

## **Testing oncological treatments in the era of personalized medicine**

David Teira

Dpto. de Lógica, Historia y Filosofía de la ciencia | UNED

Senda del rey 7, 28040 Madrid (Spain)

### *Abstract*

Should conventional randomized clinical trials provide the standard of safety and efficacy when testing targeted treatments for cancer? Should we make amendments to our current regulatory standard, stick to it, or dispense with it? I am going to maintain that, under certain circumstances, smaller phase II trials provide good enough grounds to grant regulatory approval for targeted therapies. My argument will hinge on the size of trial population, showing how this size is important not only for scientific considerations, but also for ethical and political reasons. The current system was designed to provide massive consumer protection at a point when our understanding of the biology of cancer was still relatively poor and statistical tests gave the only solid evidence about treatment effects. With targeted therapies, risks are hedged in a way that allows patients (if well informed) to make decisions for themselves, instead of deferring on pharmaceutical regulators.

*Keywords:* small trials, phase II trials, cancer, targeted treatments

# Testing oncological treatments in the era of personalized medicine

## 1. Introduction

For the last five decades, medical treatments have been tested by pharmaceutical regulators with **randomized clinical trials (RCTs)**. These are large comparative experiments in which an experimental therapy is compared with the standard alternative (or a placebo) according to a pre-defined statistical design. **Regulatory agencies**, such as the **American Food and Drugs Administration (FDA)** or the **European Medicines Agency (EMA)**, require two positive (*phase III*) trials as proof of the safety and efficacy of a treatment before patients are granted access to it<sup>1</sup>. Phase III trials are large, often involving thousands of patients. But, as this volume illustrates, molecular medicine is changing our very concepts of disease and cure and, as we are going to defend here, it forces us to rethink the sort of regulatory standard that we expect treatments to meet in order to consider them safe and effective. Let us illustrate it with a recent episode of current research on **cancer**.

A molecular diagnostic of the genetic aberrations in each individual tumor opens the door for **targeted treatments**: drugs that selectively inhibit the products of these altered genes (**Schilsky 2014**). There are about a dozen such drugs available (**Tursz and Bernards 2015**) and many more should come. As **Tsimberidou, Ringborg et al. 2013** contend, these molecular diagnostics pave the way to truly personalized treatments: e.g., 26 of 32 (81%) melanoma patients bearing the V600E BRAF mutation had responded to a treatment based on the BRAF inhibitor **vemurafenib**; 47 of 82 (57%) patients with ALK-rearranged non-small cell lung cancer responded to the ALK inhibitor **crizotinib**. However, as the figures in parentheses above show, we are often speaking of evidence that comes from very few patients, as compared to phase III trials. Oncological treatments have been so far tested, like any other drug, in RCTs in which patients are usually not selected according to their genotypes.

---

<sup>1</sup> Clinical trials are usually divided in four phases: in the first one there are experiments that seek the correct dosage, pharmacokinetics, etc.; in the second one we find trials with a small number of patients designed as pilots for the third phase trials, on which regulatory agencies ground their approval decisions. In the fourth phase, there is pharmacovigilance of the actual use of the approved treatment in the market. For a quick overview, see (**Hackshaw 2009**).

In addition, these phase III trials are not just large, but also *long*: they involve comparing a treatment with a standard alternative, following patients to a predesignated endpoint after the administration. This is usually a point in time, measured from the start of the trial: ideally, it should be *overall survival*, for how long patients who receive the treatment are still alive. In cancer, this endpoint is often five years and reaching them for a large number of patients takes time. Hence, RCTs for conventional cancer treatments are necessarily slow. But cancer patients cannot wait.

Testing targeted therapies thus poses an epistemic dilemma for pharmaceutical regulators: should they stick to large and long trials, when there are so few patients to test these treatments? Or should they decide on the basis of quicker tests? By way of motivation, we can illustrate this dilemma with two stories, of success and failure. For some new targeted therapies, the effects observed in early studies led the FDA to grant accelerated regulatory approval. For example, in 2000 **gemtuzumab ozogamicin** was approved for the treatment of CD33 positive acute myeloid leukemia in first relapse. The evidential basis for this decision was provided by small tests: three trials on 142 subjects with the required mutation (*phase II* studies). However, it was apparently correct (Tsimberidou et al. 2009): more than a decade later, it is still in the market as a treatment for the same condition it was originally approved for.

In contrast, consider **gefitinib**, a drug initially approved by FDA, for the treatment of advanced non-small-cell lung cancer (after failure of standard chemotherapy). **Gefitinib** was granted accelerated approval in May 2003, on the basis of the tumor response rate, a surrogate endpoint for clinical efficacy. This endpoint should allow us to predict patients' survival. But in 2005 the FDA withdrew its approval for use in new patients, after the completion of a large RCT comparing **gefitinib** against placebo on unselected patients with overall survival as its endpoint. It showed no evidence that the former extended patients' life. A subsequent analysis of this randomized study demonstrated that patients with EGFR mutations had higher response rates than patients without EGFR mutations (37.5% vs. 2.6%). Yet only 26 of the former received **gefitinib** in the trial (Tsimberidou et al. 2009).

These two stories motivate the questions that I am going to address in this chapter. Should conventional phase III trials provide the standard of safety and efficacy for targeted therapies? Should we make amendments to this standard, stick to it, or dispense with it? I am going to maintain that, under certain circumstances, smaller

phase II trials provide good enough grounds to grant regulatory approval for targeted therapies. My argument will hinge on the size of trial population, showing how this size is important not only for scientific considerations, but also for ethical and political reasons. The current system was designed to provide massive consumer protection at a point when our understanding of the biology of cancer was still relatively poor and statistical tests gave the only solid evidence about treatment effects. Nowadays, what has shifted is not only medical practice, due to the molecular turn, but also the way in which we deal with the uncertainty of medical treatments. With targeted therapies, risks are hedged in a way that allows patients (if well informed) to make decisions for themselves, instead of deferring on pharmaceutical regulators.

In the following section, I will briefly examine how the size of the population contributed to ground our current regulatory consensus. In section three, I will discuss how the trial of targeted cancer therapies challenges this consensus, requiring smaller trials. In section four, I will present three approaches to this challenge: a “reformist,” a “revolutionary,” and a “critical” one. In the fifth and final section, I will argue that dealing with smaller well-defined populations provides a good normative ground for impartial regulatory decisions in which patients decide their tolerance to the risks involved in targeted therapies.

## **2. The size of the trial population: why does it matter?**

Clinical trials study the effects of a treatment on a population of patients that may potentially benefit from it. The characteristics of this **population** of patients are defined in the eligibility criteria that grant admission in the trial. We test the treatment on a random sample drawn from this population, under the assumption that if the treatment works on this sample, it will do the same for every other member of the population. This assumption is grounded on the statistical design of the trial, from a frequentist standpoint (Teira 2011): we need a sample size that guarantees that there is only a very small chance of observing a statistical fluke, a treatment effect due to the particular characteristics of the patient sample that will not reappear if the trial is further replicated. In the assessment of the methodological quality of a trial, a proper calculation of its **statistical power** reliably detect a true treatment effect is considered a plus.

For most methodologists, the statistical debate on the trial population ends here.<sup>2</sup> The size of the trial population has only received attention in the context of rare diseases (Tudur Smith, Williamson, and Beresford 2014): when a “life-threatening or chronically debilitating” condition affects about 1 in 2000 people, we may not find enough patients for a properly powered trial. In these cases, the European Medicines Agency may relax its constraints on the levels of evidence required to grant regulatory approval, according to given guidelines. Different randomized designs exist for gathering causal data with a limited number of patients<sup>3</sup>, but from a regulatory standpoint, such designs provide just a second best, the exception to an approval rule based on statistical power.

In sum, our regulatory system is built on the assumption that most treatments will target big enough populations of patients, on which large trials are possible. But how big is “big enough”? In my view, we need to go beyond statistical methodology in order to adequately answer this question<sup>4</sup>. Regulatory standards of proof are not exclusively built upon epistemic principles. In order to deserve regulatory protection, populations should be also big according to financial and political standards as I am going to show.

On the one hand, the population of potential patients should be big enough to secure the financial viability of the drug development plan. Even if the actual cost of bringing a drug to the market is disputed, the figures are big enough to require substantial sales to make up for the investment: a recent rough informal estimate (by a journalist covering the industry<sup>5</sup>) put it at \$350 million the cost of launching a single drug over the last decade (2003-2013). With more drug approvals (between eight and thirteen), the cost may reach \$5.5 billion. Yet, it is not size alone that affects costs: Treatments for ultra-rare diseases may be occasionally lucrative if someone is wealthy enough to pay for them as much as \$200,000 per patient per year. Lots of neglected diseases lack a cure because patients do not have the resources to fund their treatments.

---

<sup>2</sup> Except, perhaps, for some principled Bayesians who consider trial populations an abstract entity from which we are not actually sampling. The patients in the sample usually share more traits than those explicitly stated in the eligibility criteria: their geographical location, socio-economic status, etc. See (Urbach 1993) for a quick discussion.

<sup>3</sup> For an updated discussion, check out the website of the European Union funded research project “Integrated design and analysis of small population group trials”: <http://www.ideal.rwth-aachen.de/> (accessed on September 9<sup>th</sup>, 2015)

<sup>4</sup> See (Edwards et al. 1997) for another take on this same problem. We owe this reference and a fruitful discussion of the topic to Cecilia Nardini.

<sup>5</sup> Matthew Harper (2013), “How Much Does Pharmaceutical Innovation Cost? A Look At 100 Companies”, Forbes/Pharma & Healthcare, <http://www.forbes.com/sites/matthewharper/2013/08/11/the-cost-of-inventing-a-new-drug-98-companies-ranked/> (accessed on September 9<sup>th</sup>, 2015)

The philosophical debate on alternative ways of funding biomedical research in order to address neglected diseases shows that, for drug development purposes, a population should properly combine size and purchasing power to be well served by the pharmaceutical industry –see (Reiss and Kitcher 2010) for a survey.

As to the political standards that populations of patients should meet, we should recall that pharmaceutical agencies like the EMA or the FDA are regulatory bodies established for one major goal: consumer protection. Such agencies require political support and the number of required votes depends not only on the number of consumers protected, but also on the sympathy they may elicit in their fellow citizens. For instance, (Carpenter 2010) has shown that the FDA could only strengthen its power and demand stronger tests when the Congress was persuaded that the victims of the 1930s sulfanilamide and the 1960s thalidomide tragedies were *influential* enough to deserve increased protection. Sulfanilamide was an antibacterial compound to treat streptococcal infection that, in the late 1930s, was marketed in a toxic solution that caused more than 100 deaths in the United States. The supporters of granting stronger powers to the FDA framed the scandal in terms of the group of most likeable victims: white, virginal kids avoiding any mention of the black, male, and possibly sexually licentious consumers of sulfanilamide. The thalidomide tragedy affected mostly pregnant women, whose babies suffered phocomelia as a result of the ingestion of the sedative. Were it not for such influential groups of patients, the FDA may have not reached the level of regulatory powers it now enjoys.

In sum, when, in the 1960s, the FDA adopted randomized clinical trials as regulatory yardsticks to judge the safety and efficacy of medical treatments, there was an implicit twofold assumption about the populations targeted by such treatments: their members were numerous and politically significant enough as to deserve administrative protection. Hence, size, qualified by and political influence, matters in pharmaceutical regulation.

The connection between these methodological and regulatory constraints is deeper than it may initially seem (Teira 2014). Scientists want trials to be *unbiased*, because they seek the truth about a treatment effect, uncontaminated by systematic interferences arising from the preferences of the participants in the experiment. The regulator wants to control for these preferences as well, because she needs the trial to be impartial regarding the interests in conflict about the tested treatments. The producer, its

competitors, the medical community, patients, health-care providers usually do not share their expectations about the outcome of a trial. Yet, they should all agree on the outcome and the **impartiality** of the experiment (regarding their conflicting interests) is a pre-requisite for their agreement.

For instance, nobody would agree on the outcome of a trial in which the sponsor had fiddled with the trial population in order to achieve a given outcome. If we are looking for a statistically significant difference between two treatments, one way to enlarge this difference is to delete the data of patients on which the treatment did not have a big enough effect. Such a manipulation can be justified with a revision of the eligibility criteria implemented, declaring *ex post* those low-effect patients not eligible for inclusion (**Gøtzsche** 2013). In order to control for this sort of manipulation, the eligibility criteria should be fully specified *ex ante* and the raw data registered in a publicly accessible database, so that everybody can verify that no patient has been lost for illegitimate reasons. A proper definition of the trial populations is a pre-requisite for an unbiased and impartial trial, both in the (public) interest of the experimenter and the regulator.

Summing up, from a methodological standpoint, frequentist clinical trials are scientific experiments that presuppose big populations. If every disease had been rare, that is, a phenomenon of minorities, instead of majorities, RCTs might have not provided the gold standard for testing treatments. But precisely because diseases generate big clusters of patients, they are a commercially interesting target, at least in rich developed countries. And these big populations are equally necessary to explain why the commercialization of treatments has been so strongly regulated in democratic countries. The point I am making here is that our current social consensus on RCTs as regulatory devices hinges on the confluence of methodological, commercial and political approaches to the phenomenon of curing sickness for big numbers of people. With the possibility of individuating disease at the scale of the individual patient, these patients may cluster into significantly smaller groups, for which our consensus on RCTs as regulatory standards may not hold any further.

### **3. Re-defining cancer populations**

As of today, the size of a trial population ultimately depends on how we define the condition targeted by a treatment, i.e., the eligibility criteria to enter a trial, and these depend in turn on our clinical and scientific understanding of such condition. Take for

instance, the current controversy on statins and stroke prevention: depending on how we define the “healthy” levels of cholesterol, the number of patients who may benefit from the treatment may significantly increase – by the millions (González-Moreno, Saborido, and Teira 2015). At the time phase III trials became a regulatory yardstick, in the 1960s, our grasp of the biological mechanisms by which a drug cured a disease like cancer was often poor. Hence, the population targeted could be defined according to conflicting criteria, at least in the short run: if the aim was to test a pathobiological hypothesis under strict experimental conditions, for pure research purposes, the eligibility criteria would be more restrictive; if the trial was to test the effectiveness of a given treatment in regular clinical practice, the population would be more heterogeneous. In the former case, we are conducting, e.g., basic research on a set of malignant cells without direct clinical implications, since cancer in real patients is a more complex pathology. When we try to find out how to treat these latter, we need to take into account eligibility criteria that capture such complexity. Nonetheless, according to some historians (Keating and Cambrosio 2012)), from 1970s onwards the organization of large multicenter cancer trials allowed the gathering of big sample sizes bringing the two approaches closer. In other words, the trial protocols created “criteria that attempted to generate homogeneous patient populations with regard to a constantly growing number of significant variables concerning response to therapy and the evolution of the disease under study” (Keating and Cambrosio 2012). Different treatment regimens could be thus tested with enough statistical power to detect significant effects. Bigger trials brought about a better understanding of the biology of cancer and a more precise definition of the target populations of its different treatments. From the 1970s onwards, the size of the populations targeted in cancer trials met the three requirements stated above: the eligibility criteria were scientifically sound enough; the number of patients targeted was big enough to deserve commercial and regulatory attention.

However, for the last two decades, the genomic revolution in biomedical research has challenged the equilibrium of methodological, commercial and regulatory considerations about population sizes. The possibility of targeting therapies according to genetic biomarkers involves, first, a change in the very definition of a trial population: we can now clearly define who is a potential participant, with a perfect match of clinical and biological criteria. According to the Biomarkers Definitions Working Group (group 2001), a biomarker is “a characteristic that is objectively measured and evaluated as an



indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”. Following (Buyse et al. 2011), we should further distinguish between *prognostic* and *predictive* biomarkers. Whereas the former, predict the likely course of disease in a defined clinical population, irrespective of treatment, the latter forecast instead the likely response to treatment. Assuming that a biomarker is properly *validated*, from a biological and statistical standpoint, we can use it in a trial to allocate targeted treatments to the patients who may benefit most from them.<sup>6</sup> In a targeted trial design, only biomarker-positive patients are randomized. As Buyse et al. (2011) point out, “such trials have the capacity to confirm the usefulness of the marker in identifying a population in which there is a treatment benefit” (although they provide no information regarding the lack of benefit among marker-negative patients).<sup>7</sup> Granting access to a trial according to the presence or absence of a *validated biomarker* is as rigorous, if not more, than any of our current eligibility criteria. The patient population in cancer trials has been usually defined in terms of the organ of origin, the extent of disease and the previous treatment history. These latter two variables are in principle more open to interpretation, and therefore bias, than a biomarker assay. However, reliability comes at a price for the patient and this will impact on the trial recruitment process (de Gramont et al. 2015). The amount of tissue needed might make the biopsy more or less difficult to bear to some of them. And it might happen, of course, that the screening reveals that the patient presents a target for which there is no effective therapy. Finally, even a reliable test does not guarantee that even the patients who present a given molecular alteration at the beginning of the treatment will all equally react to the therapy: the heterogeneity of tumors might generate various degrees of resistance in each individual –see Boniolo, this volume

Leaving aside the ethical issues regarding the sort of consent patients should grant, we should notice how crucial this consent is in order to reach an appropriate *sample size*. If patient accrual was already a problem in regular cancer trials (fewer than 5% of adult patients with cancer participate), the situation gets worse in targeted trials, since only a sub-sample of the tested patients will turn out to be eligible for the study.

---

<sup>6</sup> How to carry out this validation is a controversial topic in itself. We will assume, for the sake of the argument that we can have properly validated biomarkers. Without them, the *reformist* and *revolutionary* positions we will examine in the next section become, in my view, untenable.

<sup>7</sup> Again, we are going to focus on pre-treatment biomarkers, measured prior to initiation of therapy, that allow us to estimate the drug efficacy for a particular class of patients. There is an increasing advocacy for post-treatment predictive biomarkers, but we won't discuss it here: see, e.g., (Stone and Schmitt 2014)

(Rodon et al. 2012) cite three identification strategies that might alleviate the problem: first, performing a retrospective screening for certain biomarkers once a regular trial is concluded; second, pre-screening patients' tissue before they are considered for inclusion in the trial; finally, we might screen patients who are receiving standard treatment. In the first strategy, a high number of patients may be put at risk of exposure to a study drug despite not presenting the target of interest. In the other two, their tolerance to a biopsy will be crucial.

In sum, populations become *de facto* smaller, since we are not targeting anymore the undifferentiated cancer patient of previous eras. From a methodological standpoint, these smaller populations challenge the possibility of conducting properly powered trials. These smaller trials have though a clear antecedent in current trial design. As noted above, a clinical trial is a research process conducted in four stages (see footnote 1). Targeted therapies only admit a phase II trial, since they run short of patients for a properly powered phase III experiment. Hence, we face the challenge of assessing the **safety and efficacy** of targeted therapies on the basis of an experiment that might not conclusively capture the true outcome of the treatment for lack of a large enough sample.

We will set aside the discussion of the commercial implications of the redefinition of cancer brought about by the genomic revolution, since producing targeted treatments according to genetic profiles requires an entire different financial outlook for the pharmaceutical industry. Suffice it to say that they are willing to invest in it. We will focus instead, for the rest of the paper, on the challenge faced by the regulator: under which conditions smaller phase II targeted trials are acceptable as a proof of efficacy and safety? Do we need them at all? Let us spell out these methodological and normative challenges in more detail.

#### **4. Three approaches to the regulation of targeted therapies**

We are going to distinguish three approaches to the regulatory use of small phase II trials for targeted therapies: reformists, revolutionaries and critics. The *reformists* accept the current regulatory system, but argue that under certain circumstances pharmaceutical agencies can make exceptions and grant market access without phase III trials. *Revolutionaries* advocate for a radical reform of drug regulation in order to exploit the full potential of biomarkers. *Critics* question that biomarkers have

provided so far any evidential grounds for a reform, moderate or radical. We will present these three positions, providing our own argument for a moderate reform.

Starting with the reformists, (Sharma and Schilsky 2012) argue, for instance, that there is evidence that we can make good regulatory decisions without top quality evidence. The FDA has approved 31 oncology drugs between 1973 and 2006 without properly randomized trials and with a median number of patients per drug approval of 79. Had this decision been mistaken, the FDA would have later withdrawn the drugs from the market for lack of safety and efficacy. But 29 of these 31 treatments are still on the market.

Sharma and Schilsky (2012) argue that, in a targeted therapy tested on a small but well-validated sample, it is worth foregoing phase III trials if: “the response rate and average response duration should indicate a clinically meaningful improvement over that which would be expected based on historical data for the existing standard of care in the same subset of selected patients”, provided that these two outcome measures are interpreted in the context of the disease setting and there is no life-threatening safety concern about the therapy. The assumption in this argument (let us call it *assumption  $\alpha$* ) is that this meaningful improvement, in the context of a cancer targeted therapy, is more likely to be caused by the action of this therapy on the cellular signaling pathways altered in malignant cells than by mere chance. With conventional therapies, if a small trial detected a large effect, we could not conclusively tell whether it was actually caused by the treatment or by a random coincidence: e.g., the particular sample of patients it was tested on. With targeted therapies, our causal understanding of the biology of the tumor allows us to explain why such a large effect has risen in such a small group of patients. In other words, our biological background knowledge becomes as good as the statistical evidence provided by a phase III trial for establishing the safety and efficacy of a treatment.

However, Sharma and Schilsky do not question the normative inspiration of our current regulatory system: the protection of future pharmaceutical consumers is worth the costs of delaying the introduction of new treatments until their safety and efficacy is shown by a standard phase III trial. These costs are mainly the treatment opportunities that current patients lose for not having early access to untested treatments, provided they were willing to take the risks. In this regard, our current pharmaceutical regulation is clearly *paternalistic*: agencies such as the FDA interfere with the liberty and/or

autonomy of individual patients, without their consent, for the sake of the patients' health. The social drive behind this normative position seems to be our fear of toxicity scandals, discussed above: the actual victims of the 1930s sulfanilamide and the 1960s thalidomide tragedies are politically more relevant (in terms of pushing forward a paternalistic regulation) than the patients who suffer for lack of access to untested treatments that may potentially benefit them (Wardell and Lasagna 1975). Could we arrange our regulatory system in a non-paternalistic manner that gives these latter patients a chance?

This is an old debate that has been revived in our current controversy on the trial of targeted therapies, where anti-paternalist critics are making a comeback, advocating for a revolution in our regulatory system (Stewart, Whitney, and Kurzrock 2010) (Stewart and Kurzrock 2013). On the one hand, simulations allow us to estimate how many agents discarded in a standard phase III trial on unselected populations could have been shown effective in small targeted subpopulations. Hence, the old "lack of access" argument is now supported on a solid counterfactual: patients with a given genotype are losing access to treatments they could have benefitted from only because these latter are tested on the wrong populations. Phase III trials would be now picking the most common target as winner, not the best drug for each subpopulation: they would more likely detect a very small advance affecting a high proportion of patients than to detect a very large advance affecting a small proportion of them. According to this argument we would not only be losing effective treatments, but we would be spending more than we should both in running standard phase III trials and in delivering effective care according to the patients real needs.

This is the position of our revolutionaries: e.g., Stewart, Whitney and Kurzrock argue for a different regulatory system for lethal diseases such as cancer, in which higher level of risk than for benign and nonlethal diseases are accepted. Unlike in the previous case the argument is not only methodological (about the evidential grounds for properly testing treatments, drawing on the assumption  $\alpha$ ), but also normative: instead of protecting consumers as if they were an undifferentiated population, we should personalize the protection according to their condition, genetic profile and risk aversion. This position combines libertarian and individualist intuitions. As to the latter, the protection offered by regulatory agencies should not be judged in principle, but rather by its actual output: we should assess the safety and efficacy of treatments not on

average populations but on genetic profiles as close as possible to the actual patient. As to the former, since we are now dealing with individuals, they should have their say on the degree of risk they are willing to tolerate: as I mentioned at the beginning of this section, with their consent we may have smaller and quicker targeted phase I/II trials that might make phase III trials dispensable.

Therefore, the reformist and the revolutionary concur in promoting a different evidential standard for the assessment of at least certain treatments. Where they differ is in the normative goals of regulation: whereas the reformist defends exceptions to a general paternalistic approach, considered on a case by case basis, the revolutionary defends a different regulatory approach to a whole class of treatments in which the patients should be protected in an individualized manner.

Before we discuss these two positions, it is useful to consider a critical stance regarding them both. Reformists and revolutionaries accept what I have called assumption  $\alpha$ . In standard RCTs we can remain agnostic as to the causal mechanisms behind the tested treatments; we just rely on the statistical power of the test to detect the true difference between the effects of both therapies. The power of a trial to detect what a treatment really does depends crucially on the size of the sample. If the number of patients is not big enough, we may be unable to differentiate a true treatment effect from a random spike that may disappear once the sample size grows. Reformists and revolutionaries assume  $\alpha$  instead: our superior understanding of the biological underpinnings of a targeted therapy would allow us to distinguish true effects from random spikes even with small samples. John Ioannidis and various coauthors have been challenging this assumption in a series of papers: in a relevant sample of targeted trials (those reported in highly cited papers) the effect estimates for postulated associations are larger in the trial outcome than in subsequent meta-analyses evaluating the same associations. In other words, the effect disappears as the sample size grows.

According to Ioannidis and his coauthors, this is not a problem just with targeted therapies but with all sorts of very large treatment effects of medical interventions (Pereira, Horwitz, and Ioannidis 2012). They highlight two major points in their analysis. First, these very large effects typically become smaller or lose their statistical significance once additional evidence is obtained, since they usually arise in small trials with few events. According to (Ioannidis 2008), biomedical researchers tend to claim discoveries based exclusively on p-values, disconnecting significance from statistical

power. But those statistically significant outcomes are difficult to reproduce without the backup of a proper sample size.

Moreover, following still (Pereira, Horwitz, and Ioannidis 2012), statistically significant outcomes may not have a clear clinical interpretation. As I mentioned in the introduction, the ideal endpoint of a cancer treatment should be survival, but it takes a lot of time to follow patients for years. This is why we consider alternative endpoints: according to (Pignatti et al. 2015), “there has been a tendency to recognize progression-free survival as a clinical benefit endpoint in itself, leading to standard approvals”. I.e., the time during and after the treatment of the disease in which the patient lives with it without getting worse. This surrogate point, though, is always context-dependent: what are the expected clinical benefits of delaying progression and how big should they be? And, again, in order to judge this size, we need to a proper sample size.

According to Ioannidis, such large effects (at least for mortality conditions) are “exceedingly rare”. Targeted therapies will not be an exception: as (Ioannidis and Khoury 2013) put it: “Most of the emerging genomic information that is meandering its way toward health applications is still either non-validated noise or true signals with validated small effects, which are not suitable for applying to clinical practice.” Against reformists and revolutionaries, Ioannidis and coauthors argue first that assumption  $\alpha$  has yet to receive statistical confirmation in properly powered studies. Without it, small targeted trials do not provide firm enough grounds for unbiased regulatory decisions. Hence, the normative case for a change in our regulatory system rests on purely theoretical conjectures and, as (Ioannidis and Koury) argue, we might better improve it investing in larger conventional RCTs, with reliable measures and clinically relevant outcomes.

## 5. Where do we stand?

The three approaches presented above are just a snapshot of an ongoing debate in a field in very quick progress: the sort of biomarkers discussed above is just a first step in the personalization of oncology, clustering patients according to a particular genetic aberration depending on the organ affected (Boniolo, Boem, and Pavelka 2015). But there is not just inter-patient variability, but also intra-patient intra-tumor heterogeneity, and the very pressure of the treatment on the affected cells, making them evolve, may question the value of a single sample to capture the complete genomic

landscape of a patient's cancer (Dienstmann, Rodon, and Taberero 2015). In the words of Donald Berry (2015), "soon every cancer patient will have an ultra-orphan disease". If personalized medicine keeps progressing, treatments will target smaller and smaller populations, forcing us to reconsider our experimental standards for safety and efficacy. Citing again (Berry 2015), "large clinical trials in narrowly defined diseases are impossible". For the time being, sample size is being reached through alternative trial designs for which the terminology is still evolving –see e.g., (Ocana et al. 2013)<sup>8</sup>.

However, even if these designs allow us to increase our samples, they may do it slowly enough to put regulatory agencies in a difficult position regarding the approval of targeted therapies for cancer. When and on which basis should they make their decision? I think that a *reformist approach* to the regulatory use of targeted trials is defensible. On the one hand, I think that the critics are correct from a purely *methodological* standpoint: we need large phase III trials in order to grasp conclusively the true effects of targeted therapies. On the other hand, the revolutionaries are right from a *normative* standpoint: with targeted therapies, there is no need for pharmaceutical paternalism. We can combine these two points in a reformist approach as follows: regulatory agencies should make exceptions and approve targeted therapies on the basis of small phase II trials, with two provisos. First, we need to conduct larger trials in order to validate the decision (*pace* revolutionaries). Second, those targeted therapies should only be administered to patients with the proper biomarkers who are informed about the uncertainty about the treatment (*pace* critics). Is this third way tenable?

Let us first take stock of the discussion so far. Our current consensus on the size of regulatory RCTs is grounded on a combination of scientific, commercial and normative considerations. As to the former, we started testing cancer treatments at a point, forty years ago, in which our causal understanding of the disease was often poor and there was a huge element of chance in finding treatment with large effects. Today, our understanding of cancer is solid enough to make assumption  $\alpha$  compelling for a great number of cancer scholars. It is not statistically validated yet, since we have not observed a big enough number of large effects in smaller trials that did not vanish in larger ones. So, in this particular regard it is not wise to re-arrange part of our regulatory

---

<sup>8</sup> A major contender to overcome the dilemma of size is the Bayesian approach to clinical trials: see (Berry 2012) for a review and (Teira 2011) for a discussion.

system on the basis of smaller trials alone. Here I stand with the critics: we still need large phase III trials in order to reach a final conclusion as to the safety and efficacy of treatments.

However, if we agree on the quality of the basic science and on its potential for real pharmaceutical innovation, we need to consider what incentives would make the industry invest in such a risky business as targeted therapies. Investing in drugs for smaller populations will only make sense if the effect is large enough to catch the attention of consumers (and third-party payers), but we are not seeing such effects yet. If we want the industry to invest in targeted therapies, we need indeed quicker trials that might separate winners and losers early on, and the price to pay is, of course, to see some false positives go through. Eventually some unexpected adverse effects may harm patients.

This is primarily a normative question: are we willing to take the risks involved in the development of targeted therapies? Our current consensus on the undesirability of adverse effects dates back from the 1960s, but, in my view, there are enough grounds to revise it, at least when it comes to the regulation of targeted treatments. On the one hand, we are not dealing anymore with the protection of big size populations such as the victims of the **Thalidomide** scandal (potentially any pregnant woman requiring a sedative). As I have argued above, we are now focusing on patients with a given genetic profile identifiable with reasonable precision. And, by definition, these patients cluster in increasingly smaller groups. Targeted therapies will not pose massive public health threats; or, at least, we might identify how big is the mass, according to the DNA profile, and require bigger trials if necessary. From a political standpoint, the risk of a toxicity scandal (even if not always correctly estimated in targeted therapies) is smaller and manageable, at least if we stick to the following two principles. First, we should only give access to targeted therapies to the patients who have the proper biomarkers to benefit from them. Second, they should provide their informed consent: they should know that our understanding of the disease pathways is good enough to expect them to benefit from the treatment, even if we lack conclusive statistical evidence about it.

In this respect, *I think the revolutionaries are correct*: there is room for relaxing our paternalistic approach to pharmaceutical regulation and leave patients can consent either to take part in a trial or receive treatment tested in a small one according to their own degree of risk aversion. However, in exchange for this access, the liabilities that



may arise from unanticipated adverse effects should be negotiated in advance. If an adverse effect occurs, patients may not be able to sue the physicians or manufacturers. It will all depend on the terms of the agreement they reached to start the treatment. There are limits, however, to the relaxation of pharmaceutical paternalism. As **Dan Carpenter** reminds us<sup>9</sup>, in the US “access to medicines and technological advances is an important value, but it’s neither a constitutional nor a legal right”. As long as this is the case, we are just working within the current scheme of the **FDA** for **accelerated approval**: this is just a contract between the **FDA** and a pharmaceutical company: “in return for promises of further clinical studies, the company receives provisional approval and rapid market access” (**Carpenter, Kesselheim, and Joffe** 2011) The rights of the patients are created by this contract and, if further studies contradict the initial evidence, the regulator is entitled to withdraw the approval, depriving future patients of the therapy.

Hence, it is possible to strike a reformist compromise between critics and revolutionaries: we still need large trials for methodological reasons, but there are grounds to relax our current regulatory paternalism. However, there can be no compromise about the impartiality of our regulatory trials, be they small or large. We expect regulatory agencies to act in the **public interest**. That is, they should make decisions on impartial grounds, unprejudiced by the particular interests of the industry, patients or any other stakeholder in a trial. If we are going to see more and more cases of accelerated approval, can we expect pharmaceutical regulators to preserve their impartiality?

Whereas phase III **RCTs** should ideally provide conclusive evidence as to the effects of a treatment, small phase II trials with surrogate outcomes provide much less conclusive grounds for a regulatory decision. Such uncertainty can be exploited in the interest of the sponsor, who might be more willing to take risks (for financial reasons) than any other stakeholder in the trial. My previous case applies here: on the one hand, if the patient is equally willing to take his risks, it is just a matter of negotiating liabilities under the supervision of the regulator; on the other hand, accelerated approval is only conditional and should be withdrawn if the initial decision is proven incorrect.

---

<sup>9</sup> E. **Silverman**, “Avastin & FDA were both on trial: Dan explains”, *Pharmalot*, June 30<sup>th</sup> 2011, available at: <http://people.hmdc.harvard.edu/~dcarpent/fdaproject/avastin-fda-were-both-on-trial-carpenter-explains.pdf> (accessed on September 9<sup>th</sup>, 2015)

Nonetheless, as a general principle, if large effects in small trials are false positives, we want them to be random events, not fakes engineered by spurious trial design or data analysis. There are many different sources of bias, of course, but targeted trials control a major one: the definition of the population we are dealing with through precise biomarkers. As I mentioned above, a notorious strategy of pharmaceutical disease-mongering has been to fiddle with the trial populations in order to achieve positive outcomes. When the inclusion and exclusion criteria are open to interpretation, the *ex post* elimination of a few patients at the stage of data analysis might bring about a statistically significant result. If we reach a consensus on the proper **validation of biomarkers**, there will be less room for this particular bias in targeted trials than in conventional ones.

We should open a debate on which other **biases** might creep in smaller trials, but probably the most contentious issue is their endpoint. A potential compromise in order to make smaller trials more acceptable is perhaps to focus on the hardest clinical outcomes at a first stage, until we obtain a better understanding of targeted therapies. These are more attractive in principle for patients and regulators to accept the risk.

To close, I summarize my case as follows: small phase II trials provide enough grounds for regulatory agencies to grant advanced access to targeted treatments if we observe the following principles: (i) we need to make sure that these trials are impartial; (ii) we should restrict the access to the therapies patients who have the proper biomarkers, under informed consent agreements about the possible side effects; and (iii) we need to conduct larger trials to validate the advanced access. Revolutionaries and critics will surely find this compromise objectionable, but the development of personalized medicine we need a consensus that somehow brings together the best of both approaches.

- Berry, D. A. 2012. Adaptive clinical trials in oncology. *Nat Rev Clin Oncol* 9 (4):199-207.
- Berry, D. A. 2015. The Brave New World of clinical cancer research: Adaptive biomarker-driven trials integrating clinical practice with clinical research. *Mol Oncol* 9 (5):951-9.
- Boniolo, G. , F. Boem, and Z. Pavelka. 2015. Stratification and biomedicine. How philosophy stems from medicine and biotechnology. In *The Future of Scientific Practice: 'Bio-Techno-Logos*, edited by M. Bertolaso. London: Pickering & Chatto.

- Buyse, M., S. Michiels, D. J. Sargent, A. Grothey, A. Matheson, and A. de Gramont. 2011. Integrating biomarkers in clinical trials. *Expert Rev Mol Diagn* 11 (2):171-82.
- Carpenter, D., A. S. Kesselheim, and S. Joffe. 2011. Reputation and precedent in the bevacizumab decision. *N Engl J Med* 365 (2):e3.
- Carpenter, Daniel P. 2010. *Reputation and power : organizational image and pharmaceutical regulation at the FDA, Princeton studies in American politics*. Princeton: Princeton University Press.
- de Gramont, A., S. Watson, L. M. Ellis, J. Rodon, J. Tabernero, and S. R. Hamilton. 2015. Pragmatic issues in biomarker evaluation for targeted therapies in cancer. *Nat Rev Clin Oncol* 12 (4):197-212.
- Dienstmann, R., J. Rodon, and J. Tabernero. 2015. Optimal design of trials to demonstrate the utility of genomically-guided therapy: Putting Precision Cancer Medicine to the test. *Mol Oncol* 9 (5):940-50.
- Edwards, S. J., R. J. Lilford, D. Braunholtz, and J. Jackson. 1997. Why "underpowered" trials are not necessarily unethical. *Lancet* 350 (9080):804-7.
- González-Moreno, M., C. Saborido, and D. Teira. 2015. Disease-mongering through clinical trials. *Studies in History and Philosophy of Biological and Biomedical Sciences* 51:11-18.
- Gøtzsche, Peter C. 2013. *Deadly medicines and organised crime : how big pharma has corrupted healthcare*. London: Radcliffe Health.
- group, Biomarkers definition working. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69 (3):89-95.
- Hackshaw, Allan K. 2009. *A concise guide to clinical trials*. Chichester, UK ; Hoboken, NJ: Wiley-Blackwell.
- Ioannidis, J. P. 2008. Why most discovered true associations are inflated. *Epidemiology* 19 (5):640-8.
- Ioannidis, John, and Muin Khoury. 2013. Are randomized trials obsolete or more important than ever in the genomic era? *Genome Medicine* 5 (4):32.
- Keating, Peter, and Alberto Cambrosio. 2012. *Cancer on trial : oncology as a new style of practice*. Chicago: The University of Chicago Press.
- Ocana, A., E. Amir, F. Vera-Badillo, B. Seruga, and I. F. Tannock. 2013. Phase III trials of targeted anticancer therapies: redesigning the concept. *Clin Cancer Res* 19 (18):4931-40.
- Pereira, T. V., R. I. Horwitz, and J. A. Ioannidis. 2012. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 308 (16):1676-1684.
- Pignatti, F., B. Jonsson, G. Blumenthal, and R. Justice. 2015. Assessment of benefits and risks in development of targeted therapies for cancer--The view of regulatory authorities. *Mol Oncol* 9 (5):1034-41.
- Reiss, J., and P. Kitcher. 2010. Biomedical Research, Neglected Diseases, and Well-Ordered Science. *THEORIA. An International Journal for Theory, History and Foundations of Science* 24 (3):263-282.
- Rodon, Jordi, Cristina Saura, Rodrigo Dienstmann, Ana Vivancos, Santiago Ramon y Cajal, Jose Baselga, and Josep Tabernero. 2012. Molecular prescreening to select patient population in early clinical trials. *Nat Rev Clin Oncol* 9 (6):359-366.
- Schilsky, R. L. 2014. Implementing personalized cancer care. *Nat Rev Clin Oncol* 11 (7):432-8.

- Sharma, Manish R., and Richard L. Schilsky. 2012. Role of randomized phase III trials in an era of effective targeted therapies. *Nat Rev Clin Oncol* 9 (4):208-214.
- Stewart, David J., Simon N. Whitney, and Razelle Kurzrock. 2010. Equipoise Lost: Ethics, Costs, and the Regulation of Cancer Clinical Research. *Journal of Clinical Oncology* 28 (17):2925-2935.
- Stewart, David, and Razelle Kurzrock. 2013. Fool's gold, lost treasures, and the randomized clinical trial. *BMC Cancer* 13 (1):193.
- Stone, A., and N. Schmitt. 2014. Can a treatment be licenced on the basis of post-treatment predictive biomarkers? *Pharm Stat* 13 (4):214-21.
- Teira, D. 2011. Frequentist versus Bayesian Clinical Trials. In *Philosophy of Medicine*, edited by F. Gifford. Amsterdam: Elsevier.
- Teira, D. 2014. On the impartiality of British clinical trials. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (3):412-418.
- Tsimberidou, A. M., F. Braiteh, D. J. Stewart, and R. Kurzrock. 2009. Ultimate fate of oncology drugs approved by the us food and drug administration without a randomized Trial. *J Clin Oncol* 27 (36):6243-50.
- Tudur Smith, C., P. R. Williamson, and M. W. Beresford. 2014. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol* 28 (2):247-62.
- Tursz, T., and R. Bernards. 2015. Hurdles on the road to personalized medicine. *Mol Oncol* 9 (5):935-9.
- Urbach, P. 1993. The value of randomization and control in clinical trials. *Statistical science* 12:1421-31.
- Wardell, William M., and Louis Lasagna. 1975. *Regulation and drug development, Evaluative studies 21*. Washington: American Enterprise Institute for Public Policy Research.