DISEASE-MONGERING THROUGH CLINICAL TRIALS

M. González-Moreno, C. Saborido & D. Teira

dteira@fsof.uned.es

ABSTRACT

Our goal in this paper is to articulate a precise concept of at least a certain kind of disease-mongering, showing how pharmaceutical marketing can commercially exploit certain diseases when their best definition is given through the success of a treatment in a clinical trial. We distinguish two types of disease-mongering according to the way they exploit the definition of the trial population for marketing purposes. We argue that behind these two forms of disease-mongering there are two well-known problems in the statistical methodology of clinical trials (the reference class problem and the distinction between statistical and clinical significance). Overcoming them is far from simple

1.Varieties of disease-mongering

Disease-mongering generally refers to a purported commercial strategy of the pharmaceutical industry, consisting in tinkering with the definition of a given disease (sometimes to the point of creating a new one) in order to promote the sales of one of their drugs. Disease-mongering has been featured prominently in special issues of the *British Medical Journal* (2002) or *Plos Medicine* (2006), although its existence is for some

1

controversial –and it has probably been so for more than four decades, since the earliest discussions about *medicalization* or the more current debates about *pharmaceuticalization* (Abraham 2009, 2010, Williams et al. 2009). The controversy is fuelled, of course, by the huge advertising budgets of the pharmaceutical industry and the growing influence of their marketing arms in the drug development process. It starts at its very inception, with identification of the most interesting target patient from a commercial standpoint, and it certainly conditions the way in which clinical trials for drug approval are designed, conducted and published. It is open to discussion though whether the advertising power of the pharmaceutical industry goes as far as some authors claim (e.g., Payer 1992, Moynihan *et al*. 2002). For instance, the transformation of a collection of minor medical phenomena into a treatable condition: e.g., turning baldness into a generalized anxiety process (Moynihan et al. 2002), female sexual dysfunction (Lexchin 2006) into so-called *premenstrual dysphoric disorder* (Moynihan 2003), or shyness into *social anxiety disorder* (Wolinsky 2005).

In this paper we want to articulate a more precise concept of DM. We want to show how pharmaceutical marketing can commercially exploit certain diseases when their best definition is given through the success of a treatment in a clinical trial. We will distinguish two types of disease-mongering according to the way it exploits the definition of the trial population for marketing purposes. We are going to argue that behind these two forms of disease-mongering there are two well-known problems in the statistical methodology of clinical trials and overcoming them is far from simple. But let us first introduce the discussion step by step.

Clinical trials are comparative experiments in which hypotheses on treatments are tested, usually with a methodology grounded in one particular view of probability, frequentism[1]. Namely, we judge the outcome of the experiment by assessing, e.g., the size of the observed difference between treatments with the distribution of outcomes that we would observe in an infinite series of repetitions of the trial, under the hypothesis that there is no difference between treatments. If we observe a large one, either our hypothesis is false or we have observed a very rare event.

In a frequentist approach, the probabilities of observing a given outcome are tied to one particular experimental design: if we repeat the same experiment time and again, we will observe a distribution of outcomes that will make our initial hypothesis about this distribution more or less credible. One crucial point, in making our experiment repeatable, is to define the population of patients that we are sampling in the trial. We are trying to ground an inference about the effect of the treatments in this population from the outcome we observe in the group of patients on which we are conducting the test. The probability of observing this outcome is indeed tied to a given reference class, the population of patients defined by the eligibility criteria in the trial protocol. Outside this population, the trial does not say how the treatment will work. The probability of observing a difference between treatments provides the significance of the test. If the probability is very low, the event is

---

[1] RCTs are a tool for causal inference, but the problem we analyze is created by the particular statistical rendition of RCTs that we find in medicine. Here the current regulatory standard hinges on frequentist trials. If Bayesian trials were accepted, the problem we are tackling in this paper would change dramatically: a Bayesian probability is not tied to the replication of an experiment, but rather to a degree of belief conditional on the evidence available, wherever it comes from. For a Bayesian, conducting the same experiment on different populations may yield one single probability. For a frequentist, the probability is tied to one experimental design on a given population, so we would have a different p-value whenever we change the population, even if the rest of the experiment rests the same. Hence, inasmuch as disease mongering, in our sense, depends on tinkering with populations, Bayesians and frequentists would have different types of disease-mongering. See [ANONYMIZED] for a presentation and discussion of the difference between Bayesian and frequentist trials.

rare enough to deserve a reconsideration of our initial hypothesis (there was no difference between treatments) and declare one of these treatments superior.

From a purely commercial standpoint, the industry wants any treatment to: (a) work on a given class of patients, in order to earn regulatory approval and get market access; and (b) ensure that this class is as large as possible, in order to increase sales. Pharmaceutical marketing has exploited a methodological misconception about trials that prevails among both physicians and patients. Namely, that they provide a general assessment of treatments independently of the reference class they are tested on. Hence, physicians may prescribe them off-label, assuming that a patient will benefit from them as much as the participants in the trial, even if this patient would not have been eligible.

However, sometimes the definition of the trial population is so loose that physicians can be persuaded that it would suit most patients they see. We call this *mild disease-mongering*, since it does not target the trial as such, but medical prescription based on its outcome (b). However, there is also *strong disease-mongering,* where the very definition of the patient population (a) is targeted for marketing purposes. The goal here is to find a growing group of patients where we can reach a statistically significant difference between treatments. Inasmuch as the latter is obtained, there will be grounds to get regulatory approval for the drug and sell it to this larger audience.

Why call these two marketing strategies *disease-mongering*? We are going to defend the claim that randomized clinical trials (RCTs, from now on) have provided an implicit definition of at least some diseases in terms of (successful) treatments. In the 1950s RCTs came to provide a statistical proof of the safety and efficacy of medical treatments. At this

point, physicians often did not know much about the full range of biological mechanisms by which a drug succeeded in healing individual patients. Under this veil of ignorance, RCTs provided at least statistical evidence about the safety and efficacy of the drug in a given population. Once RCTs became the regulatory standard to judge the effects of a drug, pharmaceutical research adopted an operational definition of disease that extensionally captured the group of patients targeted by the drug in the trial: a disease is just the condition cured in a trial on a given group of patients by a certain treatment.

Under such circumstances, RCTs can be used for either research or marketing purposes. As to the former, we can use RCTs to study different groups of patients on which the treatment may be effective, refining thus the working definition of the disease provided in the trial. This approach would resemble the epidemiological search for multifactorial definitions of disease (e.g., Broadbent 2011). As to the latter, RCTs become marketing tools when the definition of the population is intentionally loose, so that it can be expanded for commercial rather than clinical purposes. However, if we judge the trial protocol alone, it is difficult to tell whether it is mainly conducted for marketing or research purposes: after all, they may well overlap.

The only conclusive evidence for the true intentions of the industry in sponsoring a trial is often found in confidential documents that become publicly accessible in court, in the case of litigation over a treatment[2]. Short of this, evidence of disease-mongering is always indirect and open to debate. By way of illustration, let us consider the comparison treatment

---

[2] For instance, Sismondo 2009 has made this point about ghost-writing practices in the pharmaceutical industry.

in a trial, which is often considered a reliable sign of the intentions of the industry. But the interpretation of this comparison is, of course, controversial.

For instance, Pierre Azoulay (2002) has suggested distinguishing between RCTs as *market-expanding science* if they use a placebo or any active substance other than the antiulcer drugs competing in the market under analysis; if RCTs compare any of these competing drugs, they constitute *comparative science*. Whereas the former feature the more innovative products, complying directly with regulatory requirements, Azoulay suggests that the latter may well originate in the firms' marketing departments, since there is statistical evidence for their differential effects on sales[3]. However, even placebo-controlled trials are often suspect of commercial maneuvering: for instance, if we are dealing with subjective outcomes (often the case in psychiatric trials), Peter Gøtzsche (2013) has argued that a poor blinding either of the treatment (e.g., a placebo that does not properly mimic the effects of the active treatment) or the assessment may distort the comparison, making the treatment substantially more effective than it actually is. In summary, if we lack documental evidence for the intentions of the industry in sponsoring a trial, the discussion of whether it is conducted for commercial or research purposes should proceed on a case-by-case basis, as we will do below, and will not always yield an incontrovertible conclusion.

Be that as it may, there is a significant amount of direct and indirect evidence of the pharmaceutical industry manipulating trials for commercial purposes –again see Gøtzsche

---

[3] Azoulay (2002, pp. 583-584) measured the cumulative citations of both types of trials, analyzing how responsive the demand elasticity was to each of them. He found that comparative trials proved to be "a particularly effective business-stealing weapon" for the second drug to enter the antiulcer market he studied. At the same time, he also shows that pharmaceutical investment in medical detailing grew with the increasing stock of citations in market-expanding trials (2002, pp. 579-580). Even if the data show as well that trials were not the main drive behind the marketing strategies and sales of pharmaceutical companies, Azoulay concludes that RCTs "represent investments whose effects on the product market are both substantial and long-lived" (2002, p. 582)

2013 for an updated review. For the sake of the argument, we may at least consider the possibility that the industry may commercially exploit the definition of a disease, when this latter is implicitly given by a successful RCT

When this happens, we will bump into a couple of classical problems in statistical methodology. Behind the two types of disease-mongering we have distinguished we will find two such problems: the reference class problem and the dichotomy between substantive and statistical significance. As to the former, we might quote Alan Hajek (2007, p. 563): "The reference class problem arises when we want to assign a probability to a single proposition, X, which may be classified in various ways, yet its probability can change depending on how it is classified." In our terms: what is the probability that a patient might benefit from a given treatment? In principle, this is provided by a clinical trial regarding a given population: for the members of the latter we know how likely it is that the treatment will have an effect. The reference class (in its statistical sense) is thus the population on which we test the treatment: both the intervention and the control group come from the same reference class. *Mild disease mongering* occurs when the set of eligibility criteria is vague enough as to allow the inclusion of patients that, under a more strict definition of the disease, would not qualify as members of the trial population. *Strong disease-mongering* exploits the confusion between clinical and statistical significance: if we judge how good a treatment is only on the basis of the statistical difference between the outcomes observed in a trial, we might declare a treatment effective without a real assessment of how important that difference is from a clinical standpoint. E.g., we might shorten in, say, 3 days the recovery of a patient with a given treatment and that might be statistically significant, but how important are those three days from a clinical perspective?

It might depend, for instance, on the total number of days each treatment requires (a reduction of 3 days looks different when the total is 9 than when it is 90), etc.

In order to substantiate our claims, we present our two types of disease-mongering through two case studies. The best-selling tranquilizer Valium will illustrate how sales can be boosted by over-prescription beyond the populations originally targeted in a trial. The development of the first generation of statins will show how the trial population was expanded in search of statistically significant results that allowed sales to increasing audiences. By way of conclusion, we will discuss some possible solutions to both mild and strong disease-mongering.

2. Mild disease-mongering: the case of Librium and Valium

Before we present these pioneering benzodiazepines, let us briefly recall the basics of drug testing at the point they reached the market[4]. Between 1900 and 1950, expert clinical judgment was the main approach in the assessment of the properties of pharmaceutical compounds in Britain and the US. An experienced clinician would administer the drug to a series of patients he considered more apt to benefit from it. His conclusions would be presented as a case report, with the details of each patient's reaction to the treatment. The alternatives were first laboratory experiments and then controlled clinical trials (from which RCTs would later emerge). The former would proceed either *in vitro* or *in vivo* (on animals and patients): considered superior by clinicians with a scientific background, its scope was usually restricted to safety considerations. This soon gave way to comparative trials, in which two treatments were alternated on the same patient or administered in two groups of

---

[4] For an extensive presentation and discussion of these claims about RCTs and the Streptomycin trial, see [ANONYMIZED]

patients (simultaneously or not). Controls were used to ensure the comparability of the two treatments and these adopted different forms. The following items counted as controls in these trials: the patients' eligibility criteria, the way treatments were allocated (alternation and randomization), uniformity in administration of treatments and patients' blinding. They were not necessarily all used at once. Statistical reports from controlled trials conveyed their results with different degrees of sophistication. Statistical significance testing only features occasionally before 1950.

In 1947, the British Streptomycin trial set a standard experimental template for drug testing, hinged on randomization and significance testing as presented in the work of R. A. Fisher. In the coming two decades, additional controls were added to correct different sorts of biases and further statistical methods were implemented (e.g., confidence intervals) in order to substantiate the conclusions further. In the 1960s, with the amendment of the Food and Drug Administration Act, RCTs became the regulatory yardstick for efficacy and safety of a compound in the United States and soon after in many other countries as well.

It is important to note that although experts could often assess the biological activity of a compound in the laboratory (e.g., the antibiotic properties of penicillin and the first generation of wonder drugs in the 1950s), there was great uncertainty as to how these compounds would operate in the human organism. Streptomycin owes its name to the actinobacterium *Streptomyces griseus*. It was the first of a class of drugs, the *aminoglycosides*, discovered in 1943. Its bactericidal action is a protein synthesis inhibitor. It binds to the 30S subunit of the bacterial ribosome causing abnormal protein synthesis, and, ultimately, death of microbial cells through mechanisms that are still not understood today. Ribosomes were discovered in 1958, so the use of Streptomycin against tuberculosis

was grounded on laboratory trial and error for at least a decade. However, even if the full range of biological mechanisms triggered by a drug was often unknown, RCTs allowed physicians to hedge their bets. Once a therapy and the population of patients that might receive it had been defined, RCTs could establish its effects on a purely statistical basis.

From the standpoint of the pharmaceutical industry, this uncertainty about the clinical effects a compound might have gave rise to a sort of *molecular lottery*[5], in which biomedical researchers synthesized molecules with potential biological activity, tested them on animals in research of any kind and, if anything of interest showed up in the laboratory, further research was undertaken until it eventually went to a clinical trial. This is how the first benzodiazepines, minor tranquilizers used for the treatment of anxiety, were synthesized.

Leo Sterbach, a chemist at the American branch of the Swiss company Hoffman-La Roche (just *Roche* from now on), was commissioned to find a substance that could compete in the emerging market of tranquilizers[6]. He was familiar with a family of molecules that he had investigated in the search for industrial dyes in a previous stage of his career. Suspecting that they may have biological activity, in 1957 he synthesized 24 versions, one of which happened to cause muscle relaxation and sedation in animals. It was later commercialized as *Librium*. The following year he synthesized another version that came to be sold as *Valium*. The first trial of Librium began in 1958, accessing the market after FDA approval in 1960, followed by Valium in 1963. These two benzodiazepines became the best-selling drugs of the 1960s and 1970s, making Roche stock the most expensive in the world by

---

[5] We take the expression *molecular lottery* from Lynn 1993.
[6] Our account of this case study is based on two splendid volumes: Herzberg 2009 and Tone 2009; plus a recent synoptic survey, Horwitz 2013. We have also used the more promotional Baenninger 2004.

1972. Yet, only in 1977 the biological mechanism by which they achieved their tranquilizing effects began to be understood. In other words, the world's best-selling drug of the decade had succeeded in the market despite our causal uncertainty about its operating mechanisms (Tone 2009, 135). This was not at all uncommon, as historian D. Herzberg (2009, 195) puts it, "[t]hroughout most of this era, the therapeutic use of psychotropes did not depend on any particular theory about why they worked". RCTs, however, provided the statistical proof that they actually worked.

Roche's strategy to test these first benzodiazepines was to see how far they worked on anxiety, trying them in a variety of target populations and settings (prisons, hospital, outpatient, colleges, offices), showing uniformly tranquilizing effects across all of them (Tone 2009, 133-134). However, not all these trials were equally rigorous even by the standards of the time. Some Valium tests showed that it was effective for anxiety, but not for depression; others showed efficacy for anxiety and depression; still others showed it working on mild depression and agitation; others did not find effects on agitation, but did on tension. When strict placebos were introduced though, the effects were less noticeable[7]. As *The Medical Letter* ("A non-profit publication on drugs and therapeutics") noted already in 1963 (about Librium) and 1964 (about Valium), there wasn't enough evidence to prove the superiority of benzodiazepines over barbiturates in the treatment of anxiety[8].

Had these trials been conducted mainly for research purposes, we might have expected Roche to organize the commercialization of its pioneering benzodiazepines accordingly.

---

[7] See, e.g., Gundlach et al. 1966 for a review.
[8] "Librium", *The Medical Letter* 5.24 (1963), pp. 93-96; "Valium", *The Medical Letter* 6.5 (1964), pp. 17-20 Even in a new review in 1969, the evidence was not considered conclusive: "Librium and Valium", *The Medical Letter* 11.20 (1969), pp. 81-84.

That is, with great caution in targeting the patients who may benefit from them. Yet, in the words of a member of the marketing team involved in the promotion of Valium (a more potent and palatable version of Librium), there was "almost nothing you couldn't use it for". Valium was indeed advertised, as one famous slogan put it, as "psychic support for the always weary". And it was strongly advertised, for that matter: Valium was launched with a whole range of new marketing tools: targeted symposia presenting informal tests to specialists (as soon as 1959); films and cover stories in popular magazines ("The drug that tamed lions"); direct mailing to physicians and paid advertisements in journals. A crucial point is that Valium was a prescription drug: physicians had to diagnose the patient before s/he could benefit from the treatment. This is why we think that this is a case of *mild* disease-mongering: prescribing physicians followed the marketing lead and assumed that their patients could benefit from Valium as much as the patients originally treated in the trials. As a matter of fact, "just one-third of prescriptions were associated with a diagnosis of a specific mental illness", and general practitioners –not psychiatrists– were issuing most of them (Horwitz 2013, 123).

Putting it in statistical terms, physicians had to decide whether a given patient was a member of the intended trial population, so that the effects would be the same as those observed in the RCT. This is, of course, a fallible decision in which everything depends on expert clinical judgment. It is another version of the reference class problem: depending on how the physician diagnoses the patient, the probability of him/her benefiting from the treatment will vary. We may constrain these diagnoses with as much objective evidence as possible, but it is ultimately a matter of judgment: mistakes are always possible.

The problem in the objective evidence about anxiety is that *it was difficult to define anxiety independently of the trials*, i.e., as a collection of symptoms that Valium could alleviate. And the trial definition was broad enough to be applicable to every patient with a vague set of symptoms. Following again Herzberg[9], anxiety was defined at different levels. In the 1952 version of the *Diagnostic and Statistical Manual* (DSM) of the American Psychiatric Association, a whole array of neurotic symptoms were explained as a person's dysfunctional efforts in coping with anxiety (Herzberg 2009, 32). The etiology of such dysfunctions was interpreted in psychoanalytic terms, anxiety here being rooted in the personal story of the patient, treated by verbal therapy. In the following decade, with the 1968 version of the DSM, psychologists, sociologists and journalists produce an even broader definition of anxiety, as "the chief characteristic of the neuroses". According to Horwitz (2013, 128-129), *neurosis* was just a catch-all term for many non-psychotic disturbances, for which the DSM-II provided just short and cursory definitions. In other words, the boundaries between normal and pathological anxiety became even blurrier.

Librium and Valium contributed in several ways to make these definitions evolve. For a start, they merged the treatment of two previously distinct types of anxiety: in the 1952 DSM, anxiety was considered typical of neuroses, namely every non-psychotic syndrome (Herzberg 2009, 32); yet benzodiazepines turned out to be effective with anxiety in psychotic patients (Tone 2009, 131). Moreover, their vast tranquilizing effect suggested a biological foundation for their action, relatively independent of the diagnostic categories. Donald Klein, a pioneer of this pharmacological approach to anxiety in the late 1950s, made this case explicitly (Horwitz 2013,132-133):

---

[9] See also Horwitz 2013, 127-136

Perhaps the most important thing learned from this experience was the power of the experimental technique of pharmacological dissection whereby one can pierce through the fascinating, confusing web of symptoms and dysfunctions to tease out the major participant variables by attending to specific drug effects.

Again, as a research strategy, it made perfect sense (and it was actually incorporated into the DSM III: see Horwitz 2013, 134-136). However, rather than restricting the definition of anxiety according to the trial results, the Roche marketing teams interpreted the trials in terms of broader biological interpretations of anxiety and its etiology (originating, for instance, in the difficulties for our paleo-brain to cope with the stress of modern life). As a matter of fact, Roche promoted such broad definitions in its own textbooks on anxiety for physician's use (Branch 1965). In other words, if a patient presented a collection of symptoms identified in trials, a physician could legitimately consider him/her sick and prescribe the drug to control them. It actually succeeded, to the great satisfaction of the medical community.

This is a crucial point for our analysis: we are not claiming that Roche's disease-mongering was about single-handedly redefining anxiety on the basis of a marketing campaign. Anxiety was badly defined before the arrival of tranquilizers, and even if their trials could have contributed to make its definition more precise, they failed to do so, because the sponsoring company used them to promote a broader use of benzodiazepines. But the medical profession did not help to refine it either –as we have seen with the DSM-II definition. Physicians and patients alike contributed to this commercial success without

complaint for years, for reasons well beyond pharmaceutical marketing that cannot be explored at length here[10].

We are focusing on a more restricted sense of disease-mongering. For our purposes, the sort of disease that Valium could cure was defined by the apparent success of a number of trials. The eligibility criteria defined a class of patients who could potentially benefit from the treatment and the prescribing physician had to decide whether his own patients were similar enough to this reference class. Yet, thanks to pharmaceutical advertising, this population became a moving target (an *interactive kind*, in Hacking's terms). The anxious patient is defined as the target of a successful trial. Once this success is properly advertised, physicians identify anxious patients or they even learn how to behave like one. In either case, Valium relieves their anxiety to everybody's satisfaction. There is no way of questioning the diagnosis, since we are observing in the patients the same effects that we saw in the trial. And we do not have a better definition of anxiety to assess the patients independently of the one implemented in the trial protocol.

A trial does give you the probability of observing certain effects if you treat the member of a particular group with a given drug. But it does not give you the probability of anyone in particular being a member of this group. Hence, prescribing physicians must justify their clinical judgment themselves. And, pharmaceutical advertisers will deliver arguments to help them do it. This is what we call mild disease-mongering.

---

[10] See Herzberg 2009, Horwitz 2013 and Tone 2009: all of them concur in their analyses of the major role played by pharmaceutical advertising in the widespread use of benzodiazepines during the 1960s and 1970s, although they have different analyses of what made this period and era of anxiety. Baenninger 2004, a book on the story of Valium whose copyright is owned not by its authors, but by Hoffmann-La Roche, attributes the massive use of Valium during this period to the "indiscriminate prescription by general practitioners" (103), without mentioning pharmaceutical marketing at any point.

Our case study did not end up well: The problem with benzodiazepines, and with Valium in particular, was that they created addiction and there was a risk of overdose. Given the popularity of the drug (and the feminist outrage with many women's use of it), the late 1970s were full of public outrage, congressional hearings, etc. (Tone 2009, 203-232) Sales declined internationally, but the only sanctions against Hoffman-La Roche, the producer, were for abusive pricing (Lynn 1993). After all, with their massive prescriptions the medical profession had sanctioned the treatment and they had been entitled to do so. Indeed, there is no cure for the reference class problem, other than clinical judgment –even if it proved to be too vulnerable to pharmaceutical marketing.

3. Strong disease-mongering: the case of statins

We differentiated at the beginning between mild and strong disease-mongering, depending on whether marketing began before or after the trial. If the trial population was defined on purely scientific terms and the marketing campaigns tried later to make it grow, we said that disease-mongering was mild. But if there is a marketing goal in the definition of the trial population itself, we will speak of strong disease-mongering. Yet, we should recall again that the diseases treated in these latter trials are not inventions.

Risk factors are usually physiological variables that reliably predict the onset of a disease[11]. Coronary risk factors such as high blood pressure, high blood cholesterol or diabetes have been considered good predictors of a stroke, ever since they were first identified as such in the Framingham cohort study, where epidemiologists followed the population of this small American town in order to identify variables correlated with the probability of a coronary

---

[11] Our case study relies on Jeremy Greene's wonderful monograph *Prescribing by numbers* (2007, 149-219)

disease. The intuition behind the study was that if you could control the risk factors, you would lower the probability of having a coronary ailment.

As Jeremy Greene (2007) shows, treating a risk factor as if it were a chronic condition implied a radical rethinking of disease for both patients and physicians alike. "Patients" with a high risk factor were asymptomatic regarding the condition prevented by the treatment and therefore had no subjective experience of it. Many physicians initially resisted treating them. On the other hand, it is necessary to have a successful treatment in order to prove that controlling the risk factors lowered the probability of coronary conditions. The proof arrived in 1987, after more than two decades of controversial interventions, with the first generation of statins. Meanwhile, accusations of disease-mongering were made against those who defended the medicalization of risk factors.

Unlike Valium, there was a clear understanding of the mechanism by which high cholesterol triggered coronary conditions (atherosclerosis, the formation of fatty materials in the artery walls), but it was not precise enough. One controversial point in this regard was the definition of normal levels of cholesterol. That is, the identification of a population of patients who might be targeted with a treatment to lower them.

Drawing on hospital and insurance data, in the 1960s the first bell curves of LDL cholesterol were estimated (with a mean of 195 mg/dL and 2.5 % tails in 100 mg/dL and 300mg/dL). The first treatment was Mer/29 launched by Richardson-Merrel in 1960 and withdrawn a year later for serious adverse effects. For more than two decades, many other treatments followed and failed in trials, with an ongoing controversy, particularly strong in the 1980s, about "normal" cholesterol levels. This was a vicious circle: in order to approve

a drug as a cure, the FDA requires an indication reflecting an identifiably pathological state, but there was no consensual target; on the other hand, without a treatment succeeding in a trial, it was difficult to argue that hypercholesterolemia was a proper pathology.

The trial of the statin Mevacor in 1985 ended the controversy. It successfully targeted treating genetic severe hypercholesterolemia, unanimously acknowledged as pathological. In subsequent trials, Merck, the manufacturing company, aimed at patients with gradually lower levels of LDL cholesterol: in 1986 a trial lowered it in patients with a level higher than 240mg/dL; in 1988, Mevacor was used with patients with a level lower than 200mg/dL. In 1994, Merck launched Zocor (simvastatin), a preventive treatment addressed at patients who had LDL cholesterol below 130mg/dL. In 1995 Bristol-Myers' Pravachol (pravastatin) succeeded in a trial aimed at preventing coronary conditions in healthy men. According to the 1997 NCEP guidelines, these were those men with less than 100mg/dL; 95 mg/dL less than the first estimates from the previous decades.

Here we see strong disease-mongering in action. The condition is defined and re-defined through a series of trials, in which the population targeted grew bigger and bigger, at the expense of a consensual criterion of clinical significance. As Jeremy Greene (2007, 210) puts it, "the 2001 guidelines nearly tripled the proportion of the U.S. population that was eligible for lipid-lowering therapy to a market of 36 million people". Unlike Valium, with statins we do not see market expansion through derivative prescription to patients who are considered similar enough to those in the trial. Here, the eligibility criteria make the population grow with each trial. Prescription is even more legitimate here, because unlike anxiety, high cholesterol patients are identifiable as members of the target population through a physiological test.

The crux of the matter is, of course, whether this moving target, the cholesterol levels, are fixed according to scientific or commercial criteria. As we mentioned in the introduction, the statin trials may be interpreted as guided by research concerns alone: a causal search guided by well-specified interventions on the effectiveness of the treatment, for instance[12]. How can we tell whether research or commercial goals were primed in the conduct of the statin trials? We are going to ground our answer on their clinical significance, but this criterion is far from uncontroversial.

As Greene (2007, 210) puts it, regarding the conflicts of interests involved in fixing these criteria, "when confronted with the observation that most of the guideline committee members had financial ties to the companies that produced statin drugs, the committee chair, Scott Grundy, responded that it was impossible by the end of the twentieth century to find any medical expert who did not have strong industry ties."

Industrial biomedical research is driven by profit, painfully illustrated in the case of neglected diseases. One way or another, the prospect of future sales has always been present in any pharmaceutical research plan. However, we usually expect to identify the targeted disease independently of the commercial agenda of the industry. Since industry-

---

[12] As a reviewer, for instance, puts it: "given the roughly straight-line relationship between cholesterol and Coronary Heart Disease (CHD) risk, it isn't wholly unreasonable to aim at lower and lower cholesterol level groups. In fact Geoffrey Rose (*The Strategy of Preventive Medicine*, 1980) argues that, for public health purposes, the 'most at risk' - e.g. the top 10% - are not the group you should be intervening on. This is a potential source of objection that needs to be dealt with. Scientifically, there is no such thing as *high cholesterol*, so perhaps it makes sense to define in terms of interventions". However, the trial interventions under discussion are for CHD prevention in healthy patients, for which we do not have clear data: we do not have the death rate among those who did not take statins and the information on adverse effects is poor. Actually, there are a number of worrisome side-effects (Gøtzsche 2013, 47-48, 128-129). Therefore, we need to assess the clinical significance of these prophylactic interventions, as we will discuss below.

sponsored RCTs gradually became more and more embedded in this agenda, it is complicated to separate their marketing and scientific goals. And if the disease we are considering cannot be properly identified independently of such RCTs, the shadow of disease-mongering looms large.

We should now recall the methodological point we mentioned at the beginning. RCTs are tests to assess the effect of treatments on purely statistical grounds. We are calculating probabilities for a given population, but there is no purely statistical way to identify this population. Our cholesterol level will be more or less healthy depending on how we define hypercholesterolemia and, as we have seen, this definition depends on how successful we are at treating it in a trial. But the target population should be defined by clinical criteria: the fact that we have a drug that can lower your cholesterol levels does not imply by default that you ought to lower them. The trials are giving us probabilities for succeeding in prevention of coronary diseases for a given population, but it is a matter of medical judgment to decide how significant this achievement is. Statistical and clinical significance should be properly differentiated.

This is a controversial methodological point that we may illustrate with an example of S. Ziliak and D. McCloskey (2008)[13]. Assume that we have to choose between two diet cures, based on pill A and pill B. Pill A has an average effect of making you lose 10 pounds, with an average variation of 5 pounds. Pill B will make you lose 3 pounds on average, with an average variation of 1 pound. Which one leads to more significant loss? Naturally, we opt for pill A, in spite of the higher variation, because the effect size is so much larger.

---

[13] We take the discussion of this example from our own [ANONYMIZED]

But suppose we compare these two pills in a trial, with the null hypothesis being the default claim that there is no difference between them. Using the test statistic *z*, we may compare the relative weight of the evidence for A and B: observing a three pounds weight loss after taking pill B, with a known standard error of one pound, provides stronger evidence (for the efficacy of B) than observing a ten pounds weight loss after taking pill A, with a known standard error of five pounds:

$$z_A(10) = \frac{10 - 0}{5} = 2 \qquad\qquad z_B(3) = \frac{3 - 0}{1} = 3$$

With *z* we measure statistical significance of the trial outcome, understood as "the strength of the signal relative to background noise" (Hoover and Siegler 2008b, 58). On this score, pill B indeed performs better than pill A, and reasonably so because there is quite some noise in the effects of pill A. But should we assess the difference between A and B in terms of their statistical significance (B wins) or should we consider instead the clinical significance of achieving such a large weight loss (A wins), even if this latter is a noisy outcome? This is McCloskey and Ziliak's question.

Ben Goldacre has forcefully presented this point regarding statins –see also Gøtzsche 2013, 128-130: trials have been designed focusing on a surrogate endpoint (cholesterol reduction) where statistical significance in the comparison is achievable; but we lack data on a hard trial endpoint (death), where the clinical significance would be more obvious:

> Trials have been done comparing a statin against placebo, and these have shown that they save lives rather well. Trials have also compared one statin with another: but these all use cholesterol as a surrogate outcome. Nobody has ever compared the

21

statins against each other to measure which is best at preventing death. This is a truly staggering oversight, when you consider that tens of millions of people around the world have taken these drugs, and for many, many years. If just one of them is only 2 per cent better at preventing heart attacks than the others, we are permitting a vast number of avoidable deaths, every day of the week. (Goldacre 2012, 148)

Our case about strong disease-mongering will hold if the reader agrees on this criterion of clinical significance for statins –and disagreement is always possible. Summing up, we will speak of strong disease-mongering if trials are designed to target populations in order to achieve statistical significance in the comparison, with little regard for the clinical significance of the outcome. This is a problem that plagues the use of significance testing across disciplines, but may have dramatic consequences in medicine. However, there is no such clear-cut criterion for clinical significance as a p-value for statistical analysis. This is a second methodological gap that pharmaceutical marketing exploits and for which statistics again will not provide a solution. Let us discuss the alternatives briefly by way of conclusion.

4. What is the solution for disease-mongering?

We have defined disease-mongering as a strategy of pharmaceutical marketing that intervenes on the definition of a disease with a view to increasing the sales of its treatment. We have argued that this is possible when a disease is operationally defined through the success of a clinical trial: if a drug makes a statistically significant difference in the treatment of a group of patients, the eligibility criteria defining the population of patients provides a working definition of the treated disease. One way of making sales grow is thus

to expand the population of patients that might benefit from it. We have differentiated between mild and strong disease-mongering depending on how the population is expanded. The case of valium-treated anxiety illustrates the former: the trial targets a given population and the advertising campaigns try to persuade physicians that as many patients are possible are members of the same group and may benefit from the same treatment. In strong disease-mongering, illustrated by the fight against hypercholesterolemia, a series of trials expand the size of the population that might benefit from the treatment. In the two cases considered, the drugs had minor or major side-effects that ignited a controversy on the legitimacy of the prescribing guidelines inferred from the trials. But it may happen that disease-mongering is applied to a drug that might be, for many patients, just an active placebo: it neither cures you nor kills you[14]. Nonetheless it will cost someone money and this is a good enough reason to discuss disease-mongering: ideally, we would want trials to provide clinically relevant information for our decisions, and not just promote the commercial interests of a company.

There are two methodological problems that arise in the two forms of disease-mongering we have distinguished and both should be prevented. On the one hand, there is the reference class problem: in order to prescribe a drug a physician should decide how similar her patient is to the population targeted in a trial (i.e., what is the probability that he benefits from the tested treatment as they did); but the trial itself does not give an answer to this problem. On the other hand, clinical significance should be distinguished from statistical significance, and we should promote the former rather than the latter. This

---

[14] We are grateful to Maël Lemoine for this objection

amounts, in both cases, to a vindication of expert clinical judgment. This judgment should assess this clinical benefit of a treatment independently of the commercial interests of the manufacturing company.

There are in principle two ways of grounding such judgments, none of them new: diseases should be defined independently of clinical trials; and trials should be independent from the interests of the industry. As to the former, the safest option for a clinical definition of a disease is probably to go to its biological foundations. From this standpoint, we will consider that a treatment works if it somehow intervenes in the causal mechanisms triggering the pathology. However, we admit that in the case of psychiatric conditions this may be too difficult a goal to attain.

As we mentioned before, RCTs were established as a standard to decide whether a drug works at a point in which very little was known about such mechanisms and the industry often sought new compounds in the blind (in a *molecular lottery* ultimately decided in a clinical trial). Six decades later, the situation is dramatically different and treatments are now sought following a very precise understanding of the biology of the disease. The last two decades have seen the emergence of targeted therapies designed for particular genetic profiles, tracing the mechanisms that control the onset or development of a given condition. This approach should make us reconsider the very process of drug testing, since we are not playing a molecular lottery any more: we can explain why a targeted therapy works[15].

However, we are not there yet. Even if, from a philosophical standpoint, we should praise the contribution that mechanisms make to the assessment of medical treatments, there is

---

[15] For a discussion, see our [ANONYMIZED]

still enough uncertainty about the way they work: as Jeremy Howick *et al.* (2013) recently put it, current knowledge of mechanisms is often mistaken, the mechanistic knowledge itself can lack external validity, mechanisms can behave paradoxically, and the mechanist solution does not overcome the problem of the extrapolator's circle – *vid.* also Campaner 2012. Hence, we cannot dispense with clinical trials in testing medical treatments.

Leaving aside for a moment the threat of disease-mongering, these are clear signs that the pharmaceutical industry is using clinical trials for advertising purposes, focusing less on innovative research than on sale strategies. If drugs were freely available to consumers without prescription, the pharmaceutical companies would advertise to them directly. Prescription requires an investment in communicating to physicians the characteristics of new drugs, and clinical trials appear to be the best information source for this audience. Their results are primarily delivered to this audience in scientific journals, but also through conventional advertising, visits of pharmaceutical representatives and informative sessions in medical centers, etc. –see Jain 2007, for an overview.

Even if trials start out with a solid definition of the disease targeted, we need to make sure that the goal of the comparison between treatments somehow serves the interest of the patients. In the two cases of disease-mongering considered here, this implies focusing on clinically significant endpoints and target groups of patients that are as relevant as possible for the final prescription. As a matter of fact, most trials are designed with good definitions of the targeted disease, but often the outcome does not meet either of these two goals (e.g., Bero & Rennie 1996). The problem with trials is thus greater than that of disease-mongering, in the sense considered here. We just want to show that a general solution already in place would also apply to the problem discussed: to design and conduct second

and third phase clinical trials in publicly funded institutions, either for a fee or directly from the public purse. The industry would be free to organize their own trials, but for regulatory purposes only publicly financed trials would count for approval. If expert clinical judgment is both fallible and inevitable in the design and interpretation of trials, we should, therefore, arrange an institutional setting in which the incentives for these experts do not collide with the public interest.

To close, we hope we have contributed a more precise definition of disease-mongering, even if it does not exhaust the many forms of pharmaceutical marketing that are nowadays referred to as such. Mild and strong disease-mongering exploit the limitations of the statistical methodology underlying trials, but there is no internal statistical solution for either of them. In order to prevent disease-mongering, we need trials grounded on more solid science and designed with a clinically relevant goal. This is a platitude on which, we hope, everybody will agree although the precise details must still be worked out. Here, however, we have tried to define the problem and its magnitude as accurately as possible.

5. References

Abraham, J. (2009). Partial Progress: Governing the Pharmaceutical Industry and the NHS, 1948–2008, *Journal of Health, Politics, Law and Policy* 34: 931–77.

Abraham, J. (2010). Pharmaceuticalization of Society in Context: Theoretical, Empirical and Health Dimensions, *Sociology* 44(4): 603–622

Azoulay, P. (2002). Do Pharmaceutical Sales Respond to Scientific Evidence? *Journal of Economics and Management Strategy*, 11, 551-594.

Baenninger, A. (2004). *Good chemistry: the life and legacy of valium inventor Leo Sternbach*. New York: McGraw-Hill.

Bero, L. A., & Rennie, D. (1996). Influences on the Quality of Published Drug Studies. *International Journal of Technology Assessment in Health Care*, 12, 209-237.

Branch, C. H. H., & Roche Laboratories. (1965). *Aspects of anxiety*. With a pref. by C. H. Hardin Branch. Philadelphia,: Lippincott.

Broadbent, A. (2011). Causation and models of disease in epidemiology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 40.4, 302-311.

Campaner, R. (2012). *Philosophy of medicine. Causality, Evidence and Explanation*. Bologna: Archetipo Libri.

Goldacre, B. (2012). *Bad Pharma*. London: Fourth State.

Gøtzsche, P. (2013). *Deadly Medicines and Organised Crime: How big pharma has corrupted healthcare*. London: Radcliffe Health.

Greene, J. A. (2007). *Prescribing by numbers : drugs and the definition of disease*. Baltimore: Johns Hopkins University Press.

Gundlach, R., Engelhardt, D., Hankoff, L., Paley, H., Rudorfer, L., & Bird, E. (1966). A double-blind outpatient study of diazepam (Valium) and placebo. *Psychopharmacologia*, 9, 81-92.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, 156, 563-585.

Herzberg, D. L. (2009). *Happy pills in America : from Miltown to Prozac.* Baltimore: Johns Hopkins University Press.

Horwitz, A. (2013). *A Short History*. Baltimore: The Johns Hopkins University Press.

Howick, J., Glasziou, P., & Aronson, J. (2013). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics*, 34, 275-291.

Jain, S. (2007). *Understanding Physician-Pharmaceutical Industry Interactions. A concise guide*, N. York: Cambridge University Press.

Lexchin,J. (2006). Bigger and better: How Pfizer redefined erectile dysfunction, *PLoS Medicine*, 3(4), 429−432

Lynn, M. (1993). *The Billion Dollar Battle: Merck Versus Glaxo.* London: William Heinemann Ltd.

Moynihan, R., Gøtzsche, P. C., Heath, I., & Henry, D. (2002). Selling sickness: the pharmaceutical industry and disease mongering. Commentary: Medicalisation of risk factors. *British Medical Journal*, 324, 886-891.

Moynihan, R. (2003) The making of a disease: female sexual dysfunction. *British Medical Journal*, 326 (45), 658

Payer, L. (1992). *Disease-mongers : how doctors, drug companies, and insurers are making you feel sick.* New York: J. Wiley.

Sismondo, S. (2009). Ghosts in the Machine: Publication Planning in the Medical Sciences. *Social Studies of Science*, 39, 171-198.

Tone, A. (2009). *The age of anxiety: a history of America's turbulent affair with tranquilizers.* New York: Basic Books.

Williams, S.J., Gabe, J. and Davis, P. (2009). The Sociology of Pharmaceuticals. In S. Williams, J. Gabe and P. Davis (Eds), *Pharmaceuticals and Society*. Chichester: Wiley−Blackwell, pp. 1–11

Wolinsky,H. (2005). Disease mongering and drug marketing. *European Molecular Biology Organization Reports*, 6(7), 612−6

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: how the standard error costs us jobs, justice, and lives.* Ann Arbor: University of Michigan Press.