regulatory epistemologies in drug regulation. It does not consider the integration of historical cases and overlooks contextual complexities. Furthermore, as we have illustrated through our analysis of Aducanumab's approval, focusing solely on ongoing cases does not eliminate conflicting interpretations of those cases, which is one of the alleged benefits of Sung and Holman's approach.

Nonetheless, we align ourselves with Sung and Holman's overarching perspective that the discourse surrounding medical evidence should concentrate on distinct domains in the real world. We also agree that pharmaceutical regulation provides fertile ground to assess evidential pluralism as well as other views on medical epistemology in practice.

In this section, we propose an alternative empiric approach to evaluate the reliability of evidential pluralism in drug regulation. As we have argued, the AA program adopts an approach to evidence that closely resembles evidential pluralism. Our proposal aims to assess the reliability of the AA program and, consequently, the implications for evidential pluralism. We consider that, given their resemblance, AA can provide crucial insights for assessing the prospects of the pluralist framework. This approach, as Sung and Holman's proposal, does not rely on theoretical arguments or particular case studies.

Given that the paramount aim of drug regulation is to ensure patient well-being, a natural course of action is to examine the tangible outcomes of AA. Such a perspective is rooted in the notion that the most reliable measure of a regulatory approach's value lies in its capacity to safeguard patients from ineffective and harmful interventions (Marks, 1997). The main question, therefore, is how we can empirically evaluate pharmaceutical regulation policies designed to protect patients.

To evaluate the merit of regulatory initiatives we must establish an empirical index capable of discerning which decision procedure, or epistemological framework, is best suited to achieve the overarching goal of patient protection. This empirical index would serve as a crucial yardstick to judge the efficacy and reliability of the various regulatory policies in the realm of pharmaceuticals.

One could look at many different data to perform an empirical analysis of regulatory outcomes. In the context of AA, the most common empirical index for determining the performance of the program is the rate of conversion. As mentioned before, the FDA, once granted the accelerated approval, requires the companies to run full-fledged RCTs to confirm the results. If these RCTs are positive the FDA converts the accelerated approval to a standard approval, otherwise the regulators may withdraw the drug from the market. Therefore, the conversion rate represents a direct and immediate benchmark to assess the success of the AA decisions. If the evidential pluralism holds, we should expect a high rate of conversion and few withdrawals of drugs. Such predictions can be empirically tested. For instance, in a recent article, Beakes-Read et al. (2022) examined the performance of the AA program since its inception in 1992. The authors found that, as of December 2021, the FDA has granted AA to 278 drugs and 50% (139) of them received conversion to standard approval. In the same period, the FDA has withdrawn 32 drugs, meaning a withdrawal rate of

11%. This data however also includes the drugs that are still pending completion of confirmatory trials or FDA review (107 − 38%). If we look instead at the data without including the drugs for which a decision has not been made yet, then the withdrawal rate increases to 19%.[8]

Moreover, the authors noted that there have been very few delinquent cases (i.e., drugs for which the sponsor has failed to complete the confirmatory trials by the agreed-upon deadline) or dangling cases (i.e., drugs remaining on the market despite confirmatory trials not confirming clinical benefit) and concluded that the program has been largely successful.

However, when considered in isolation, these data are indicative but lack significant information regarding the reliability of evidential pluralism. In fact, we cannot draw any informative conclusions from them. To draw a normative conclusion, specifically asserting that evidential pluralism is at least as effective as EBM epistemology in the context of drug regulation, a direct comparison between the two approaches is essential.

Andreoletti and Teira (2019) have suggested adopting withdrawal rates as an empirical benchmark to evaluate the efficacy of regulatory initiatives. Building on that suggestion, a more effective approach to assessing the AA program can be developed by comparing its withdrawal rate with that of traditional approval. Drug withdrawals may indeed highlight failures in the initial regulatory evaluation of evidence. The most recent estimate, available in the scientific literature, suggests a withdrawal rate of less than 2% for drugs that received marketing authorization in the US between 1950 and 2011 (Onakpoya et al., 2016). An older analysis of drug discontinuations in the US, spanning the years 1980 to 2010, concluded that out of the 740 newly introduced drugs in that timeframe, 118 (15.9%) were terminated (Qureshi et al., 2011). However, both these estimates include the withdrawals of drugs that received AA. Moreover, the inconsistency in the results suggests that acquiring dependable and consistently standardized data on drug withdrawals is a challenging task. Due to these reasons, these figures cannot serve as a reliable basis for comparison.

Data on drug discontinuations for safety reasons before the introduction of the AA program may offer a more useful comparison. For example, according to a pioneering review by Bakke et al. (1995), between 1974 and 1993, 377 new chemical entities were approved in the United States. Of these, 10 were withdrawn from the market for safety reasons. Thus, one can estimate approximately a 3% safety-related withdrawal rate for standard approval. This finding aligns with the outcomes of another study examining the regulatory impacts of the Prescription Drug User Fee Act[9] (Berndt et al., 2005), which revealed that between 1980 and 1992, 320 drugs were approved with 9 withdrawals (2.8%), and from 1993 to 2002, 361 drugs were approved with 8 withdrawals (2.2%). This combined to a total of 661 approvals and 17 withdrawals (2.5%) (Berndt et al., 2005). These figures are significantly lower than those reported by Beakes-Read et al. (2022) for AA.

[8] The data for this calculation is available at https://www.fda.gov/media/151146/download?attachment.

[9] The Prescription Drug User Fee Act (PDUFA) allows the FDA to collect fees from drug manufacturers submitting new drug applications for evaluation.

If we consider such data for analysis, which include a 2.8% withdrawal rate for traditional approval decisions ($N=320$) before AA was introduced and an 11% withdrawal rate for the AA decisions ($N=278$), then a straightforward statistical analysis for comparing proportions reveals a statistically significant difference in withdrawal rates between the two regulatory regimes [$z = -4.0248$, $P<.00001$].

Henceforth, at first glance, the available data indicate that the AA performance, particularly concerning drug withdrawals, is inferior to that of the standard approval. However, determining the acceptability of this disparity necessitates a more normative approach, which goes beyond the scope of this paper. Notably, the AA program intends to expedite the market approval of new drugs for severe diseases, accommodating greater uncertainty. Evaluating the acceptability of such a trade-off entails ethical reflections (Andreoletti & Blasimme, 2023).

Similar conclusions can be derived regarding the evidential pluralist epistemology underlying the AA. The data suggests that its application results in suboptimal decisions compared to standard approval grounded in EBM epistemology. Nevertheless, as mentioned above, given the constraints of this preliminary analysis, the data serves as suggestive rather than conclusive.

4.1 Objections

Regarding the proposed empirical approach, nonetheless, some objections may be raised. Firstly, it may be argued that regular approval and accelerated approval serve distinct regulatory objectives. Therefore, direct comparisons, particularly in terms of withdrawal rates, might not be appropriate. The reason behind this disparity lies in the fact that the acceleration of the approval process often involves the acceptance of less established evidence, leading to an increased likelihood of errors. Consequently, a higher number of withdrawals could be anticipated. Nevertheless, this objection assumes a primacy of the EBM epistemological approach. Those who advocate for evidential pluralism, however, contend that the limitations of statistical evidence are counterbalanced by evidence of mechanisms and deny EBM's primacy. This consideration thereby suggests that the regulatory outcomes are indeed comparable. Consequently, no notable differences should be expected or assumed.

A second potential objection hinges on inherent dissimilarity between the types of drug withdrawals in regulatory programs, making a direct comparison problematic. Critics may correctly argue that accelerated drugs are often withdrawn due to a lack of efficacy, whereas drugs granted regular approval are typically withdrawn due to adverse effects.[10] This difference is certainly a product of the distinctive regulatory goals each pathway serves. Withdrawals are influenced by "structural" factors, and they mostly depend on the political authority granted to the FDA. In this sense, the FDA can withdraw drugs that have been "traditionally" approved only for safety rea-

[10] The relevance and reliability of evidence of mechanisms may diverge across scenarios. Pérez-González and Rocca (2022) have recently analysed and compared the roles of evidence of mechanisms in inferences about efficacy and about safety. They argue that evidence of mechanisms is more reliable for disconfirming the efficacy than for confirming the efficacy of an intervention. Regarding safety, however, they claim that evidence of mechanisms is more reliable for confirming (than for disconfirming) the presence of adverse effects of an intervention.

sons, for example, severe drug reactions not anticipated in the trials. This is because, once approved, the efficacy of drugs is considered established. Whereas in the AA framework, the FDA can withdraw drugs if they fail the confirmatory trial (and therefore not converted to traditional approval). So, in this latter case, the withdrawal is for efficacy reasons. Once converted, though these drugs can be withdrawn only for safety reasons. This is because, again, after conversion, the efficacy of the drug is supposed to be established.[11] However, even when evaluated within the context of evidential pluralism, the prevalence of withdrawals prompted by the absence of therapeutic benefit for accelerated drugs is not inherently expected to be high. This divergence, in fact, further underscores the importance of evaluating the efficiency of the AA pathway. It compels us to question whether the observed withdrawal rates of accelerated drugs align with the intended goals of the pathway. In this regard, withdrawal rates, despite the inherent differences, still offer a meaningful metric for assessing the viability of the AA pathway.

Finally, the viability of the suggested alternative—comparison to historical approvals dating back to 1974—raises questions about its relevance in the current regulatory landscape. Notably, the regulatory paradigm has undergone substantial changes over the years, rendering a pre-post comparison potentially inappropriate. The use of historical controls in research settings has been extensively discussed by methodologists and is acknowledged as problematic across various fields due to potential biases. Drawing comparisons between present and historical populations may result in confounded outcomes, given, for example, substantial shifts in patient demographics, disease characteristics, or other variables that can impede meaningful comparability between the two groups. Nonetheless, while recognizing these limitations, the use of historical controls might still provide useful information. For example, in the analysis of the impacts of regulatory innovation, historical comparisons have been employed (Berndt et al., 2005).

All these objections rightfully raise considerations about the extent of direct comparability between regular approval and AA withdrawals. However, such limitations do not inherently invalidate the empirical approach. Instead, they underscore the complexity of regulatory outcomes and reinforce the necessity of adopting a more nuanced evaluation strategy. While the withdrawal rates may not serve as a perfectly analogous yardstick for both pathways, it remains possible to identify alternative comparators. An illustrative example lies in assessing the therapeutic value of drugs—a measure that accounts for the tangible benefit brought to patients' lives. This approach evaluates the drugs' real-world impact on patients' health and well-being, ultimately reflecting the success of the regulatory approach in delivering efficacious treatments (see, e.g. Vokinger et al., 2022a, b).

[11] It is important to note that drugs may also be withdrawn for reasons beyond safety or efficacy, such as the commercial choices made by the producers (see, e.g., Prasad, 2014).

5 Conclusions

In this paper, we have explored the intricate interplay between evidential pluralism and drug regulation. EBM, which privileges statistical evidence, has driven clinical practice and regulatory decision-making for decades. However, as its limitations became evident, evidential pluralism emerged as an alternative, acknowledging the complementary nature of statistical evidence and evidence of mechanisms. To illuminate the potential contribution of evidence of mechanisms in drug regulation, we have analysed the AA program adopted by the FDA. We argued that the AA program follows a pluralist approach to the evidence on which regulatory decisions are made. This approach (and the diverse roles played by evidence of mechanisms) has been illustrated through the accelerated approval of Crizotinib.

To assess and evaluate the reliability of evidential pluralism in drug regulation, however, it is crucial to consider real-world outcomes. Sung and Holman's innovative open-case approach presents a notable step toward addressing the empirical gap. By concentrating on ongoing cases where regulatory outcomes are uncertain, their framework aims to align the theoretical considerations of evidence with real-world consequences. Nevertheless, we have shown that, despite its potential, this approach is not without limitations. The selective focus on ongoing cases discounts the invaluable insights derived from historical outcomes and risks overlooking cases with known conclusions. While Sung and Holman offer valuable insights into Aducanumab's approval, a closer look reveals some limitations in their discussion. This highlights that focusing on open cases does not resolve the challenge of interpreting such cases.

Recognizing the complexity of empirical evaluation, this paper has proposed an alternative and more systematic method focused on the outcomes of regulatory programs. By examining the withdrawal rates of drugs granted AA and comparing them with those granted traditional approval, a quantifiable benchmark emerges. The comparison elucidates the real-world impact of evidential pluralism in terms of regulatory outcomes.

A definitive resolution to the question of the reliability of AA and evidential pluralism is out of the scope of this paper. A comprehensive empirical endeavour would indeed necessitate a more extensive commitment. Nevertheless, our proposed framework represents a preliminary step towards shedding light on this matter and contributes to the ongoing discourse in the philosophy of science and medicine (see Andreoletti & Teira, 2019; Luján & Todt, 2021; Sung & Holman, 2023). Moreover, as the discussion continues to evolve, our endeavour not only contributes to the academic discourse but also highlights the need for a more systematic approach to understanding the complex relationship between epistemology and drug regulation.

Acknowledgements We thank the anonymous reviewers for their insightful comments and suggestions.

Funding MA has been supported by ERA-NET ELSA_NEURON_026 and by the Swiss National Science Foundation (SNSF — 10NE17_199434). SPG has been funded by Spanish Ministry of Science and Innovation (Grant n. PID2021-125936NB).
Open access funding provided by Swiss Federal Institute of Technology Zurich

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

References

Abdin, A. Y., Auker-Howlett, D., Landes, J., Mulla, G., Jacob, C., & Osimani, B. (2019). Reviewing the mechanistic evidence assessors E-Synthesis and EBM+: A case study of Amoxicillin and Drug reaction with Eosinophilia and systemic symptoms (DRESS). *Current Pharmaceutical Design*, *25*(16), 1866–1880. https://doi.org/10.2174/1381612825666190628160603

Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses, 21 CFR, § 314.500 et seq (1992). https://www.ecfr.gov/current/title-21/part-314/subpart-H

Andersen, L. M., & Kjær, J. N. (2019). Book Review: Evaluating evidence of mechanisms in Medicine: Principles and procedures. *Journal of Evaluation in Clinical Practice*, *25*(6), 1226–1227. https://doi.org/10.1111/jep.13093

Andreoletti, M. (2021). Why do we need randomized controlled trials? Medical scandals and the evolution of drug regulation. *Future of Science and Ethics*, *6*(1), 54–63.

Andreoletti, M., & Blasimme, A. (2023). Accelerated drug approval: Meeting the ethical yardstick. *Bioethics*, *37*(7), 647–655. https://doi.org/10.1111/bioe.13191

Andreoletti, M., & Teira, D. (2019). Rules versus standards: What are the costs of epistemic norms in drug regulation? *Science Technology & Human Values*, *44*(6), 1093–1115. https://doi.org/10.1177/0162243919828070

Anjum, R. L., Copeland, S., & Rocca, E. (2020). Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. *BMJ Evidence-Based Medicine*, *25*(1), 6–8. https://doi.org/10.1136/bmjebm-2018-111092

Aronson, J. K. (2005). Biomarkers and surrogate endpoints. *British Journal of Clinical Pharmacology*, *59*(5), 491–494. https://doi.org/10.1111/j.1365-2125.2005.02435.x

Aronson, J. K., Caze, A. L., Kelly, M. P., Parkkinen, V. P., & Williamson, J. (2018). The use of mechanistic evidence in drug approval. *Journal of Evaluation in Clinical Practice*, *24*(5), 1166–1176. https://doi.org/10.1111/jep.12960

Auker-Howlett, D., & Wilde, M. (2020). Reinforced reasoning in medicine. *Journal of Evaluation in Clinical Practice*, *26*(2), 458–464. https://doi.org/10.1111/jep.13269

Bakke, O. M., Manocchia, M., de Abajo, F., Kaitin, K. I., & Lasagna, L. (1995). Drug safety discontinuations in the United Kingdom, the United States, and Spain from 1974 through 1993: A regulatory perspective. *Clinical Pharmacology & Therapeutics*, *58*(1), 108–117. https://doi.org/10.1016/0009-9236(95)90078-0

Beakes-Read, G., Neisser, M., Frey, P., & Guarducci, M. (2022). Analysis of FDA's Accelerated approval program performance December 1992–December 2021. *Therapeutic Innovation & Regulatory Science*, *56*(5), 698–703. https://doi.org/10.1007/s43441-022-00430-z

Berndt, E. R., Gottschalk, A. H. B., Philipson, T. J., & Strobeck, M. W. (2005). Industry funding of the FDA: Effects of PDUFA on approval times and withdrawal rates. *Nature Reviews Drug Discovery*, *4*(7), 545–554. https://doi.org/10.1038/nrd1774

Booth, C. M., Eisenhauer, E. A., Gyawali, B., & Tannock, I. F. (2023). Progression-free survival should not be used as a primary End Point for Registration of Anticancer drugs. *Journal of Clinical Oncology*, *41*(32), 4968–4972. https://doi.org/10.1200/JCO.23.01423

Camidge, D. R., Bang, Y. J., Kwak, E. L., Iafrate, A. J., Varella-Garcia, M., Fox, S. B., Riely, G. J., Solomon, B., Ou, S. H. I., Kim, D. W., Salgia, R., Fidias, P., Engelman, J. A., Gandhi, L., Jänne, P. A., Costa, D. B., Shapiro, G. I., LoRusso, P., Ruffner, K., & Shaw, A. T. (2012). Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: Updated results from a phase 1 study. *The Lancet Oncology*, *13*(10), 1011–1019. https://doi.org/10.1016/S1470-2045(12)70344-3

Carome, M. (2022). Public Citizen's Advocacy Campaign Opposing FDA Approval of Aducanumab for Alzheimer's Disease: The Fight Against Regulatory capture. *Health Matrix: The Journal of Law-Medicine*, *32*(1), 31–54.

Ciani, O., Buyse, M., Drummond, M., Rasi, G., Saad, E. D., & Taylor, R. S. (2017). Time to review the role of surrogate end points in Health Policy: State of the art and the Way Forward. *Value in Health*, *20*(3), 487–495. https://doi.org/10.1016/j.jval.2016.10.011

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, *33*(2), 339–360. https://doi.org/10.1007/s11245-013-9220-9

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Ellingson, B. M., Wen, P. Y., Chang, S. M., Van Den Bent, M., Vogelbaum, M. A., Li, G., Li, S., Kim, J., Youssef, G., Wick, W., Lassman, A. B., Gilbert, M. R., De Groot, J. F., Weller, M., Galanis, E., & Cloughesy, T. F. (2023). Objective response rate targets for recurrent glioblastoma clinical trials based on the historic association between objective response rate and median overall survival. *Neuro-Oncology*, *25*(6), 1017–1028. https://doi.org/10.1093/neuonc/noad002

Epstein, S. (1996). *Impure Science: AIDS, activism, and the politics of knowledge*. University of California Press.

FDA(2022). *Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure*. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure

FDA (2023). *Clinical Trial Considerations to Support Accelerated Approval of Oncology Therapeutics: Guidance for Industry*. https://www.fda.gov/media/166431/download

FDA (2020b). *Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff*. https://www.fda.gov/media/133511/download

FDA (2020a). *Final Summary Minutes of the Peripheral and Central Nervous System Drugs Advisory Committee Meeting*. https://www.fda.gov/media/145690/download

FDA (2019). *Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products: Guidance for Industry*. https://www.fda.gov/media/133660/download

Fleming, T. R. (2005). Surrogate endpoints and FDA's Accelerated approval process. *Health Affairs*, *24*(1), 67–78. https://doi.org/10.1377/hlthaff.24.1.67

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

Gillies, D. (2017). Evidence of mechanism in the evaluation of streptomycin and thalidomide. *Studies in history and philosophy of Science Part C: Studies. History and Philosophy of Biological and Biomedical Sciences*, *66*, 55–62. https://doi.org/10.1016/j.shpsc.2017.06.003

Gillies, D. (2018). *Causality, probability, and Medicine*. Routledge. https://doi.org/10.4324/9781315735542

Herrup, K. (2015). The case for rejecting the amyloid cascade hypothesis. *Nature Neuroscience*, *18*(6), 794–799. https://doi.org/10.1038/nn.4017

Holman, B. (2019). Philosophers on drugs. *Synthese*, *196*(11), 4363–4390. https://doi.org/10.1007/s11229-017-1642-2

Howick, J. (2011a). Exposing the vanities—and a qualified Defense—Of mechanistic reasoning in Health Care decision making. *Philosophy of Science*, *78*(5), 926–940. https://doi.org/10.1086/662561

Howick, J. (2011b). *The philosophy of evidence-based medicine*. Wiley-Blackwell.

Howick, J., Glasziou, P., & Aronson, J. K. (2013). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics*, *34*(4), 275–291. https://doi.org/10.1007/s11017-013-9266-0

Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, *25*(2), 139–157. https://doi.org/10.1080/02698595.2011.574856

Junod, S. (2008). FDA and clinical drug trials: A short history history corner. *FDLI Update*, *2008*(2), 55–57.

Kazandjian, D., Blumenthal, G. M., Chen, H. Y., He, K., Patel, M., Justice, R., Keegan, P., & Pazdur, R. (2014). FDA approval Summary: Crizotinib for the treatment of metastatic non-small cell Lung Cancer with anaplastic lymphoma kinase rearrangements. *The Oncologist*, *19*(10), e5–e11. https://doi.org/10.1634/theoncologist.2014-0241

Khachaturian, Z. S. (2022). The 'Aducanumab Story': Will the last chapter spell the end of the 'Amyloid hypothesis' or Mark a New beginning? *The Journal of Prevention of Alzheimer's Disease*, *9*(2), 190–192. https://doi.org/10.14283/jpad.2022.36

Kim, D. W., Ahn, M. J., Shi, Y., De Pas, T. M., Yang, P. C., Riely, G. J., Crino, L., Evans, T. L., Liu, X., Han, J. Y., Salgia, R., Moro-Sibilot, D., Ou, S. H. I., Gettinger, S. N., Wu, Y. L., Lanzalone, S., Polli, A., Iyer, S., & Shaw, A. T. (2012). Results of a global phase II study with crizotinib in advanced *ALK*-positive non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology*, *30*(15), 7533–7533.

Kwak, E. L., Bang, Y. J., Camidge, D. R., Shaw, A. T., Solomon, B., Maki, R. G., Ou, S. H. I., Dezube, B. J., Jänne, P. A., Costa, D. B., Varella-Garcia, M., Kim, W. H., Lynch, T. J., Fidias, P., Stubbs, H., Engelman, J. A., Sequist, L. V., Tan, W., Gandhi, L., & Iafrate, A. J. (2010). Anaplastic lymphoma kinase inhibition in non–small-cell Lung Cancer. *New England Journal of Medicine*, *363*(18), 1693–1703. https://doi.org/10.1056/NEJMoa1006448

Lasagna, L., & Wardell, W. M. (1975). The FDA, Politics, and the Public. *Journal of the American Medical Association*, *232*(2), 141–142. https://doi.org/10.1001/jama.1975.03250020015015

Lathia, C., Amakye, D., Dai, W., Girman, C., Madani, S., Mayne, J., MacCarthy, P., Pertel, P., Seman, L., Stoch, A., Tarantino, P., Webster, C., Williams, S., & Wagner, J. (2009). The Value, qualification, and Regulatory Use of surrogate end points in Drug Development. *Clinical Pharmacology & Therapeutics*, *86*(1), 32–43. https://doi.org/10.1038/clpt.2009.69

López-Mas, R., & Luján, J. L. (2023). Comparing regulatory options: The role of epistemic policies and pragmatic consequences. *Science and Public Policy*, scad077. https://doi.org/10.1093/scipol/scad077

Luján, J. L., & Todt, O. (2021). Evidence based methodology: A naturalistic analysis of epistemic policies in regulatory science. *European Journal for Philosophy of Science*, *11*(1), 26. https://doi.org/10.1007/s13194-020-00340-7

Marks, H. M. (1997). *The progress of experiment: Science and therapeutic reform in the United States, 1900–1990*. Cambridge University Press.

Mullard, A. (2021). Landmark Alzheimer's drug approval confounds research community. *Nature*, *594*(7863), 309–310. https://doi.org/10.1038/d41586-021-01546-2

New drug (1992). Antibiotic and biological drug product regulations: Accelerated approval, 57 FR, 13234–13242.

Onakpoya, I. J., Heneghan, C. J., & Aronson, J. K. (2016). Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: A systematic review of the world literature. *BMC Med*, *14*, 10. https://doi.org/10.1186/s12916-016-0553-2

Park, J. O., Lee, S. I., Song, S. Y., Kim, K., Kim, W. S., Jung, C. W., Park, Y. S., Im, Y. H., Kang, W. K., Lee, M. H., Lee, K. S., & Park, K. (2003). Measuring response in solid tumors: Comparison of RECIST and WHO Response Criteria. *Japanese Journal of Clinical Oncology*, *33*(10), 533–537. https://doi.org/10.1093/jjco/hyg093

Parkkinen, V. P., & Williamson, J. (2020). Extrapolating from Model Organisms in Pharmacology. In A. LaCaze & B. Osimani (Eds.), *Uncertainty in Pharmacology: Epistemology, Methods, and Decisions* (pp. 59–78). Springer Nature. https://doi.org/10.1007/978-3-030-29179-2

Parkkinen, V. P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B., & Williamson, J. (2018). *Evaluating evidence of mechanisms in medicine: Principles and procedures*. Springer.

Pérez-González, S., & Rocca, E. (2022). Evidence of biological mechanisms and health predictions: An insight into clinical reasoning. *Perspectives in Biology and Medicine*, *65*(1), 89–105.

Prasad, V. (2014). The Withdrawal of drugs for commercial reasons: The Incomplete Story of Tositumomab. *JAMA Internal Medicine*, *174*(12), 1887–1888. https://doi.org/10.1001/jamainternmed.2014.5756

Qureshi, Z. P., Seoane-Vazquez, E., Rodriguez-Monguio, R., Stevenson, K. B., & Szeinbach, S. L. (2011). Market withdrawal of new molecular entities approved in the United States from 1980 to 2009. *Pharmacoepidemiology and Drug Safety*, *20*(7), 772–777. https://doi.org/10.1002/pds.2155

Rocca, E. (2016). Bridging the boundaries between scientists and clinicians—mechanistic hypotheses and patient stories in risk assessment of drugs. *Journal of Evaluation in Clinical Practice*, *23*(1), 114–120. https://doi.org/10.1111/jep.12622

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170. https://doi.org/10.1080/02698590701498084

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *Bmj*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Sevigny, J., Chiao, P., Bussière, T., Weinreb, P. H., Williams, L., Maier, M., Dunstan, R., Salloway, S., Chen, T., Ling, Y., O'Gorman, J., Qian, F., Arastu, M., Li, M., Chollate, S., Brennan, M. S., Quintero-Monzon, O., Scannevin, R. H., Arnold, H. M., & Sandrock, A. (2016). The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature*, *537*(7618), 50–56. https://doi.org/10.1038/nature19323

Shaw, A. T., Yasothan, U., & Kirkpatrick, P. (2011). Crizotinib. *Nature Reviews Drug Discovery*, *10*(12), 897–898. https://doi.org/10.1038/nrd3600

Shaw, A. T., Kim, D. W., Nakagawa, K., Seto, T., Crinó, L., Ahn, M. J., De Pas, T., Besse, B., Solomon, B. J., Blackhall, F., Wu, Y. L., Thomas, M., O'Byrne, K. J., Moro-Sibilot, D., Camidge, D. R., Mok, T., Hirsh, V., Riely, G. J., Iyer, S., & Jänne, P. A. (2013). Crizotinib versus Chemotherapy in Advanced *ALK* -Positive Lung Cancer. *New England Journal of Medicine*, *368*(25), 2385–2394. https://doi.org/10.1056/NEJMoa1214886

Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., Bando, M., Ohno, S., Ishikawa, Y., Aburatani, H., Niki, T., Sohara, Y., Sugiyama, Y., & Mano, H. (2007). Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature*, *448*(7153), 561–566. https://doi.org/10.1038/nature05945

Soda, M., Takada, S., Takeuchi, K., Choi, Y. L., Enomoto, M., Ueno, T., Haruta, H., Hamada, T., Yamashita, Y., Ishikawa, Y., Sugiyama, Y., & Mano, H. (2008). A mouse model for *EML4-ALK* -positive lung cancer. *Proceedings of the National Academy of Sciences*, *105*(50), 19893–19897. https://doi.org/10.1073/pnas.0805381105

Stephens, T., & Brynner, R. (2001). *Dark remedy: The impact of thalidomide and its revival as a vital medicine*. Basic Books.

Sung, D., & Holman, B. (2023). Against evidentiary pluralism in Pharmaceutical Regulation. *Philosophy of Science*, 1–16. https://doi.org/10.1017/psa.2023.40

Tampi, R. R., Forester, B. P., & Agronin, M. (2021). Aducanumab: Evidence from clinical trial data and controversies. *Drugs in Context*, *10*, 1–9. https://doi.org/10.7573/dic.2021-7-3

Teira, D. (2020). A defence of Pharmaceutical Paternalism. *Journal of Applied Philosophy*, *37*(4). https://doi.org/10.1111/japp.12413

Temple, R. (1995). Development of drug law, regulations, and guidance in the United States. In P. L. Munson, R. A. Muller, & G. R. Breese (Eds.), *Principles of pharmacology: Basic concepts and clinical applications* (pp. 1643–1663). Chapman and Hall.

Vokinger, K. N., Hwang, T. J., Glaus, C. E. G., & Kesselheim, A. S. (2022a). Therapeutic value assessments of Novel Medicines in the US and Europe, 2018–2019. *JAMA Network Open*, *5*(4), e226479. https://doi.org/10.1001/jamanetworkopen.2022.6479

Vokinger, K. N., Kesselheim, A. S., Glaus, C. E. G., & Hwang, T. J. (2022b). Therapeutic value of drugs granted accelerated approval or conditional marketing authorization in the US and Europe from 2007 to 2021. *JAMA Health Forum*, *3*(8), e222685. https://doi.org/10.1001/jamahealthforum.2022.2685

Williamson, J. (2019). Establishing Causal claims in Medicine. *International Studies in the Philosophy of Science*, *32*(1), 33–61. https://doi.org/10.1080/02698595.2019.1630927

Worrall, J. (2007). Why there's no cause to randomize. *The British Journal for the Philosophy of Science*, *58*(3), 451–488.

Zhang, A. D., Schwartz, J. L., & Ross, J. S. (2019). Association between Food and Drug Administration Advisory Committee Recommendations and Agency actions, 2008–2015. *The Milbank Quarterly*, *97*(3), 796–819. https://doi.org/10.1111/1468-0009.12403