

The Babel of Drugs: On the Consequences of Evidential Pluralism in Pharmaceutical Regulation and Regulatory Data Journeys



Niccolò Tempini and David Teira

Abstract Throughout the last century, pharmaceutical regulators all over the world have used various methods to test medical treatments. From 1962 until 2016, the Randomized Clinical Trial (RCT) was the reference test for most regulatory agencies. Today, the standards are about to change, and in this chapter we draw on the idea of the data journey to illuminate the trade-offs involved. The 21st Century Cures Act (21CCA) allows for the use of Electronic Health Records (EHRs) for the assessment of different treatment indications for already approved drugs. This might arguably shorten the testing period, bringing treatments to patients faster. Yet, EHR are not generated for testing purposes and no amount of standardization and curation can fully make up for their potential flaws as evidence of safety and efficacy. The more noise in the data, the more mistakes regulators are likely to make in granting market access to new drugs. In this paper we will discuss the different dimensions of this journey: the different sources and levels of curation involved, the speed at which they can travel, and the level of risk of regulatory error involved as compared with the RCT standard. We are going to defend that what counts as evidence, at the end of the journey, depends on the risk definition and threshold regulators work with.

1 Introduction

Since the early 1900s, the US Food and Drug Administration (FDA) has been using laboratory and clinical experiments to test toxicity and safety of pharmaceutical compounds before approving their release to the market (Carpenter 2010). In other words, the FDA sets the threshold of risk that patients can take when they decide

N. Tempini

Department of Sociology, Philosophy and Anthropology & Exeter Centre for the Study of the Life Sciences (Egenis), Exeter, UK

Alan Turing Institute, London, UK

e-mail: n.tempini@exeter.ac.uk

D. Teira (✉)

Department of Logic, History and Philosophy of Science, UNED, Madrid, Spain

e-mail: dteira@fsof.uned.es

© The Author(s) 2020

S. Leonelli, N. Tempini (eds.), *Data Journeys in the Sciences*,

https://doi.org/10.1007/978-3-030-37177-7_11

between treatment options. However, for the last 50 years, only data from highly standardized experiments, Randomized Clinical Trials (RCTs), has counted as legitimate regulatory evidence for market approval. This regime is changing. The 21st Century Cures Act (21CAA) invites the FDA to consider new evidentiary standards in assessing treatments, including data from Electronic Health Records (EHR).¹ For more than 50 years pharmaceutical regulators dealt with evidence mainly originating in single-purpose drug tests. The 21CCA allows them to use data (1) generated with goals different than testing and, (2) *repackaged* and *re-purposed* to assess the safety and efficacy of a treatment. *Travelling data* (Leonelli 2016) enter the field of pharmaceutical regulation.² This chapter tries to understand *what the use of different kinds of data for pharmaceutical regulation means for the assessment and comparison of risks linked to different drugs*.

The FDA is a unique institution. Its gatekeeping power is based on scientific evidence: successful tests are a pre-condition for market access. It is the most influential regulatory agency, setting the regulatory paradigm and benchmarks that others all over the world follow. Any shift has great magnitude, as FDA decisions shape a global market for prescription drugs: in 2016, worldwide sales have been estimated in 768\$bn (EvaluatePharma 2017).

Pharmaceutical regulators have to strike a balance between very powerful and conflicting institutional principles: access to worthy new compounds should be granted quickly, while guaranteeing strict thresholds of safety and efficacy. They also need to face conflicts of interest in the industry (e.g., patent versus generic manufacturers) and also among patients (e.g., depending on their risk aversion in trying new drugs). Regulatory tests have enforced the epistemic standard that arbitrates these conflicting principles: if a promising drug fails, the sponsor will lose the investment and patients their hope, but the result will be accepted.

The 21CCA introduces evidentiary pluralism in drug testing: instead of a single source of *regulatory objectivity* (Cambrosio et al. 2006), the FDA will define different standards about what counts as evidence of safety and efficacy in a treatment for each research design allowed, including how to evaluate EHR that can now be put in motion and journey towards yet new uses (Leonelli, [this volume](#), Chap. 1). The implications are not clear. Some welcome the initiative as necessary for bringing

¹Signed by President Obama on December 13, 2016, the 21CCA is law aimed at expediting “the discovery, development, and delivery of new treatments and cures and maintain America’s global status as the leader in biomedical innovation.” Section 2061 invites the FDA “to issue guidance that addresses using alternative statistical methods in clinical trials and in the development and review of drugs.” Section 2062, on which this papers focus, claims: “To support approval of a drug for a new indication, the FDA must evaluate the use of evidence from clinical experience (in place of evidence from clinical trials) and establish a streamlined data review program”. For the full text, see: <https://www.congress.gov/bill/114th-congress/house-bill/6> (Accessed on August 24, 2107)

²Sabina Leonelli theorizes the travelling of data as the achievement of purposeful strategies aimed at enabling the reuse of scientific data in situations and for uses that can be different from those originating the same data (Leonelli, [this volume](#)). To travel, it is paramount that data keep the capacity to hold evidential value in new stages of the inquiry (Leonelli 2016). She calls these strategies of data repurposing *packaging*.

new treatments to the market, others denounce it as paving the way for pharmaceutical fraud.³ We contribute to this ongoing debate with an analysis of the epistemic and political implications of the use of EHR for drug assessment: what can we expect from regulatory decisions based on these data?

EHR are systems digitalizing medical records in standardized formats, to gather information regarding a) patients, as obtained during their visits to medical facilities (e.g. clinical interview, anamnesis and assessment, or diagnosis in an emergency room); b) the complementary evidence generated by those visits (e.g., medical imagery, test results); and c) data gathered by measurement devices that patients wear and use while away from the point of care (e.g., glucose sensors inserted under the skin). EHR data are generated in the context of routine activities that are shaped not only by standards of care but also statistics, administration and billing. They are not a record of scientific observation and intervention performed in isolation. They are a product of hybrid accounting practices, a record of clinical care just as much as of auditable administration (cfr. Ramsden, [this volume](#), Chap. 17; Power 1997). It is very difficult to clean the data from the traces of interactions that are not of interest, and as such the re-use of EHR requires complex arrangements and specialised expertise (Tempini and Leonelli 2018).

To use EHR for regulatory activity requires different standards of practice and evidence than those involved in the evaluation of RCT results. Drawing on the Guidance documents so far issued by the FDA and on our own fieldwork in EHR reuse,⁴ in Sect. 4 we argue that the successful use of EHR in tests depends on adequate data management. In order to control for bias, experts need information about potential confounders. This should be included in the travel package (cfr. Leonelli 2016) of the EHR, or be otherwise available. Against a popular belief, we shall here rehearse an old statistical argument about how Big Data, on its own, is not going to correct for such biases (also Boyd and Crawford 2012). The implication is that the evidential standards of the new regulatory pluralism will be different: *what counts as evidence will depend on the risk threshold one works with*. Risk involved in using drugs approved through new testing standards might be passed downstream, to patients and their carers. They will have to decide whether to take the risks involved in taking drugs tested with inferior evidentiary standards.

In Sect. 2, we will defend the claim that the pre-21CCA regulatory regime hinged on two value judgments: the FDA should (a) behave as a strongly paternalist regulator and (b) adopt the RCT as sole source of evidence for safety and efficacy. In Sect.

³For a glimpse of this debate, see (Avorn and Kesselheim 2015) and the responses in the same issue, plus (Kesselheim and Avorn 2017)

⁴This paper draws on the qualitative fieldwork of one of us (Tempini) on two leading EHR-reuse infrastructures: the Secure Anonymised Information Linkage databank (SAIL) and the Medical and Environmental Data Mashup Infrastructure (MEDMI). Visiting and interviewing developers, managers, data scientists and clinical researchers between September 2015 and January 2017, he approached these infrastructures with a view to document the associations between organisational forms and processes, activities of infrastructure development, data practices, and scientific research practice that underlie the debate on EHRs.

3, we show how the relaxation of standards of evidence (b) has also relaxed regulatory paternalism, giving patients more access to treatments approved on different sources of evidence. Regulating with travelling EHR data would imply a further step away from paternalism: we still don't know how good this evidence is as a source of decisions about treatments. In Sects. 4 and 5, we review some of the challenges involved in standardizing evidence from EHR. We advise paternalistic caution arguing that even the most libertarian patients would want to know how reliable a testing standard is, in order to make an informed decision about treatment options.

2 Regulation with Non-travelling Data

For more than half a century, the international paradigm in drug regulation was set by the 1962 amendment to the FDA Act: a pharmaceutical company seeking approval for the commercialization of a new treatment should submit “adequate and well-controlled clinical studies” for evidence of efficacy and safety. The definition of a well-controlled study would not be clarified until 1970, when it was defined as two well-controlled clinical trials.

Testing treatments with RCTs is a long process. RCT data are gathered according to a research protocol in which statistical considerations are paramount. Treatment effects should be estimated in trials that have a given statistical *power*: the patients' sample size will determine the probability of making a type I error (accepting into the market inferior treatments). Once the administration of the treatment and the measurement intervals are pre-established in the trial protocol, the duration of the trial will mostly depend on the amount of time it takes to enroll the predesignated number of patients and consequently execute the protocol. According to (DiMasi et al. 2016), the average time from the start of clinical testing to marketing approval is 96.8 months. Administering a treatment to a patient may take weeks or months until the time comes to measure the target outcome. Gathering the data for enough completed treatment protocols may take years, as enrolment is difficult and time-consuming and depends on the condition of interest and the eligibility criteria for the participants. Even well-funded, high-impact research fields suffer from slow trials: for instance, only 3–5% of cancer patients enrol in trials (Bell and Balneaves 2015).

From the standpoint of many patients, the wait is too long. Although some of them might have early access to the drug through trial participation (e.g., via Right to Try Laws: Carrieri et al. 2018), everybody else wait till market approval to benefit from the treatment. Even the “luckiest” few, those patients who benefit from an effective drug within the trial, might have to wait for years after the protocol completion before they can access the drug again in the market. Also the industry argues that the process is too long, although for different reasons. A company will reap most profits from a compound during patent time, and this starts counting before the RCTs even start. The longer the testing and approval process, the less patent time to exploit commercially.

Why then did regulatory authorities choose this route? The main reason is normative: the 1962 Act gave the FDA a *paternalistic* gatekeeping power on pharmaceutical markets in order to protect patients from repeats of past pharmaceutical catastrophes (e.g., Thalidomide). The safety and efficacy of a product should be assessed *ex ante*, before market release. RCTs were chosen as the sole regulatory yardstick thanks to the advocacy of American pharmacologists, who defended their superiority to grasp treatment effects on the basis of a sample of patients (Marks 1997; Podolsky 2015). Other sources of evidence about treatments (e.g., case studies), until then in use to assess their effects by doctors, were discarded in regulation.⁵

The 1962 Act hinged then on two value judgments (Teira [Forthcoming](#)). First, the 1962 FDA ACT established a strongly paternalist regulatory body. Physicians and patients were deprived of treatments lacking safety and efficacy without their consent and for their own good. Second, the RCT was selected as the gold standard for determining whether a treatment was safe and efficacious. Any concerns are subordinated to the greatest good the regulator should protect: the safety and security of the pharmaceutical consumer.

With this normative justification, *the regulator exclusively evaluates non-travelling data*. Trial data are indeed designed for one-use only: testing the safety and efficacy of treatments. RCTs are not experiments to learn in which the experimenter is free to try as many things as she may see fit, in order to find out how a treatment works. RCTs are experiments to prove (Teira 2013), in which the whole test design serves the purpose to convince the regulators that a treatment is safe and effective. They are a paradigmatic example of hypothesis-driven research. RCT data are rarely re-purposed for other ends and their ‘travel equipment’ is consequently basic: datasets store the outcomes for the different variables measured, in a format suitable for statistical analysis. These data are seldom portable onwards, to the clinics where patients receive care after the trial.⁶ The situation of inquiry within which trial data are used does not usually change.

Yet, trial data move. Trials are often distributed over a number of different sites and for this reason their organization requires to carefully consider issues related to metadata and the standardization of practices. In order to speed up the testing pro-

⁵Case reports provided evidence about safety and efficacy in the first decades of the twentieth century, but they were gradually displaced by statistical trials from the 1930s onwards. The rise of Evidence Hierarchies in the 1990s consecrated the principle that statistically designed experiments could deal with the biological variability of treatment effects in patient populations. However, case reports have since attracted efforts of standardization, as [Rachel Ankeny](#) shows in her chapter [this volume](#), aimed at reclaiming their evidential value. The developments on the regulatory use of EHR we are focusing on here can be seen, in Ankeny’s terms, as part of those “emerging efforts to develop deeper understandings of appropriate, effective, and rigorous ways of using observation-based methodologies in the biomedical sciences”. In this chapter we will point out some of the limitations of these initiatives as well.

⁶As a matter of fact, experienced researchers have found difficult to simply handle them: according to ([Götzsche 2013](#)), the clinical documentation for just three drugs he tried to access in 2010 from the Swedish regulatory agency was compiled in 70 meters of binders, about half a million printed pages.

cess, trials are conducted in multiple clinical facilities, where patients are admitted and treated in accordance to a shared protocol. For the first three decades after the 1962 Act, these facilities were standard health institutions in which the trial participants were recruited among the accruing patients. From the 1990s onwards, the industry has sponsored the rise of Contract Research Organizations (CRO) that find patients wherever they are and enroll them in the trial protocol on a dedicated site, not always a conventional medical facility. By 2005, only 25% of all pharmaceutical research was conducted in academic medical centers (as opposed to 80% before 1990) (Fisher 2009).

The mobility of data is monitored by regulatory bodies with careful audit rules (Helgesson 2010). For almost 20 years now, the FDA has developed guidance documents establishing the ALCOA principles of data quality to be observed in either electronic or paper records (e.g., CDER 2018). Data should be Attributable, Legible, Contemporaneous, Original and Accurate. The records should document who created or changed a record; they should be readable (to third parties); they must contain a time stamp of its generation; they should be the first place where the data are recorded; and they should be faithful to the actual measurement. The major problem with trial data is that their mobility stops as soon as they reach the sponsoring company headquarters: according to a recent study, an astonishing 45,2% of the outcomes of the approximately 25.927 RCTs registered at ClinicalTrials.gov by major trial sponsors have not been published (Powell-Smith and Goldacre 2016). There have been prominent campaigns advocating for a legal mandate to register all the conducted trials and release the raw outcomes (e.g., AllTrials.net) and the European Union is about to implement a systematic policy in that regard.⁷

Yet, even if trial data were routinely released to the public, it would be mostly for replication and validation of the sponsor analyses. As of today, there are no systematic plans of curating these data into databases for general research purposes.⁸

3 Regulation with Travelling Data

The 1962 FDA Act established a paternalistic pharmaceutical regulator with a single standard of evidence for testing safety and efficacy. *But if the FDA approaches pharmaceutical regulation with different value judgments, we may let other kinds of data to travel and be used as evidence in regulation.* Already in the 1970s, libertarian critics of the FDA made this possibility explicit (Wardell and Lasagna 1975). If patients were allowed access to experimental treatments (under the prescription of a qualified physician and an informed consent form) regulatory agencies would ‘simply’ need to collect adverse event reports as promptly as possible. They could then proceed, when necessary, to withdraw unsafe treatments. In this anti-paternalist

⁷Through the Clinical Trial Regulation EU No. 536/2014 to be implemented in 2019.

⁸See, for instance, <http://www.alltrials.net/find-out-more/all-trials/> (Accessed on August 26, 2017)

approach, physicians and patients are free to explore treatment options. Pharmaceutical regulators exploit the data users generate with whatever statistical tools available. Adverse event data from any source should travel to the regulator's desk.

However, pharmaceutical regulation is not only about *ends*: it is also a matter of *means*. In the 1970s, such a reporting system would have been probably paper-based and relatively slow in processing and acting upon information. Contergan, the German brand name of the sleeping pill sold in the US as Thalidomide, was withdrawn from the German market 'only' a couple of months after its adverse effects were noticed in a medical journal. But at that point, 4000 children had already been born with severe deformations (Gaudillière and Hess 2012, pp. 1–2). Even a libertarian regulator could be averse to the possibility of a pharmaceutical catastrophe with too many patients harmed for delayed reporting, detection and reaction.

Both ends and means have shifted throughout the last five decades. First of all, *regulatory paternalism has been gradually relaxed*, mostly after the participants' revolts during the antiretroviral AZT trials in the 1980s (Epstein 1996). AIDS patients advocated for their freedom to decide which treatment to take, against trial designs that imposed placebos on some of them. In response to their demands, the FDA introduced an early access system based on quicker trials with surrogate endpoints: instead of following the treatment until its final outcome, the trials tracked a variable that predicted this outcome, shortening the testing process. However, this prediction may fail. Critics of the pharmaceutical industry have denounced that treatments tested in trials with surrogate outcomes have a different level of safety and efficacy than compounds tried in standard RCTs (Gonzalez-Moreno et al. 2015; Pease et al. 2017). In other words, the FDA offers different levels of patient protection according to the testing standard it chooses. Nonetheless, patients (with or without the support of the pharmaceutical industry) have continued to advocate their right to try experimental treatments, even when there is no solid RCT evidence to support them.⁹ Although the FDA remains the gatekeeper to the pharmaceutical marketplace, its paternalism has been implicitly softened with the relaxation of its testing standards.

As to the means, during the last decade, the rise of computing and digital networks enabled the diffusion of the *electronic health record* (EHR): according to the regulator, EHR systems are "electronic platforms that contain individual electronic health records for patients and are maintained by health care organizations and institutions." (FDA 2016, p. 4). With EHR clinical data can start travelling more easily, but the landscape is fragmented. There are multiple sources for EHR and many different ways to exploit them. In the first place, there are hospitals and all sorts of medical institutions (from physician offices to multi-speciality practices), but also insurance claims databases and registries. The multitude of vendors providing EHR systems has made the achievement of data interoperability and comparability a long-term issue requiring sustained standardization efforts. Recently, relative advancements in standardization united with cheap availability of enormous com-

⁹See, for instance, the recent FDA decision to approve eteplirsen (Exondys 51), against the recommendation of its own scientific board but accepting the demands of patients with Duchenne muscular dystrophy (Kesselheim and Avorn 2016).

puting capabilities have made it possible for some infrastructures to achieve a scale of data integration that could only dreamed of only a decade ago.

E.g., Kaiser Permanente is a US based integrated managed care consortium with 11.7 million health plan members as of October 2017 (Wikipedia, March 1, 2018). It is now constructing a virtual data warehouse, with a view to study the effectiveness and safety of the treatments prescribed. Kaiser Permanente is just one of the sources feeding the Sentinel initiative (FDA 2018), by which the FDA is monitoring the safety of medical products already in the market, drawing on normalized and validated records from a group of data partners. As of 2017, Sentinel was accessing data from 193 million individuals. At an international level, the Observational Health Data Sciences and Informatics (OHDSI) is a collaboration between researches in 12 countries based on a Common Data Model that specifies how to encode and store clinical data. As of 2016, there were 52 databases, with a total of 682 million patient records (Hripcsak et al. 2016).

Yet, as of today, these are all pioneering initiatives: database interoperability and standardization is not the norm (Fleming et al. 2014; Ford et al. 2009; Lyons et al. 2009). EHR are extensively used in healthcare management, both for administrative and clinical purposes. The use of EHR for other purposes is mostly derivative. Scientific concerns have not been top priority in EHR design and practice. The generation and maintenance of EHR data has instead been shaped by the situated requirements of healthcare, local information infrastructure and institutional routines and reporting policies. It is thus difficult to render different sets of EHRs comparable (Demir and Murtagh 2013). This requires intensive “cleaning”, curation and external validation. Furthermore, there are serious privacy issues: EHR contain personal information and there are a number of legal and procedural principles that should be observed in their handling. Most EHR are not ready-made to travel onwards for scientific reuse.

The travelling of EHR data thus needs to be achieved through methodological, technological and organizational solutions. An increasingly frequent approach has been the creation of secure analytical environments, where researchers can transform datasets to suit their research needs (see Tempini, *this volume*, Chap. 13). Data transform operations are carried out through a combination of automated pipelines and human judgement and intervention. A deep knowledge of the idiosyncrasies of each dataset is paramount and some data infrastructures have dedicated data analysts to provide just such expertise (Tempini and Leonelli 2018).

New developments in respect to both ends and means of regulation set the foundations for the 21st Century Cures Act (21CCA). Epistemic and methodological novelties come together in section 2062 of the 21CCA, which opens up the possibility of using electronic health records to assess new indications for already approved treatments. It mandates the FDA “to use of evidence from clinical experience (in place of evidence from clinical trials)” and “establish a streamlined data review program” in order to support approval of a drug for new indications.

Drug repositioning is indeed a booming field (Institute of Medicine 2014). Once drugs are in the market, physicians are free to prescribe them as they see fit. Pharmaceutical companies cannot promote off-label prescription, since regulatory protection against any adverse effect liability extends as far as the indications

recorded in the treatment label – those tested with an RCT. Nonetheless off-label use is sometimes successful, at least *prima facie*. The 21CCA intends to capitalize on the wealth of information on off-label use captured in EHR systems to faster evaluate alternative indications. Assuming that any *safety* issue neglected in the original trial would have already emerged in the market, the 21CCA focuses on an alternative approach to *efficacy* testing: the clinical data may be sourced from many different contexts and the statistical techniques for the analysis of treatment effects should go beyond RCT hypothesis-testing.

The 21CCA has been a controversial bill: to name just a few of the lobbying groups involved, pharmaceutical, device and biotech companies reported more than \$192 million in lobbying expenses; more than two dozen patient groups reported spending \$6.4 million in disclosures that named the bill as one of their issues (Lupkin and Findlay 2016). Not all these groups were focusing on the testing standards for drug approval: the legislation is tied to a huge raise in funding for the National Institutes of Health, enough for many stakeholders in the biomedical community to support it. Yet, the 21CCA has initiated a paradigm shift in drug testing that, according to very qualified critics, will not promote “a 21st century of cures, but a return to the 19th century of frauds” (Gonsalves et al. 2016). If the new evidentiary standards for drug approval admit inferior drugs into the market, in the absence of counter-measures many patients may return to an era in which they could not tell apart good and bad treatments.

However, preferences about testing standards depend on value judgments. Many patients might want to be protected (to a given degree) by a paternalist regulator. The degree of protection they should expect depends on the testing standard for safety and efficacy. Of course, patients might be willing to take more or less risk depending on the situation, and conditions such as access to expert counselling and high quality information, ability to process complex information, and the range of options afforded by each one’s insurance plan. Our assumption is that, faced with the increasing complexity involved in evaluating treatment options, most decision takers would welcome the availability of estimates of a testing standard’s reliability. As we are going to discuss in the following section, the use of travelling data (via EHR) for drug testing poses precisely this question. Whereas so far EHR have been packaged without paying any attention to regulatory needs, the 21CCA opens up the possibility of converting EHR for regulatory purposes. If so, we may ask when do EHR provide reliable evidence for assessing new drug indications? How shall we measure and share the risks involved in a regulatory decision based on EHR?

4 How Far Can EHR Data Travel?

Perhaps it is too early for a conclusive answer. As we are writing (March 2018), the US Office of the National Coordinator for Health Information Technology is opening to public discussion how to articulate the 21CCA “trusted exchange framework”, a first step towards achieving a flow of interoperable health information

across different networks in the country. The FDA should still issue guidance documents in order to implement the new testing standards promoted by the 21CAA. It will take years until we see the full consequences of the incorporation of travelling data into drug regulatory testing.

Yet, the debate about how to use EHRs for regulatory purposes does not start from scratch. There are already a number of FDA guidance documents about the use of EHR in both clinical trials and epidemiological studies. E.g., the FDA guidance on the *Use of Electronic Health Record Data in Clinical Investigations*, issued in July 2018, refers to the use of EHR in standard regulatory trials (CDER 2018). In accordance with the ALCOA principles mentioned above, the main goal of the document is to guarantee the auditability of every data record presented for regulatory use. More relevant for our purposes is the *Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data*, an FDA guidance issued in May 2013 (CDER and CBER 2013). It implements the same approach to data audit, but adds some significant methodological caveats:

Investigators should demonstrate a complete understanding of the electronic healthcare data source and its appropriateness to address specific hypotheses. Because existing electronic healthcare data systems were generated for purposes other than drug safety investigations, it is important that investigators understand their potential limitations and make provisions to use the data systems appropriately. (FDA 2013, p. 13)

The implicit principle is that RCT (non-travelling) data are the evidentiary benchmark for assessing the appropriateness of EHR for drug safety investigation. If this is the case, the regulator may expect that the packaging of the EHR should include enough information for the investigator to perform the assessment of their *limitations* (as compared to RCTs), and/or human expertise to be otherwise available. But the Guidance assumes no standardized curation defining a suitable EHR. It rather leaves in the investigator's hands, the internal and external evaluation of the EHR. As to the internal assessment, the investigator should evaluate best strategies for data coding (a key step of data repurposing): "Safety outcomes that cannot be identified using International Classification of Diseases (ICD) codes cannot be appropriately studied using data sources that rely solely on ICD codes in claims data" (FDA 2013, p. 14). As to the external assessment, using again an example from the Guidance, administrative claims data generated to support payment for care should be used taking into account the payor's policies governing the approval and denial of such payments, in order not to introduce a selection bias in the analysis (e.g., patients who should have been included for clinical reasons do not leave a record if payment is denied and treatment discontinued). Investigators will need to dispose of a wealth of informal knowledge about the practices and institutional shifts that shape clinical reporting. As other auditing practices, the assessment of EHR remains opaque as to the definition of its own core matter, and is shaped by economic constraints and attitudes towards cost-benefit trade-offs (see Power 1997).

The Guidance lists, without any pretence of completeness, a number of dimensions in the EHR that investigators should consider when assessing their databases.

In order to grasp the general principle behind this assessment, we need to understand first how the comparative benchmark works. RCTs are experiments designed to generate data that would allow for a clear test of a pharmacological hypothesis: for a given population, is the new treatment for condition X equal or better than an already established alternative? The design should exclude whatever potential confounders may interfere in the outcome variables measured. RCTs should compare like with like: the circumstances of the patients in every arm of the trial should be the same except for the interventions under study, so that if any difference is observed in the outcome, we may safely attribute it to the treatment administered. Sameness between trial arms is constructed by adopting control measures for a list of potential confounders (e.g., since the expectations of patients on the treatments they receive may play a role in the outcome, they should be administered in a way that no patient can discern which of the treatments in the test they are receiving – *blinding*). Throughout the last five decades, trialists have accumulated a good understanding of the different sources of bias in their experimental setups and have organized checklists to score the reliability of a test (Higgins et al. 2011).

In order to match this level of experimental control of the treatment effect, observational studies should measure potential confounders associated with the outcome and conduct an adjusted statistical analysis that accounts for differences in the distribution of these factors between intervention and control groups. Exceptionally, the size of the treatment effect might be so large as to swamp all the potential confounders (Glasziou et al. 2007). But most RCTs do not observe very large effects – and when they do, they are not necessarily reliable (Nagendran et al. 2016). The regulator using EHR data must expect that they are packaged for travel with enough information about potential confounders as to grant a solid assessment. It is then necessary that experts construct one such list of items for EHR studies, scoring the degree of control on the treatment effects allowed by a given EHR dataset. They will need to ensure that procedures are in place to operationalize expert knowledge of the specific datasets in ways that are accountable to the regulator. For these data to travel, researchers will need to explicitly account for the uniquely contextual features of each data ‘assemblage’ used in a study.

As of now, best experiences in EHR re-use (exemplified in infrastructures like SAIL in the United Kingdom – Ford et al. 2009; Lyons et al. 2009) suggest that dedicated data analysts are in the best position to develop deep knowledge of the potential sources of bias in the EHR they specialise in. Specialised data analysts flank researchers in the selection, modelling, extraction and analysis of the dataset while at the same time relying on the clinical expertise of the researchers (Fleming et al. 2014). Knowing directly about the quality of the data and getting past the initial selection of variables from a list of available sources is of paramount importance (see Tempini and Leonelli 2018). Issues such as missing, unknown or uncertain values are endemic in EHR datasets. Data collection practices in the health care system are greatly variable. Even a high quality curated EHR cohort can sometimes offer only limited coverage for important confounders. For example, in the creation of an electronic cohort of children living in Wales (the Welsh Electronic Cohort for Children), data about maternal smoking could be missing for up to 50% and contrib-

uted to the redefinition of the sample set (in this case, the cohort came to comprise children born in Wales, because children who moved to Wales after birth had comparatively poorer data). Source databases could disagree on the sex of a child, requiring researchers to harmonize even ‘basic’ data. More generally, missing data can be detected a) at the level of the individual records; b) at the institutional level (values can be missing from all the records contributed by one organization); or c) at the infrastructural level (all records from a particular EHR software vendor). US and UK systems are fragmented into multiple infrastructures marketed by competing vendors, though industry concentration is increasing. EHR data to be made available for re-use and travel are sometimes selected by vendor, again with an uncertain effect on sampling.

There is a growing literature on the EHR study biases (Pivovarov et al. 2014; Rusanov et al. 2014; Vawdrey and Hripcsak 2013). A key concern is with the *event-based nature* of EHR data (see Jorm 2015): data are collected in the occasion of patient encounters with the healthcare system. The timing and reason for these encounters have not been pre-emptively stipulated by a study protocol and are instead associated to patient needs. Data about healthier patients are thus scarcer, and this can have implications for sample selection. Shifting reporting policies (administrative, accounting and fiscal frameworks) and other circumstances of health care coverage can mean certain phenomena are under- or over-reported (Dixon et al. 2015; Fleming et al. 2014). In addition, algorithms used for curation and modelling need to be validated: coding can be simplistic and/or overlapping, often requiring researchers to create custom code-lists and control for duplication. A complex ecosystem of practices, solutions and institutions is necessary to make scientific reuse of EHR possible (Hripcsak and Albers 2013). In 2014, a review of the state of EHR implementations in the US found that only a small proportion of systems meet “meaningful use guidelines”; while most systems met basic standards for data collection, only by 40–60% of systems satisfied criteria for the sharing of data between points of care and with public health agencies (Adler-Milstein et al. 2014). Underperforming systems are not randomly distributed and have a higher share in small and rural hospitals.

How should we think about these caveats? A quick rejoinder would contest the status of RCT data as benchmark in this comparison: many philosophers of science have defended evidentiary pluralism regarding medical causality, contesting the gold standard status of RCTs – for a review, see (Reiss 2017). Although, a priori, RCTs allow a high degree of causal control on an intervention, the theoretical assumptions behind this superiority may not hold empirically and, depending on the context, observational studies might be equally defensible. In other words, biases may equally harm RCTs and observational studies – see (Senn 2013) for a discussion. Thanks to EHR, observational studies may reach a sample size that no RCT can match and, with adequate data mining processes, true treatment effects may be detected.

Following (Senn 2008), it is worth recalling here that observational studies can improve only to a limited extent solely thanks to the addition of data. In assessing the reliability of a statistical estimator (e.g., of a treatment effect), we depend on two magnitudes. On the one hand, we have the underlying biases in the measurement

process, arising from the methodological limitations in the study we discussed above: e.g., not properly blinded patients may distort the treatment outcome. On the other hand, there is the *standard error* of the measurement process, arising from the sheer variability between the subjects measured: not every patient reacts in the same way to the treatment and we need to find a reasonable average. The standard error is (roughly) inversely proportional to the number of subjects in the study. Here comes the power of *big data*: the bigger the number of EHR, the lower the standard error. But even if the standard error tends to zero as the sample size grows, the bias will remain constant. The only reliable approach to controlling for biases is in the design of the study, and here is where RCTs dominate. This seems to be the position of the FDA research staff as of the end of 2016:

EHR and claims data are not collected or organized with the goal of supporting research, nor have they typically been optimized for such purposes, and the accuracy and reliability of data gathered by many personal devices and health-related apps are unknown. (Sherman et al. 2016)

5 “Delivering the Proof in the Policy of Truth”

There is thus no *a priori* reason to expect that EHRs can be as reliable as conventional RCTs for regulatory purposes. If so, the 21CCA is set to push further the methodological relaxation of the FDA’s regulatory paternalism. The FDA will still act as a gatekeeper, but it will allow into the pharmaceutical marketplace drugs with as many different levels of safety and efficacy as the testing standards that are used. Just as trials with surrogate outcomes turned out to be often less reliable than old-fashioned RCTs, the assessment of new indications for already approved drugs with EHRs may introduce a new safety and efficacy threshold, and one that lowers the levels of protection for pharmaceutical consumers. The incommensurability between FDA approvals based on RCT vs EHR evidence generates a risk for clinicians and patients taking a therapeutic decision between heterogeneous options. As we put forth in the introduction, *what counts as data depends on the risk threshold one works with*. The FDA, we argue, is lowering its risk threshold, and this is allowing different kinds of data to travel and be used as evidence. Clinicians and patients will then have to set their own threshold in turn.

The residual risk involved by EHR-based regulation (i.e. risk that is not controlled by FDA regulatory activity) is thus passed downstream to patients. Each patient will take decisions based on different risk thresholds and standards of evidence. Their decision will also depend on standards of care patients are able to access and the specialists they will consult with. Will patients accept drugs based on different kinds of evidence? All evidence points to a positive answer, especially if we consider the influence that pharmaceutical marketing, once deployed to promote newly approved uses, can exert on the entire cultural frame in which health is understood and evaluated (Dumit 2012). The trend, Dumit shows, is for more and the most profitable drugs to succeed.

It remains to be seen whether and how international regulatory agencies other than the FDA will revise their position with respect to the use of EHRs in regulation. At stake, we have competing forces in pharmaceutical markets. The 21CCA is supposedly addressing the crisis of innovation in the pharmaceutical industry and bringing new treatments to patients. In adopting evidentiary pluralism, the 21CCA implicitly sanctions a popular hypothesis on the causes of the crisis: it is partly due to the high amount of treatments that are lost to inadequate testing standards. Using different sources of evidence, regulatory agencies will be able to minimize these treatment losses. From a political standpoint, the question with this regulatory shift is whose interests it serves. If all evidentiary standards were equally reliable, the interests of the industry and of many patients might be aligned. But if we are right in our diagnosis, and we are left with uncertainty as to the comparability of RCT vs EHR tests, patients may have a dilemma. In the best case scenario, new tests would bring more cures to the market, but some of these may be ineffective or even harmful. Is it worth having more treatment options available even if not all of them are equally reliable?

These shifts point negotiate the core of liberal democratic polities: the move of the FDA further away from regulatory paternalism marks a retreat of the State from protecting its citizens from harm (through legal and bureaucratic devices such as regulatory activity).¹⁰ A drug regulation framework stipulates what is the acceptable evidence of risk magnitude and risk structure (e.g., bias vs standard error) for granting approval of experimental treatments. Thus the 1938 and 1962 acts created an *ex ante* protection from harm. Before then, and since the last 150 years, only tort law was available (and still is) (Gibbs and Mackler 1987) – an *ex post* reparation for the injury caused by a compound. Following Agamben (1998), we can interpret both tort law and the FDA regulatory frameworks as core institutions of the liberal democratic polity: they protect one citizen from harm inflicted by another.

The protective power of the FDA mark of approval is complex. On the one hand, the prohibition of releasing and administering drugs that have not been tested is an example of how regulation anticipates potential harm and protects from it. On the other hand, when a drug is tested according to stipulated testing regimes and considered safe enough after evaluation of the supporting evidence, its approval by the FDA is voucher for the limited liability of the manufacturer for the harms that a pharmaceutical might still cause.¹¹ Harm eventually inflicted by the off-label use of

¹⁰The political metaphysics of the 21CCA shift in pharmaceutical regulation are stark. According to Agamben (1998), the *dispositif* of law is intended to demarcate a field of social relations that are protected from harm. Harm inflicted within the field will be punished through sanctions. However, a field cannot be demarcated without creating an opposing field outside protection. Key to the cultural and legal evolution of Western polities at least since Roman law has thus been the existence of a field of social relations that is outside the scope of law, and most importantly the continuous negotiation and redefinition of its boundaries. This is the field of *nuda vita*. Agamben argues that the existence of such a field is a necessary precondition of the social contract – complete protection of all the living would immobilize society. Note that here we are not taking a strange turn: this is the theme of the crisis of innovation, elaborated from a different perspective.

¹¹The policy defines a case logic whereby the harm inflicted despite fulfilment of due process is not to be sanctioned in the same way that harm otherwise inflicted is culpable. Any changes in drug

a drug is consequently more easily sanctioned (and pharmaceutical companies do not promote such uses for this reason).

There are of course exceptions where the manufacturer can still be liable after drug approval. This can be for negligence (defect in design, testing, manufacturing or labelling); strict liability (injury caused by avoidable reasons: does not apply if “the product is properly prepared and accompanied by proper directions and warning”); and breach of contract (does not usually apply to drugs). Key here is the assumption by the legislator that drugs have *unavoidable risks*: perfect knowledge about them is impossible. To receive satisfaction, the plaintiff should then argue that the risks were avoidable for either improper design, manufacturing or labelling.

However, for the last five decades, the existence of FDA approval has ruled out improper design or testing as a litigation pathway. The courts rarely failed manufacturers for the harm caused by properly produced and labelled FDA-approved drugs. The FDA regulation has been until now authoritatively demarcating what epistemic risks (as implied by each accepted test design) will be treated as *unavoidable* and therefore not culpable.

A framework less centred on safety such as that the 21CCA is introducing, we argue, increases patient choice while shifting some of the harm involved in taking drugs from *ex ante* to *ex post* protection devices. Whether the 21CCA will then open a new space of litigation (thus undermining the evidentiary power of FDA approvals), or if instead it will simply mean that more harms will be unsanctionable (if American courts continue the practice of accepting FDA approval as evidence of test quality), it remains to be seen. With potentially inferior testing standards regulating access to market, it becomes possible that some harm is inflicted because of failure of the inferior safety standard and which could have been avoided by a superior standard. Until the reliability of the new standards is fully grasped, patients will have to suffer the eventual consequences of lesser State protection.

In the while, we expect that individual risk aversion will shape market outcomes: some patients (and their caretakers and doctors) will welcome uncertain but more abundant treatment choices; others will not. The attitudes of US citizens towards pharmaceutical risks changed throughout the twentieth century to support increasingly strict safety regulations, at least if we judge it by Congress decisions (Carpenter 2010). Is there a public demand for more cures offsetting this previous risk aversion? And is it a well-formed demand or does it rather reflect the marketing pressure of pharmaceutical lobbies, as critics contend?

In sum, the 21CCA paves the way for the regulatory use of EHR. We have argued that, before these data start their *journey* to the regulator’s desk, it is crucial that we debate how to *package* EHR in order to make the best use of the information they provide. In designing the journey, one crucial point is how to convey the information about its potential limitations for regulatory use in a standardized format. Only

testing regulation are thus concerned with redefining the boundary between one type of harm risk (not culpable) and the other (culpable). Paternalistic policies following scandals such as Thalidomide expanded the field of protected life. With the 21CCA’s prospect of easier approval to more drugs through inferior testing standards, the boundary between the two harms is moving again but in the opposite direction.

with some degree of package standardization, we can estimate the reliability of EHR in making regulatory decision, how often they yield to error, as compared to other sources of evidence. And this is the sort of information that a robust public sphere needs to debate whether the sort of evidentiary pluralism promoted by the 21CCA is welcome. If the journey of EHR data becomes so long as to require clinicians and patients to evaluate the evidence in favour of a treatment option, it might be travelling that eventually can go too far.

Acknowledgements Tempini's work was supported by ERC grant award 335925 (DATA_SCIENCE), and by EPSRC grant EP/N510129/1.

References

- Adler-Milstein, Julia, Catherine DesRoches, Michael Furukawa, et al. 2014. More Than Half of US Hospitals Have at Least a Basic EHR, But Stage 2 Criteria Remain Challenging for Most. *Health Aff (Millwood)* 33: 1664–1671. <https://doi.org/10.1377/hlthaff.2014.0453>.
- Agamben, Giorgio. 1998. *Homo Sacer*. Stanford, CA: Stanford University Press.
- Ankeny, Rachel A. this volume. Tracing Data Journeys Through Medical Case Reports: Conceptualizing Case Reports Not as “Anecdotes” but Productive Epistemic Constructs, or Why Zebras Can Be Useful. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Avorn, Jerry., and Aaron S. Kesselheim. 2015. The 21st Century Cures Act — Will It Take Us Back in Time? *New England Journal of Medicine* 372: 2473–2475. <https://doi.org/10.1056/NEJMp1506964>.
- Bell, Jennifer A., and Lynda G. Balneaves. 2015. Cancer Patient Decision Making Related to Clinical Trial Participation: An Integrative Review with Implications for Patients' Relational Autonomy. *Support Care Cancer* 23: 1169–1196. <https://doi.org/10.1007/s00520-014-2581-9>.
- Boyd, danah m., and Kate Crawford. 2012. Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15: 662–679.
- Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2006. Regulatory Objectivity and the Generation and Management of Evidence in Medicine. *Social Science & Medicine* 63: 189–199. <https://doi.org/10.1016/j.socscimed.2005.12.007>.
- Carpenter, Daniel P. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton: Princeton University Press.
- Carrieri, Daniele, Fabio A. Peccatori, and Giovanni Boniolo. 2018. The Ethical Plausibility of the ‘Right To Try’ Laws. *Critical Reviews in Oncology/Hematology* 122: 64–71. <https://doi.org/10.1016/j.critrevonc.2017.12.014>.
- Center for Drug Evaluation and Research (CDER). 2018. Use of Electronic Health Record Data in Clinical Investigations. *FDA*. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry>. Accessed 14 Sept 2019.
- Center for Drug Evaluation and Research (CDER) & Center for Biologics Evaluation and Research (CBER). 2013. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. *FDA*. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic>. Accessed 14 Sept 2019.
- Demir, Ipek, and Madeleine J. Murtagh. 2013. Data Sharing Across Biobanks: Epistemic Values, Data Mutability and Data Incommensurability. *New Genetics and Society* 32: 350–365. <https://doi.org/10.1080/14636778.2013.846582>.

- DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. 2016. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *Journal of Health Economics* 47: 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
- Dixon, Brian E., P. Joseph. Gibson, Karen F. Comer et al. 2015. *Measuring Population Health Using Electronic Health Records: Exploring Biases and Representativeness in a Community Health Information Exchange*. Paper presented at the 15th World Congress on Health and Biomedical Informatics, MEDINFO 2015.
- Dumit, Joseph. 2012. *Drugs for life: how pharmaceutical companies define our health*. Durham, NC: Duke University Press.
- Epstein, Steven. 1996. *Impure Science. Aids and the Politics of Knowledge*. Berkeley/Los Angeles: University of California Press.
- EvaluatePharma. 2017. *World Preview 2017, Outlook to 2022*. Evaluate. <http://info.evaluategroup.com/rs/607-YGS-364/images/WP17.pdf>. Accessed 14 Sept 2019.
- FDA. 2013. *Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data*. Washington: U.S. Department of Health and Human Services.
- . 2016. Use of Electronic Health Record Data in Clinical Investigations. In *Draft Guidance*. Washington: U. S. Department of Health and Human Services.
- Fisher, Jill A. 2009. *Medical Research for Hire: The Political Economy of Pharmaceutical Clinical Trials*. New Brunswick, NJ: Rutgers University Press.
- Fleming, Lora E., Andy Haines, Brian Golding, et al. 2014. Data Mashups: Potential Contribution to Decision Support on Climate Change and Health. *International Journal of Environmental Research and Public Health* 11: 1725–1746. <https://doi.org/10.3390/ijerph110201725>.
- Food and Drug Administration (FDA). 2018. FDA's Sentinel Initiative. FDA. <https://www.fda.gov/safety/fdas-sentinel-initiative>. Accessed 14 Sept 2019.
- Ford, David V., Kerina H. Jones, Jean-Philippe Verplancke, et al. 2009. The SAIL Databank: Building a National Architecture for E-Health Research and Evaluation. *BMC Health Services Research* 9: 157. <https://doi.org/10.1186/1472-6963-9-157>.
- Gaudillière, Jean Paul, and Volker Hess. 2012. *Ways of Regulating Drugs in the 19th and 20th Centuries*. Basingstoke, UK: Palgrave Macmillan.
- Gibbs, Jeffrey N., and Bruce F. Mackler. 1987. Food and Drug Administration Regulation and Products Liability: Strong Sword, Weak Shield. *Tort & Insurance Law Journal* 22: 194–243.
- Glazious, Paul, Iain Chalmers, Michael Rawlins, et al. 2007. When Are Randomised Trials Unnecessary? Picking Signal from Noise. *The BMJ* 334: 349–351. <https://doi.org/10.1136/bmj.39070.527986.68>.
- Gonsalves, Gregg, Daniel P. Carpenter, and Joseph S. Ross (Producer). 2016. Lawmakers Must Ask Tough Questions About the 21st Century Cures Act. *The Hill*. <http://thehill.com/blogs/congress-blog/healthcare/307020-lawmakers-must-ask-tough-questions-about-the-21st-century>. Accessed 14 Sept 2019.
- Gonzalez-Moreno, Maria, Cristian Saborido, and David Teira. 2015. Disease-Mongering Through Clinical Trials. *Studies in History and Philosophy of Science* 51: 11–18. <https://doi.org/10.1016/j.shpsc.2015.02.007>.
- Gøtzsche, Peter C. 2013. *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. London: Radcliffe Publishing.
- Helgesson, Claes-Frederick. 2010. From Dirty Data to Credible Scientific Evidence: Some Practices Used to Clean Data in Large Randomised Clinical Trials. In *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*, ed. C. Will and T. Moreira, 49–64. Farnham: Ashgate.
- Higgins, Julian, Douglas G. Altman, Peter. Gøtzsche, et al. 2011. The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials. *The BMJ* 343. <https://doi.org/10.1136/bmj.d5928>.
- Hripscak, G., and D.J. Albers. 2013. Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association* 20 (1): 117–121.
- Hripscak, George., Patrick B. Ryan, John D. Duke, et al. 2016. Characterizing Treatment Pathways at Scale Using the Ohdsi Network. *Proceedings of the National Academy of Sciences* 113: 7329–7336. <https://doi.org/10.1073/pnas.1510502113>.

- Institute of Medicine. 2014. *Drug Repurposing and Repositioning: Workshop Summary*. Washington, DC: The National Academies Press.
- Jorm, L. 2015. Routinely collected data as a strategic resource for research: Priorities for methods and workforce. *Public Health Research and Practice* 25 (4): e2541540.
- Kesselheim, Aaron S., and Jerry Avorn. 2016. Approving a Problematic Muscular Dystrophy Drug: Implications for FDA Policy. *JAMA* 316: 2357–2358. <https://doi.org/10.1001/jama.2016.16437>.
- . 2017. New “21st Century Cures” Legislation: Speed and Ease vs Science. *JAMA* 317: 581–582. <https://doi.org/10.1001/jama.2016.20640>.
- Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago/London: The University of Chicago Press.
- Leonelli, Sabina. this volume. Learning from Data Journeys. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Lupkin, Steven, and Sydney Findlay, (Producer). 2016. Grab Bag of Goodies In 21st Century Cures Act. *Kaiser Health News*. <http://khn.org/news/grab-bag-of-goodies-in-21st-century-cures-act/>. Accessed 14 Sept 2019.
- Lyons, Ronan A., Kerina H. Jones, Gareth John, et al. 2009. The SAIL Databank: Linking Multiple Health and Social Care Datasets. *BMC Medical Informatics and Decision Making* 9: 3. <https://doi.org/10.1186/1472-6947-9-3>.
- Marks, Harry M. 1997. *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Cambridge/New York: Cambridge University Press.
- Nagendran, Myura, Tiago V. Pereira, Grace Kiew, et al. 2016. Very Large Treatment Effects in Randomised Trials as an Empirical Marker to Indicate Whether Subsequent Trials Are Necessary: Meta-Epidemiological Assessment. *The BMJ* 355: i5432. <https://doi.org/10.1136/bmj.i5432>.
- Pease, Alison M., Harlan M. Krumholz, Nicholas S. Downing, et al. 2017. Postapproval Studies of Drugs Initially Approved by the FDA on the Basis of Limited Evidence: Systematic Review. *The BMJ* 357: j1680. <https://doi.org/10.1136/bmj.j1680>.
- Pivovarov, Rimma., David J. Albers, Jorge L. Sepulveda, et al. 2014. Identifying and Mitigating Biases in EHR Laboratory Tests. *Journal of Biomedical Informatics* 0: 24–34. <https://doi.org/10.1016/j.jbi.2014.03.016>.
- Podolsky, Scott H. 2015. *The Antibiotic Era: Reform, Resistance, and the Pursuit of a Rational Therapeutics*. Baltimore: Johns Hopkins University Press.
- Powell-Smith, Anna, and Ben. Goldacre. 2016. The TrialsTracker: Automated Ongoing Monitoring of Failure to Share Clinical Trial Results by all Major Companies and Research Institutions [Version 1; Peer Review: 2 Approved]. *F1000Research* 5:2629.
- Power, Michael. 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Ramsden, Edmund. this volume. Realizing Healthful Housing: Devices for Data Travel in Public Health and Urban Redevelopment in the Twentieth Century United States. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Reiss, Julian. 2017. Causality and Causal Inference in Medicine. In *The Routledge Companion to Philosophy of Medicine*, ed. Miriam Solomon, Jeremy R. Simon, and Harold Kincaid, 58–70. London: Routledge.
- Rusanov, Alexander, Nicole G. Weiskopf, Shuang Wang, et al. 2014. Hidden in Plain Sight: Bias Towards Sick Patients When Sampling Patients with Sufficient Electronic Health Record Data for Research. *BMC Medical Informatics and Decision Making* 14: 51. <https://doi.org/10.1186/1472-6947-14-51>.
- Senn, Stephen. 2008. Lessons from TGN1412 and TARGET: Implications for Observational Studies and Meta-Analysis. *Pharmaceutical Statistics* 7: 294–301. <https://doi.org/10.1002/pst.322>.
- . 2013. Seven Myths of Randomisation in Clinical Trials. *Statistics in Medicine* 32: 1439–1450. <https://doi.org/10.1002/sim.5713>.
- Sherman, Rachel E., Steven A. Anderson, Gerald J. Dal Pan, et al. 2016. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine* 375: 2293–2297. <https://doi.org/10.1056/NEJMs1609216>.
- Teira, David. 2013. A Contractarian Solution to the Experimenter’s Regress. *Philosophy of Science* 80: 709–720.

- . Forthcoming. On the Normative Foundations of Pharmaceutical Regulation. In *Philosophy of Pharmacology*, ed. A. LaCaze and B. Osimani. Dordrecht: Springer.
- Tempini, Niccolò. this volume. The Reuse of Digital Computer Data: Transformation, Recombination and Generation of *Data Mixes* in Big Data Science. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Tempini, Niccolò, and Sabina Leonelli. 2018. Concealment and Discovery: The Role of Information Security in Biomedical Data Re-Use. *Social Studies of Science* 48: 663–690.
- Vawdrey, David K., and George Hripcsak. 2013. Publication Bias in Clinical Trials of Electronic Health Records. *Journal of Biomedical Informatics* 46: 139–141. <https://doi.org/10.1016/j.jbi.2012.08.007>.
- Wardell, William M., and Louis Lasagna. 1975. *Regulation and Drug Development*. Washington: American Enterprise Institute for Public Policy Research.
- Wikipedia. 2018. Kaiser Permanente. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Kaiser_Permanente&oldid=828279882 Accessed 14 Sept 2019.

Niccolò Tempini is Senior Lecturer in Data Studies at the University of Exeter, Department of Sociology, Philosophy and Anthropology, and a Turing Fellow at the Alan Turing Institute. He is an interdisciplinary social scientist interested in questions of information, data, technology, organization, value and knowledge. He researches Big Data research and digital infrastructures, investigating the specific knowledge production economies, organization forms and data management innovations that these projects engender with a focus in their social and epistemic consequences. He studies the practices of data scientists, software developers, researchers and nonprofessionalised experts to understand how different forms of knowledge and value intersect with each other when different actors come to grips with new methods and new forms of data, information technology and organization. His research has been published in international journals across science and technology studies, information systems, sociology and philosophy (more information at www.tempini.info).

David Teira is Professor at the Department of Logic, History and Philosophy of Science at UNED, the Spanish Open University. He has worked on the uses of statistical methods in economics and medicine. In particular, he has studied how the different stakeholders in a clinical trial can agree on an experimental outcome, despite their conflicts of interests (<http://www.uned.es/personal/dteira>).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

