theFuture ofScience andEthics

Rivista scientifica a cura del Comitato Etico
della Fondazione Umberto Veronesi

Volume 6 ■ 2021 ■ ISSN 2421-3039

Fondazione
Umberto Veronesi
– per il progresso
delle scienze

theFuture ofScience andEthics

Fondazione
Umberto Veronesi
– per il progresso
delle scienze

theFuture ofScience andEthics

Rivista scientifica
del Comitato Etico
della Fondazione Umberto Veronesi
ISSN 2421-3039
ethics.journal@fondazioneveronesi.it
Via Solferino, 19
20121, Milano

Volume 6 ■ 2021

SOMMARIO

ARTICOLI

• **IL '*GREEN PASS*' ALLA LUCE DELL'ARTICOLO 32 DELLA COSTITUZIONE: ALCUNE BREVI CONSIDERAZIONI**
di Federico Gustavo Pizzetti **10**

• **ANTROPOCENE, PANDEMIA, GIUSTIZIA INTERGENERAZIONALE: L'ETICA PUBBLICA AL CROCEVIA FRA INCLUSIONE ED ESCLUSIONE DEL FUTURO**
di Ferdinando G. Menga **22**

• **LA VITA UMANA COME BENE DISPONIBILE**
di Giorgio Macellari **32**

• **GEOETICA: UN'ETICA PER LA RELAZIONE TRA GLI ESSERI UMANI E LA TERRA**
di Silvia Peppoloni e Giuseppe Di Capua **42**

• **WHY DO WE NEED RANDOMIZED CONTROLLED TRIALS? MEDICAL SCANDALS AND THE EVOLUTION OF DRUG REGULATION**
di Mattia Andreoletti **54**

• **MICROETHICS FOR HEALTHCARE DATA SCIENCE: ATTENTION TO CAPABILITIES IN SOCIOTECHNICAL SYSTEMS**
di Mark Graves e Emanuele Ratti **64**

• **LA BIOETICA COME PROFESSIONE E L'EXPERTISE IN MATERIA BIOETICA: RIFLESSIONI PEDAGOGICHE SULLO SVILUPPO DI UN CURRICULUM DI MASTER DI SECONDO LIVELLO IN BIOETICA E SCIENZE SOCIALI IN AMBITO ANGLOSASSONE**
di Silvia Camporesi **74**

DOCUMENTI DI ETICA E BIOETICA

• **LA FIGURA DELL'ESPERTO IN BIOETICA**
Comitato Nazionale per la Bioetica **86**
Commenti di
• Marianna Gensabella e Lucio Romano 94
• Demetrio Neri 98

• **IL TEMPO DELLA RICERCA. COMPRENDERE LA SCIENZA PER SUPERARE L'EMERGENZA COVID-19**
Comitato Etico Fondazione Umberto Veronesi **102**
Commenti di
• Raffaella Campaner e Marina Lalatta Costerbosa 112
• Federica Russo 116
• Daniele Fanelli 120
• Gianluca Attademo 124

• **SCIENCE FOR PEACE 2021: IL DIRITTO E IL DOVERE DI VACCINARSI** **128**

RECENSIONI

Consulta Scientifica del Cortile dei Gentili
PANDEMIA E GENERATIVITÀ. BAMBINI E ADOLESCENTI AI TEMPI DEL COVID
di Mons. Carlo Maria Polvani **134**

Anna Maria Bruzzone
CI CHIAMAVANO MATTI. VOCI DAL MANICOMIO (1968-1977)
di Anna Poma **138**

Maya J. Goldenberg
VACCINE HESITANCY: PUBLIC TRUST, EXPERTISE, AND THE WAR ON SCIENCE
di Teresa Gavaruzzi e Alessandra Tasso **142**

Antonella Ficorilli
NUOVI TERRITORI PER L'ETICA NELLA RICERCA SCIENTIFICA
di Matteo Galletti **146**

Agnese Collino
LA MALATTIA DA 10 CENTESIMI. STORIA DELLA POLIO E DI COME HA CAMBIATO LA NOSTRA SOCIETÀ
di Donatella Barus **150**

Armando Massarenti e Antonietta Mira
LA PANDEMIA DEI DATI. ECCO IL VACCINO
di Cinzia Caporale **152**

Laura Pepe
LA VOCE DELLE SIRENE. I GRECI E L'ARTE DELLA PERSUASIONE
di Mauro Serra **156**

Alessandro Bilotta e Dario Grillotti
LA FUNZIONE DEL MONDO. UNA STORIA DI VITO VOLTERRA
di Sandra Lucente **160**

Sara Garofalo
SBAGLIANDO NON SI IMPARA. PERCHÉ FACCIAMO SEMPRE LE SCELTE SBAGLIATE IN AMORE, SUL LAVORO E NELLA VITA QUOTIDIANA
di Andrea Grignolio Corsini **164**

NORME EDITORIALI **168**

CODICE ETICO **169**

I COMPITI DEL COMITATO ETICO DELLA FONDAZIONE VERONESI **172**

theFuture ofScience andEthics

Articoli

Why do we need randomized controlled trials? Medical scandals and the evolution of drug regulation

Perché abbiamo bisogno degli studi randomizzati e controllati? Scandali medici e l'evoluzione della regolamentazione farmaceutica

MATTIA ANDREOLETTI
mattia.andreoletti@gmail.com

AFFILIAZIONE
Università Vita-Salute San Raffaele

SOMMARIO

Perché abbiamo bisogno degli studi randomizzati e controllati (RCT)? Finora, le risposte a questa domanda si sono concentrate principalmente sulle loro virtù metdologiche. In breve, abbiamo bisogno degli RCT perché questi sono il modo migliore per valutare la sicurezza e l'efficacia dei farmaci. Ma questa risposta è solo parzialmente soddisfacente, non spiega infatti perché mai vogliamo testare i farmaci prima che possano essere commercializzati e perché vogliamo farlo in modo così rigoroso. In questo articolo analizzo le ragioni che hanno portato gli RCT a diventare il 'gold standard' della ricerca clinica. Tali ragioni sono il risultato dell'interazione di preoccupazioni storiche, organizzative e socio-politiche. Concentrandomi sulla storia della regolamentazione dei farmaci negli Stati Uniti, sostengo che i cambiamenti e le riforme sono stati attuati in risposta a grandi scandali farmaceutici, e non solo in risposta alle reali esigenze epistemiche emerse con gli sviluppi della ricerca farmaceutica. Nello specifico, mostro che gli scandali hanno giocato un ruolo cruciale nell'innescare una serie di riforme della regolamentazione dei farmaci, e quindi nel plasmare la metodologia della ricerca clinica contemporanea.

PAROLE CHIAVE

Medicina

Studi clinici

Regolamentazione dei farmaci

Farmacologia

Scandali

ABSTRACT

Why do we need Randomized Controlled Trials (RCTs)? So far, the answers to this question have mostly focused on the virtues of the methodological design. Roughly, we need RCTs because they are the best way to assess drugs safety and efficacy. But this answer is just partially satisfactory since it does not explain why, in the first place, we want to test drugs before they can be marketed, and why we want to do it in such a rigorous way. Here I clarify that the reasons that brought about the emergence of RCTs as the 'gold standard' of clinical research are the outcome of the interaction of historical, organizational, and, ultimately, socio-political concerns. Focusing on the history of drug regulation in the United States, I argue that changes and reforms were implemented in response to major pharmaceutical scandals, and not only in response to the real epistemic needs put in place by developments in drug research. More specifically, I show that scandals played a crucial role in triggering reforms in drug regulation, and hence in shaping the methodology of contemporary clinical research.

KEYWORDS

Medicine

Clinical trials

Drug regulation

Pharmacology

Scandals

DOI: 10.53267/20210105

1. INTRODUCTION

Nowadays the medical-scientific community agrees that a Randomized Control Trials (RCT) is the best research design to evaluate the efficacy of medical treatments in a specific population. Conventionally, the term 'treatment' refers to many kinds of interventions: diagnostic, screening, health education, etc. However, RCTs are systematically and extensively adopted in drug research and testing, as they are the last phase of a mandatory threefold process, which is strictly regulated by transnational laws. Of course, RCTs did not come out of the blue, nor did the rules that had made them compulsory. In this article, I dig into the history of randomized controlled trials to bring out and make clear the reasons why they became the gold standard for drug testing and regulation. More specifically, I argue that pharmaceutical scandals played a crucial role in the development of drug regulations, forcing regulators to acknowledge the weak points of previous standards and to consider more robust alternatives, ranging from laboratory tests to RCTs. The historical investigation of the evolution of methodological concepts is instrumental to warrant our claim (Schickore, 2011). Why do we need RCTs? So far, the answers to this question have mostly focused on the virtues of this methodological design. Roughly, we need RCTs because they are the best way to assess drugs safety and efficacy. But this answer is just partially satisfactory since it does not explain why, in the first place, we want to test drugs before they can be marketed, and why we want to do it in such a rigorous way. Here I clarify that the reasons that brought about the emergence of RCTs as the 'gold standard' of clinical research are the outcome of the interaction of historical, organizational, and, ultimately, socio-political concerns.

In order to bridge this explanatory gap, I suggest focusing on the history of drug regulation in the United States (Gaudillière & Hess, 2013; Marks, 2000; Temin, 1980, 1985). As I argue, changes and reforms were implemented in response to major pharmaceutical scandals, and not in response to the real epistemic needs put in place by developments in drug research. Moreover, if this is true, then one should evaluate the epistemic import of experimental designs also to the extent to which they could prevent scandals. Considering the historical and socio-political context is particularly relevant for the recent debates on the adoption of new regulatory standards. As the historian of medicine, Marcia Meldrum put it: «the RCT is a dynamic methodology, and its present and future are informed by its history» (Meldrum, 2000).

Historically, the link between scandals and policies in Western democracies is nothing new: many sociologists and political scientists have discussed it for decades (Butler & Drakeford, 2003; Thompson, 2000)[1]. In general, a scandal is defined as an event, often regarded as morally wrong, which causes public outrage. While it is clear that scandals play a crucial role in the general political scenario, it is quite uncharted whether these events could have an impact in other fields, such as clinical research. In what follows, I show that scandals played a crucial role in triggering reforms in drug regulation, and hence in shaping the methodology of contemporary clinical research.

2. THE GREAT AMERICAN FRAUD

In the last decades of the nineteenth century, laboratory science had a great boost thanks to the development of many basic research fields such as chemistry, physiology, and microbiology. These scientific advancements ended up in what historian of medicine Charles Rosenberg (Rosenberg, 1997) has called a 'therapeutic revolution', that is, the discovery of a noticeable number of effective therapeutic agents. Physicians and patients were deeply affected by this 'revolution', as they came across a continuously increasing number of new drugs.

However, physicians realized soon that many drugs did not contain any active ingredient, but pharmaceutical companies promoted inactive drugs in the same way as the ones with real and active compounds. With this regard, for instance, Samuel Hopkins Adams, an American investigative journalist (a muckraker), in 1905 coined the expression «The Great American Fraud ». In discussing therapeutic reforms, the market plays a contingent yet significant role, much as scientific progress does. Indeed, at a certain point, the medical scientific community had to face «a novel intellectual and political problem» (Marks, 2000): how to foster even further the increasing scientific progress in the laboratories while protecting the patients and the market from fake and potentially unsafe drugs. In other words, there was a need to tell apart effective and inef-

fective drugs without discrediting the entire scientific enterprise.

The American Medical Association (AMA) made the first effort towards a more rational approach to pharmaceutical therapeutics. In the spring of 1905, the AMA established the Council on Pharmacy and Chemistry, which had the task of investigating the medicines advertised in the pages of the Journal of the association (JAMA). The work of the Council was to review the scientific evidence supporting a drug and deliberate on its quality. In practice, the scientific evidence was often scarce and then the deliberation of the council reflected «a curious mixture of judgments […] and opinions» (Marks, 2000). When the council's assessment was a matter of laboratory tests, to reveal whether the drug contained an active known ingredient, the decision was quite easy. However, pharmaceutical companies also developed drugs containing ingredients that could be tested in a laboratory but whose beneficial properties were completely unknown. In these murky cases, the deliberation was more difficult or even impossible. In these latter cases, extra-scientific considerations, such as the track record of the companies, played a major role in the decision-making process.

As just mentioned, since clinical evidence was scarce or even missing, the Council relied mostly on the expertise of academic clinicians, but then often bumped into a divergence of opinions. Hence, they warned that their approval for the biological purity of the compounds did not imply clinical efficacy. In many cases, laboratory tests could not address the question of efficacy. Take for instance glandular extracts (e.g., red bone marrow, ovarian, parotid gland extracts) that were common on the market in the early 1900s. Labels reported the exact chemical composition, and this could be easily tested in the labs. However, it was unclear what all those extracts did: laboratory tests were not sufficient for that question.

Nonetheless, the U.S. Government in 1906 passed the first key legislation to control the drug market: the Pure Food and Drug Act. The new law gave to the Bureau of Chemistry (the predecessor to the FDA) in the Department of Agriculture the legal power to seize adulterated or misbranded products (Junod, 2008). But it assumed the same standards of the Council: laboratory tests to check whether a drug contained the ingredients labeled or advertised by the manufacturer. Moreover, the law did not allow anyone to screen drugs and control for potential frauds before their placing on the market: it was remedial but not preventive. The meaning and the exact enforcement of the 1906 Act were indeed questionable. In 1912, to counter this flaw, the U.S. Congress enacted the Sherley Amendment that prohibited explicitly false therapeutic claims. However, in the following years, the consequences of the new law were practically nil since it was still hard to prove something regarding the therapeutic effects of the drug through laboratory tests alone. Yet, the necessity to investigate in a more systematic way a method to test for drug efficacy was made clearer only a few years later, when some scandals emerged.

One of the most striking was the case of Banbar, an old patent medicine advertised as a cure for diabetes. The drug was not dangerous *per se*, since it contained just inactive ingredients like milk, sugar, and a grass plant known as 'equisetum'. Nonetheless, it was obviously life-threatening for those who rejected insulin, which had become a standard treatment shortly after its discovery in 1922. Meanwhile, in 1927, the Bureau of Chemistry's name was transformed into the 'Food, Drug, and Insecticide Administration', then abbreviated to the current version (FDA). In the 30s the 'new' FDA accused the producer of Banbar of fraud and took to court all the evidence about the death of patients who had refused to take insulin to get Banbar. In its defense, the producer of the drug took to the court testimonial letters, which consumers had written thanking him. Those letters were sufficient to demonstrate to the court his *bona fides* about the efficacy of the drug. So, the FDA did not get the authorization to seize the product that remained on the market (Junod, 2008).

This case clearly illustrates the major limits of the 1906 Act: it was more about basic chemical quality control (the drug had the ingredients it claimed it had) to protect consumers from frauds, rather than addressing more relevant epistemic needs such as safety and efficacy. These would come in the following decades when new scandals made it unavoidable.

3. THE 1938 FOOD, DRUG, AND COSMETIC ACT

In the wake of Banbar and other minor scandals, people started being more and more suspicious

of pharmaceutical companies and the drug trade. In those years, two books became very popular and influential among the public opinion: *100,000,000 Guinea Pigs: Dangers in Everyday Foods, Drugs, and Cosmetics* by Arthur Kallet and F.J Schilnk, and *American Chambers of Horrors: the truth about food and drugs* by Ruth deForest Lamb. The authors harshly criticized the FDA and the government for their failure in protecting people from the abuses and the frauds of drug companies. They pointed out all the weaknesses of the 1906 Act, asking for an immediate update. Instead, at the very beginning, the FDA reacted vindicating the success of all its activities.

Meanwhile, the pharmaceutical market was growing fast. In the 1930s, more than a hundred companies were manufacturing drugs containing sulphanilamide, a 'wonder' antibacterial compound used to cure streptococcal infections. The company S. E. Massengill started the production of a syrup-type sulphanilamide using diethylene glycol, an extremely toxic solvent. The syrup was then placed on the market, without any tests in animals or humans, causing at least 106 documented deaths (Wax, 1995). However, under the 1906 Act, the FDA could only prosecute Massengill for misbranding. The subsequent public outrage prompted Congress to pass a new set of laws: the 1938 Food, Drug and Cosmetic Act. The 1938 Act required companies to inform the FDA of their intention to put a new drug on the market. On the one hand, the FDA was given the power to ask for «adequate tests by all methods reasonably applicable to show whether or not the drug is safe» (Greenberg, 1999). The major concern of regulators in the 1938 Act was the safety of the drugs, whereas they did not nearly consider the problem of evaluating the efficacy, which of course they soon bumped into. On the other hand, the 1938 Act did not make FDA approval a prerequisite for market access.

It is worth focusing on the kind of 'adequate tests' required by the FDA as proof of drug safety. Although these tests remained unspecified in the act, the regulators adopted the same standards already advocated by the AMA's Council on Pharmacy and Chemistry: laboratory analysis and experts' evaluation. Moreover, animal tests, even if not formally required, were systematically requested by the FDA, and soon became a sort of gold standard for drug safety. This was one of the major novelties of the 1938 Act. Another major accomplishment was the overcoming of the 'fraud flaw' of the 1906 Act: the FDA could now remove from the market unsafe drugs without having to prove that there was the intent of fraud on the part of the producer.

Soon the new Act was put to the test. In the spring of 1938 British researchers had discovered a new sulfonamide compound (2-para-aminobenzene pyridine), apparently better than every other sulfa drug. Experimental tests in mice showed low toxicity, few adverse side effects, and more beneficial effects than its predecessors. In October 1938, Merck & Company, an American company, applied for the FDA approval of sulphapyridine for the treatment of pneumonia, for which there was no effective therapy yet. The FDA requested the opinion of the experts and clinicians who had the opportunity to test the experimental drug. Some of them were reporting adverse events, some did not. On the drug's efficacy, the data were even more unconvincing: the drug had been administered only to a few patients with pneumonia and it was still too early to judge its efficacy. Therefore, many sceptics were advising the FDA to keep the application on hold since they were concerned about the risk-benefit balance. They were also concerned about the lack of data on the effects of sulphapyridine on other infectious diseases for which it might be prescribed.

The FDA had adopted the view that the expert judgment of qualified clinical investigators should prevail over the opinion of regular clinicians. But in case of disagreement among the former, the debate would not be settled by the methodological superiority of their respective tests, but through the majority rule.

Despite the pressure of the press, asking for fast approval of the drug, and despite the incoming winter, a time when cases of pneumonia were more frequent, the FDA kept collecting and reviewing data and experts' opinions until the deadline provided for in the statute. In March 1939 the FDA decided to 'not deny' (it is worth noting that, officially, the act did not allow the FDA to *approve* a drug, but just gave to the agency the power to deny a request) the applications for sulphapyridine, provided that manufacturers explicitly reported on the labels and in advertising that the drug had to be used «under close, continuous observation of a qualified practitioner of medicine» (Marks,

2000). This is because some doubts remained about the efficacy of the drug. As noted by Theodore Klumpp, by then chief of the Drug Division in the FDA: «While a few investigators recommended that the drug be withheld from the market such recommendations upon analysis do not appear to rest upon considerations of the intrinsic safety or danger of the drug. Principally those workers were concerned with the orderly development of medical scientific knowledge, concerning the therapeutic efficacy of the drug» (Marks, 2000). The sulphapyridine was soon replaced by a more powerful drug, penicillin, so the extent of the FDA's decision is not clear. But regarding safety, 'adequate tests', laboratory analysis and experts' judgment gave the impression to perform that task well. At least, it seemed so.

What was clear among the medical community, at that point, was that the standards adopted by the FDA were far from being able to check for efficacy. Drug evaluation was left to the judgments and opinions of experts, which was considered superior to regular clinical judgment, and the medical community thought to be reliable at least in spotting adverse effects. However, another scandal soon undermined that belief and forced the FDA to reconsider its regulations and standards. Developments outside the medical field converged to make it possible.

4. METHODOLOGICAL INNOVATIONS

Physicians had been dealing with the variability of biological phenomena for centuries. They were always aware of the fundamental role of chance in medical observations: the natural course of the disease, spontaneous remissions, and response to treatments were considerably different in each patient. Clinical measurement was not as uniform as laboratory tests. Therefore, physicians relied only on their experience to handle uncertainty. This was the case also in comparative experiments. Indeed, knowledge of the variance of the diseases, and potential perturbing factors, could be exploited to perform comparative studies, trying to reduce the chance to a minimum. Of course, the management of chance was considered fundamental for any comparative experiment. Therefore, their quality depended on the experience of the researcher. Still, this approach had serious limitations because physicians' knowledge of both confounding factors and the magnitude of natural random variability might be limited. What statistics could offer to clinical researchers was an experimental design that permitted to control for biological variability and chance regardless of previous knowledge. Generally, this breakthrough is credited to the genius of a British statistician and biologist, Sir Roland Aylmer Fisher (1890-1962).

Fisher had been dealing with biological variability since 1919 when he began to work as a statistician at the agricultural experimental station in Rothamsted. Fisher had to find a reliable method to solve some practical problems in agricultural research: Which varieties of seeds are better? Which fertilizer? Which crop rotation system is best? Simple comparisons cannot provide a reliable answer. Suppose that you observe a 10 percent difference in yields between two varieties: is it due to a real difference in the quality of the seeds or to plot conditions? One way to answer this is to rely on experience: an expert farmer could tell that a 10 percent difference is never due to plot conditions alone. Nonetheless for Fisher, this strategy was far from being scientific since it relied entirely on experts' knowledge (i.e., subjective). Moreover, it would not be feasible if such previous knowledge were not available to anyone. Another option would be to replicate the experience many times, but this is rarely possible in agricultural practice. Fisher calculated that it would require approximately five hundred years to find that such a 10 percent difference is due to chance alone.

So, Fisher proposed to set up a new experimental design, dividing the experimental plots into strips to increase the number of observations in a single experiment. This way he reduced the variability of the effects due to other factors than a quality difference between grains. In other words, he increased the sample size of the experiment. But the most crucial innovation was to sow grain in strips in *random* order. The randomization of the plots ensured that all the possible perturbing factors were equally distributed among all the strips.

According to Fisher randomization is crucial not just for controlling for confounders, but also for the calculation of the probability of finding a given difference between the experimental treatments, as illustrated by the famous thought experiment of the lady tasting tea (Salsburg, 2002), which however cannot be discussed here.

Yet, Fisher's direct influence on biological and medical communities was negligible. It was Bradford Hill, a British statistician working on medical topics, who exported Fisher's experimental design to drug testing in the 1940s. Historians of medical statistics have argued, time and again, that British physicians did not grasp the statistical rationale of randomization (Armitage, 1982; Chalmers, 2011). There was instead a widespread concern among British doctors about the many ways in which personal biases could spoil the evaluation of novel therapies.

They found in the randomized allocation of treatment a device that could neutralize the personal beliefs of investigators as to who would benefit most from the therapy. Allocation bias occurs when the allocation of subjects to study groups is jeopardized by the preferences of the experimenters (e.g., the healthiest or youngest patients receive the experimental treatment). Randomization can easily succeed in neutralizing this bias. However, many other biases can occur in a comparative experiment. For instance, participants' preferences can still spoil the result, conditioning the evaluation of the outcomes. If physicians want to favor the drug under testing, they could report better outcomes for the experimental drug and so could patients as well. That is why we need another de-biasing method, such as blinding the allocation of treatments to physicians and patients. Indeed, comparative controls, such as blind assessment, are similarly instrumental for the coming into being of RCTs (see Kaptchuk, 1998; Shapiro & Shapiro, 2000). And randomization is an essential part of blinding procedures.

Despite their merits, it took more than a decade to implement both Fisher's approach and controls in medicine: the first randomized controlled trial, with significance testing and blind assessment, took place in Britain in 1947. It would take one more decade to spread among physicians and two more decades to transform it into a regulatory standard (Byar et al., 1976).

5. HOW RANDOMIZED CONTROLLED TRIALS BECAME THE GOLDEN STANDARD

In the years after the war, some breakthroughs in clinical trials design were achieved in two independent studies of streptomycin. For the first time, researchers introduced in trials' design a standardized set of controls that will soon become fundamental: a control group, the random allocation of patients, and standardized non-qualitative criteria to assess outcome. In the U.S., the Public Health Serviced (PHS) organized a research study on streptomycin to treat tuberculosis. PHS researchers did not want to make the mistakes of their predecessors, so they strictly controlled the trial funding and the limited amount of streptomycin available made it necessary to arrange comparative experiments to produce the best knowledge most efficiently.

To control for the allocation bias, the study design included the randomization of treatments. PHS researchers' main concern was to avoid individual decisions of physicians, especially those who were already convinced of the beneficial effects of streptomycin. That is also why PHS researchers planned to conduct the entire study in a double-blind fashion, but they failed to convince the involved physicians. Nonetheless, the study produced reliable and uncontested results in favor of streptomycin. However, it employed only descriptive statistics, there was no use of statistical tests of significance.

On the other side of the Atlantic, in 1947, the British Medical Research Council was conducting a very similar trial, which became known as the 'first RCT' ever, since it employed for the first time a standardized method for statistical inference. The scientist in charge was Sir Austin Bradford Hill, a relevant actor in the history of medicine. In a series of papers on medical statistics, published in the prestigious journal The Lancet, he defended the relevance of randomization and controls to ensure the objectivity of a study. Bradford Hill argued that the primary experimenter's aim is «to ensure beforehand that, as far as possible, the control and treated groups are the same in all relevant respects» (Yoshioka, 1998). Moreover, randomization was crucial to ensure the objective assessment of treatments since it removed personal responsibility from the clinician in selecting which patients would benefit. These two ideas shaped the rationale behind the design of the MRC trial. The trial enrolled 107 patients randomized in two groups: 55 assigned to the experimental group receiving streptomycin and the standard of care (bed rest) and 52 to the control group receiving only bed rest. The radiologists who interpreted x-ray chest exams were blind to the allocation of the treatments. After 6 months there were only 4 deaths

in the streptomycin group, whereas there were 15 in the control. Investigators considered that difference statistically significant, «the probability of it occurring by chance is less than one in a hundred» ("Streptomycin Treatment of Pulmonary Tuberculosis", 1948). For the first time both a method for minimizing allocation bias and statistical evaluation of collected data were employed in a clinical trial. Therefore, Hill's trial became a milestone and influenced an entire generation of physicians.

Certainly, these trials had both a great and important weight in the history of medicine and clinical research, i.e., the exclusion of subjective judgments from drug testing and evaluation. The new methodological standard provided indeed a more objective and scientific tool to appraise therapeutic innovations, rather than relying on conflicting judgments. Yet, despite its initial success, the randomized controlled experimental design was integrated into drug regulation more than a decade later, in the aftermath of further pharmaceutical scandals.

In 1959, U.S. Senator Estes Kefauver held hearings on the drug industry. His main concern was the exorbitant profit margins of pharmaceutical companies. The companies justified their profits with the high costs of research since many drugs failed during drug development. The hearings generated important evidence documenting the poor quality of clinical research supporting the marketing of many drugs. It revealed to the public what all the experts already knew: most of the clinical research was just rubbish. Kefauver's hearing placed drug regulation on the top of the agenda of U.S politics, but it was another tragedy to trigger both the enactment of the Kefauver-Harris 1962 Amendments of the 1938 Act and the subsequent Investigational New Drug Regulations in 1963.

The story of thalidomide is notorious. It was a quite popular drug in Europe and especially in Western Germany, where it was manufactured by pharmaceutical company Chemie Grünenthal since 1957 and marketed as Contergan. The drug was prescribed to treat a great number of various symptoms, mostly psychological as anxiety or tension. But it was also often administered to many pregnant women to alleviate nausea and sickness. This was the beginning of a tragedy for thousands of women around the world. Indeed, those who had taken thalidomide gave birth to children with phocomelia, a terrible congenital disorder involving limbs malformations, leading to premature death. In the U.S, the German company reached an agreement with Richardson-Merrell to market the drug, and this latter applied for approval with the FDA in 1961 when evidence of thalidomide side effects started to be reported. Of course, both the German and the American companies denied the link between the cases of phocomelia and their product. As part of the approval process, the drug was then distributed to many physicians in the U.S for testing purposes. At the FDA, one of the physicians reviewing thalidomide approval, Frances Oldham Kelsey, decided to withhold it asking for more clinical tests because of emerging evidence of serious adverse effects. Unfortunately, the testing drug samples still caused 17 reported cases of phocomelia, but Kelsey's decision was indeed a great and fortunate one and it has secured her a place in history.

So, under public pressure and after a rushed discussion, in 1962 the Congress passed a new pharmaceutical regulatory framework, inspired by Kelsey's *precautionary* attitude. First, it introduced a system of control by FDA over clinical experimentation, assigning an IND (Investigational New Drug) status to experimental drugs, and nullifying this status if clinical trial protocols were not methodologically sound or patients' rights were not respected. Second, it removed the 'automatic' approval by default after 60 days: drugs needed a 'positive' approval by the FDA to enter the market. And third, above all, it required 'substantive evidence' of effectiveness based on 'well-controlled studies', in addition to the pre-clinical demonstration of safety. The lawmakers left the task of better specifying the meaning of those expressions to FDA experts and officers, who saw the minimum standard in 'randomized controlled trials'.

Moreover, the 1963 IND rules shaped somehow the 3-phases structure of drug testing, the form DF 1571 listed for the first time three phases of trials. The testing of a new drug is indeed a complicated and time-consuming process, and it is usually divided into three phases. Phase 1 trials are the first human studies of a new drug. They usually require few healthy volunteers and are designed to obtain preliminary information on drug safety, including side effects and dosing. Phase 2 studies involve a small number of diseased people,

and they are designed to further assess adverse effects and to offer initial data on drug efficacy. Phase 3 trials (i.e., RCTs) are reserved for experimental drugs which have shown at least some evidence of effectiveness in the previous. They include many patients (several hundred to several thousand) and are designed to gather enough information on safety and effectiveness to allow an adequate assessment of a risk/benefit ratio for the drug.

This division provides additional evidence to support my claim that RCTs become the gold standard in medical research because they better serve the political goals of regulators, compared to animal experiments and experts' judgment. Indeed, it is the design of phase 3 trials that makes it possible to objectively assess the safety and efficacy of a drug, all the previous phases are pointless to this epistemic aim. However, as it was already clear at that time, RCTs were quite challenging and demanding experiments requiring many patients to allow correct statistical inferences. As Donald Mainland noted in the '60s: «the method of controlled trials is still in its infancy; that, although the principles are simple, the art is extremely difficult» (Mainland, 1960).

Intuitively, running big experiments in humans can raise a medical scandal as well, exposing many individuals to a potentially toxic drug. This possibility would result in an even bigger scandal than thalidomide, making the fears of early critics of the pharmaceutical industry true: turning people into guinea pigs. Therefore, the early phases were introduced to provide preliminary evidence of safety, before exposing many patients to the drug. From a purely epistemic point of view, these phases are negligible, but from a political point of view, they served to protect consumers from further medical disasters.

6. CONCLUSION

In conclusion, there is historical evidence to warrant the idea that a series of pharmaceutical scandals pushed the US drug regulation in the direction of tighter and tighter controls, leading ultimately to the adoption of RCTs as the current safety and efficacy standard. Recognizing that RCTs have been adopted in response to the pressure of pharmaceutical consumers in Western democracies through parliaments is paramount to many discussions on the strengths and weaknesses of such methodology. For instance, philosophers of science have long discussed the limits of RCTs to support causal claims (see e.g., Cartwright, 2010, 2011) and have proposed several alternatives which might better suit this aim. However, if the main aim of regulatory trials is not epistemic (i.e., warranting causal claims), but rather political (i.e., protecting people from pharmaceutical catastrophes), then we should also assess how RCTs perform in achieving this latter goal. With regard to this, RCTs may have performed quite well so far, since [in the last decades] we did not assist to any thalidomide-like scandal. Therefore, it seems that there is not a real need for alternatives. Moreover, as some scholars have suggested, despite the epistemic limitations in assessing causality RCTs have led mostly to accurate regulatory decisions, if for instance their accuracy is defined in terms of market withdrawals (see Andreoletti & Teira, 2019). So far, RCTs have served at best the goals of regulators.

FUNDING

The author thanks the Italian Ministry of Education, University, and Research for supporting this work with a PRIN grant: Progetti di Ricerca di Rilevante Interesse Nazionale-Bando 2017 Prot. 201743F9YE.

NOTES

1. Carpenter (Carpenter, 2010) and Porter (Porter, 2020) are partial remarkable exceptions. Hutchinson (Hutchison, 2016) has also made a similar point but focusing on nursing practice.

REFERENCES

Andreoletti, M., & Teira, D. (2019). Rules versus Standards: What Are the Costs of Epistemic Norms in Drug Regulation? *Science, Technology, & Human Values*, *44*(6), 1093–1115. https://doi.org/10.1177/0162243919828070.

Armitage, P. (1982). The role of randomization in clinical trials. *Statistics in Medicine*, *1*(4), 345–352. https://doi.org/10.1002/sim.4780010412.

Butler, I., & Drakeford, M. (2003). *Social Policy, Social Welfare and Scandal: How British Public Policy is Made. Palgrave Macmillan UK. https://doi.org/10.1057/9780230554467.*

Byar, D. P., Simon, R. M., Friedewald, W. T., Schlesselman, J. J.,

DeMets, D. L., Ellenberg, J. H., ... & Ware, J. H. (1976). Randomized clinical trials: perspectives on some recent ideas. *New England Journal of Medicine*, *295*(2), 74-80.

Carpenter, D. P. (2010). *Reputation and power: Organizational image and pharmaceutical regulation at the FDA*. Princeton University Press.

Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical Studies*, *147*(1), 59–70.

Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. *The Lancet*, *377*(9775), 1400–1401.

Chalmers, I. (2011). Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. *Journal of the Royal Society of Medicine*, *104*(9), 383–386. https://doi.org/10.1258/jrsm.2011.11k023.

Gaudillière, J.-P., & Hess, V. (2013). *Ways of regulating drugs in the 19th and 20th centuries*. Palgrave Macmillan.

Greenberg, M. D. (1999). AIDS, Experimental Drug Approval, and the FDA New Drug Screening Process. *NYUJ Legis. & Pub. Pol'y*, 3, 295.

Hutchison, J. S. (2016). Scandals in health-care: Their impact on health policy and nursing. *Nursing Inquiry*, *23*(1), 32–41. https://doi.org/10.1111/nin.12115.

Kaptchuk, T. J. (1998). Intentional Ignorance: A History of Blind Assessment and Placebo Controls in *Medicine. Bulletin of the History of Medicine, 72*(3), 389-433. doi:10.1353/bhm.1998.0159.

Junod, S. W. (2008). FDA and clinical drug trials: A short history. *A Quick Guide to Clinical Trials*, 25–55.

Mainland, D. (1960). The clinical trial—Some difficulties and suggestions. *Journal of Chronic Diseases*, *11*(5), 484–496. https://doi.org/10.1016/0021-9681(60)90013-8.

Marks, H. M. (2000). *The progress of experiment: Science and therapeutic reform in the United States, 1900-1990*. Cambridge Univ. Press.

Meldrum, M. L. (2000). A Brief History of the Randomized Controlled Trial. *Hematology/Oncology Clinics of North America*, 14(4), 745–760. https://doi.org/10.1016/S0889-8588(05)70309-9.

Porter, T. M. (2020). *Trust in numbers: The pursuit of objectivity in science and public life* (New edition). Princeton University Press.

Rosenberg, C. E. (1997). *No other gods: On science and American social thought* (Rev. and expanded ed). Johns Hopkins University Press.

Salsburg, D. (2002). *The lady tasting tea: How statistics revolutionized science in the twentieth century* (1. Holt Pp. Ed). Holt.

Schickore, J. (2011). What Does History Matter to Philosophy of Science? The Concept of Replication and the Methodology of Experiments. *Journal of the Philosophy of History*, *5*(3), 513–532. https://doi.org/10.1163/187226311X599934.

Shapiro, A. K., & Shapiro, E. (2000). *The powerful placebo: From ancient priest to modern physician. JHU Press.*

Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation. (1948). *BMJ*, *2*(4582), 769–782. https://doi.org/10.1136/bmj.2.4582.769.

Temin, P. (1980). *Taking Your Medicine: Drug Regulation in the United States*. https://doi.org/10.4159/harvard.9780674592780.

Temin, P. (1985). Government Actions In Times Of Crisis: Lessons From The History Of Drug Regulation. *Journal of Social History*, *18*(3), 433–438. https://doi.org/10.1353/jsh/18.3.433.

Thompson, J. B. (2000). *Political scandal: Power and visibility in the media age*. Polity Press Blackwell.

Wax, P. M. (1995). Elixirs, Diluents, and the Passage of the 1938 Federal Food, Drug and Cosmetic Act. *Annals of Internal Medicine*, *122*(6), 456. https://doi.org/10.7326/0003-4819-122-6-199503150-00009.

Yoshioka, A. (1998). Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ*, *317*(7167), 1220–1223. https://doi.org/10.1136/bmj.317.7167.1220.

Fondazione Umberto Veronesi
– per il progresso delle scienze