

The animal model of human disease – a core concept of medical research: historical cases, failures, and some epistemological considerations¹

Volker Roelcke

Institute of the History of Medicine
Giessen University, Germany
volker.roelcke@histor.med.uni-giessen.de

accepted for publication in Science in Context, vol. 35, 2022

Argument

The article uses four historical case studies to address epistemological issues related to the animal model of human diseases, as distinct from model organisms in biology. The knowledge derived from animal models is widely assumed to be highly valid and predictive of the reactions by human organisms. In this contribution, three significant historical cases of failure (ca. 1890, 1960, 2006), and a closer look at the emergence of the concept around 1860/70 are used to elucidate core assumptions related to the specific practices of animal-human knowledge-transfer, and to analyse the explanations provided by historical actors after each of the failures. It is argued that the epistemological status of the animal model changed from helpful methodological tool for specific questions, but with precarious validity, to an obligatory method for the production of strong knowledge on human diseases, indicating a culture of biomedical research in human disease which took the value of this methodological tool as self-evident, and more or less beyond question for more than a century.

¹ *Acknowledgments*: Preliminary versions of this article were presented at the Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, 1 April 2019; at the Edelstein Center for the History and Philosophy of Science, Technology, and Medicine, The Hebrew University, Jerusalem, 4 April 2019; and at the Workshop *Tiermodelle menschlicher Krankheit: Historische und epistemologische Perspektiven*, Humboldt Universität Berlin, 10 September 2019. I am grateful to the participants of these events, and in particular to José Brunner, Otniel Dror, Moritz Epple, Christoph Gradmann, Eva Jablonka, Lara Keuck, and Ohad Parnes for constructive criticisms of earlier versions of the text.

In summer 1997, after years of working with rats in the laboratory, the young pathologist Frank Dombrowski at the University of Bonn was able to present to his colleagues the ‘animal model’ for a human disease, namely liver cell cancer (Dombrowski et al. 1997). In order to establish this animal model, he had to proceed through a number of stages: First, he had to identify and breed a suitable strain of rats. It needed to be a strain in which he could damage the liver cells in a way that resembled the tissue changes and metabolic abnormalities of a malignant tumour in humans. Therefore, the various strains of rats had to be exposed to a range of potential kinds of damage, and consequently tested for the ensuing pathological properties. Once the suitable strain of animals and the adequate kind of damage had been identified to create the targeted tumour, the young scientist had to find technologies and living conditions for these manipulated rats which allowed the artificial cancers to be kept stable and standardised. As a result of this protracted experimental activity, the animal model of human liver cancer could be reproduced in identical form at will.

Once the animal model was established, it functioned as a starting point and tool for further experimental research on ‘human’ liver cell cancer, in particular for investigations into the biological properties of the tumour, and into potential interventions for therapy or disease prevention. For this achievement, published in the prestigious *American Journal of Pathology*, Frank Dombrowski received much acclaim; only a few years later, he was appointed to a full professorship, associated with the position of director of a renowned university institute for pathology.

Today, animal models of human diseases are considered central research sites and technologies of knowledge production in the biomedical sciences (see e.g. Gad 2007, Hau 2014). The underlying idea is that phenomena observed in laboratory experiments on animal models may be used to understand disease mechanisms and to predict the intended effects and hazards of interventions in humans. Animal models have a two-fold status as both scientific objects (or “epistemic things”) and epistemological tools, or “technical objects”: The first aspect draws upon their status as objects of research and their being part of an experimental setting. The latter refers to their specific (epistemic) function in modelling the pathological processes of human diseases (Rheinberger 1997; see also Löwy & Gaudilliere 1998).² The importance of this concept and technology is illustrated, for example, by the fact that in 1966, a central Registry of Comparative Pathology (RCP) was established at the Armed Forces Institute of Pathology, Washington, with the objective “to serve as an information exchange for scientists interested in the study of animal models of human

² A set of analytical terms to describe and analyze animal models of human disease is provided by Huber & Keuck 2013.

disease” (Leader 1969, 553).³ Since then, numerous special issues in scientific journals as well as monographs have been published which were devoted to conceptual and practical issues related to this experimental tool, as well as the evaluation of data on specific diseases produced by it.⁴ In 2008, a specialized journal *Disease Models and Mechanisms* was established which focuses on “the use of model systems to better understand, diagnose, and treat human disease”.⁵

Such *animal models* of human diseases differ from the *model organisms* used in fields like physiology or genetics: In the latter case, “naturally” occurring processes analysed in a specific model organism are seen as predictive for analogous mechanisms in animals in general (as in the case of mutations, protein synthesis, etc.), since these processes are assumed to have a structural identity, or homology, in all animals, or in relevant subgroups.⁶ In the animal model of human diseases, by contrast, not only are animals seen as predictive for humans, but in addition, artificially created diseases in animals are supposed to represent the essential features of diseases occurring in humans (Rheinberger 2005 and 2006; Huber & Keuck 2013).⁷ As has been argued previously, “the representational relation between animal model and human patient needs to be generated and validated” (Huber & Keuck 2013, 385).

As a matter of fact, researchers in the biomedical sciences assume that the knowledge derived from experiments in such animal models, if performed properly, is highly valid, and predictive of the reactions by human organisms – for example in the

³ From 1969 onwards, the RCP published the *Comparative Pathology Bulletin*, probably the first periodical with a specific focus on issues related to animal models.

⁴ For example, in 1980, the *American Journal of Pathology* published a special issue with the proceedings of a workshop on „Needs for New Animal Models“, organized by the RCP; see Hackel 1980; Leader & Padgett 1980; for more recent handbooks, see e.g. Gad 2007, Hau 2014.

⁵ *Disease Models and Mechanisms*. Introductory message on the Homepage of the journal. 4 November 2020. < <https://dmm.biologists.org/> >; see also Freeman & St. Johnston 2008, 6: “This new journal is aimed at people who set out with an explicit goal to investigate human disease using model organisms”.

⁶ On the history of model organisms, see e.g. Holmes 1993, Kohler 1994, Creager 2002, Leonelli 2007. For a critique of a kind of historiography of model organisms that focuses on a limited number of species whose very development has become canalized through processes of standardization, and the ensuing pitfalls for an adequate history of genetics and genomics, see Ankeny 2010; the differences between *model organisms* in biology and *theoretical models* which are used e.g. in mathematics and mechanistic theorizing are discussed in Levy & Curry 2015.

⁷ For a first overview on the history of animal models of human diseases, see Bynum 1990; for an early reflection on epistemological questions, see Canguilhem 1965, 19-50, commented upon in Roelcke 2009, 17-20; for an early stage in the history of the animal model, exemplified by Robert Koch’s methodology, see Gradmann 2005b, Roelcke 2009; for the emergence of the concept and related practices prior to Koch, see Roelcke 2013;

context of trials of new drugs, or in toxicology.⁸ Thus, this kind of knowledge might be termed “strong”, in contrast to more precarious, or “weak knowledge” as derived e.g. from clinical observation.⁹ Because of this assumed validity and predictive power, research on animal models has been made an explicit prerequisite in modern ethical and legal regulations regarding human subject research, including the Nuremberg Code and the Declaration of Helsinki of the World Medical Association.¹⁰ The underlying rationale is that no research should be carried out in humans without prior trials in animals, so that human subjects will not be exposed to any hazards found to affect animals. This rationale thus assumes that, at least to some extent, humans and animals share a common nature. Thus, the experimental use of animal models is a substitute for experimentation in humans (e.g. Frenkel 1969, 160; see also Rheinberger 2005 and 2006).¹¹

Nevertheless, transferring knowledge derived from research in the animal model to human beings in order to start clinical trials apparently has not been very successful: For example, in 2012, the authors of an editorial in *Nature Reviews Drug Discovery* reported that the probability of compounds from a group of major pharmaceutical companies making it to market from initiation of phase 1 trials (the first application in humans) declined from 10% in the period 2002-2004, to 5% in 2006-2008 (N.N. [editorial] 2012). Similar numbers were found in the most comprehensive survey of “clinical success rates” of newly developed drugs covering the period between 2003 and 2011 (Hay et al. 2014). This implies that for at least nine out of ten cases, the knowledge produced in the animal model (before the start of phase 1 trials) was unsuitable for application in humans (see also e.g. Hackam & Redelmeier 2006; van der Worp et al. 2010; Cummings et al. 2014; Perrin 2014). Further, the knowledge transfer from animals to humans entails some risk, as has been documented by regular cases of massive failure. As I intend to show in the following, such failures may be used to look at some presuppositions and implications of the very concept of the animal model.

⁸ For examples from different historical contexts in the 20th and early 21st century assuming such validity, see e.g. Frenkel 1969, Jones 1980, Davidson et al. 1987, Gad 2007, Hau 2014; on criteria for appropriate models and performance, see e.g. Leader & Padgett 1980; Held 1983.

⁹ See Epple 2019 (following a distinction introduced in Plato’s *Theaetetos*) for a general proposal to distinguish between bodies of knowledge considered to be “stronger” or “weaker”.

¹⁰ On the emergence and content of the Nuremberg Code and the Declaration of Helsinki, see Weindling 2001; Lederer 2004. The first legal regulations on a national level, which required that research in animal models should precede any human subject research, were the Guidelines / *Richtlinien* issued by the German Ministry of the Interior in 1931, see Roelcke 2017.

¹¹ Some of the epistemological fallacies involved in animal model research are discussed in Shanks et al 2009 and Sjoberg 2017.

The article is divided into three parts: In a first step, I shall describe three historical cases of such failure: Robert Koch's tuberculin (in 1890), the sedative thalidomide (ca. 1960), and the immunomodulatory compound TGN1412 (in 2006) which was to target conditions such as multiple sclerosis or rheumatoid arthritis. For each case, I shall look into the explicit or implicit assumptions about animal models which were related to the specific practices of animal-human knowledge-transfer, and in addition analyse the explanations provided by relevant historical actors after each of the failures. This might indicate the extent to which the validity and limits of the animal model were re-evaluated. The development and failure of Koch's tuberculin and of the sedative thalidomide have been described previously,¹² but without particular attention to the specific assumptions of the historical protagonists about the knowledge-transfer between animals and humans, nor to the attempts of historical actors to explain the failures. For the case of TGN 1412, there does not yet exist a historical reconstruction.¹³

In an additional case study, after the presentation of the three cases of failure, I shall sketch the mid-19th-century origins of the concept of an animal model of human disease processes in the laboratory of Julius Cohnheim at the University of Breslau, and his reflections on the validity and limits of such a methodological tool. I shall employ Cohnheim's explicit reservations regarding the epistemological status of the animal model to further elucidate the potential causes for failures of the animal model and draw some tentative conclusions about the mechanisms which are at work in the construction of supposedly strong knowledge.

The specific cases were chosen for two reasons: First, they are related to decisive steps in the historical trajectory of the use of animal models: their unequivocal acceptance as central methodological tools for research into human diseases (in the case of Koch); and the introduction of core regulatory norms, that is, restrictions, for the application of knowledge derived from the animal model in humans (in the aftermath of the failures of both thalidomide and TGN1412). Second, well beyond the specificities of each individual case, they all illustrate exemplary and significant presuppositions for the transfer of knowledge from the animal model to humans, and justifications for the further use of this methodological tool in spite of the previous failures:

More specifically, the case of Koch's tuberculin did not only represent a personalized dispute between microbiology and clinical pathology in the late 19th century, but it also documents the explicit arguments and implicit assumptions of the later Nobel laureates Koch, Ehrlich, and Behring about the supposedly strong nature

¹² For references, see below in the specific sections.

¹³ For a first philosophical analysis, see Lemoine 2017.

of the knowledge derived from the animal model. In this way, it illustrates the enormous authority which the animal model of human disease had gained between Cohnheim's cautious evaluation following his initial use of the method since the 1860s, and the situation in the years around 1890. The case of thalidomide triggered a broad international debate not only on the specific compound on more general issues related to testing new drugs in animals, and thus illustrates the rationalities of the biomedical scientists involved in the drug research, and of those scientists in the international community that reacted to the failure and tried to draw conclusions from it. As some of the historical actors claimed themselves, these rationalities and reflections had an exemplary character for medical research relying on animal models of human diseases. The example of TGN1412, finally, does not only describe the case-specific issues of unexpected problems in using proteins with a new level of species-specificity. Rather, this kind of issue, and the broader debates related to the unknown dimensions of complexity of the immune system in different species prior to concrete research endeavours (albeit supposedly known in hindsight) represents one of the basic challenges of the transfer of knowledge from animals to humans – as even some of the historical actors acknowledged when they talked about a „shock wave [...] for the international scientific community, the biotechnology industry and the regulatory authorities“ or a „disaster“ (Hünig 2012: 317; Tranter et al. 2013: 164). These basic concerns were associated with debates about potential regulatory implications, and led to changes in the guidelines of the European Medicine Agency.

Before embarking on the journey through the historical cases, a note on terminology appears in order: The first explicit use of the term “animal model” has yet to be clarified, but there is some evidence that it might have occurred in the mid-20th century.¹⁴ However, for the purpose of the present paper, I suggest to adopt a somewhat pragmatic approach to allow a larger historical framework which integrates the period of the “laboratory revolution” in medicine, and what is conventionally termed the emergence of experimental medicine in the second half of the 19th century – a period in which the use of animals in medical research was rapidly rising. Claude Bernard's classic *Introduction à l'étude de la médecine expérimentale* (1865) exemplifies this historical context and the integral use of animals in this kind of understanding of medical research, although it was focused on physiology rather than disease processes.¹⁵ If we define “animal models” as living animal organisms

¹⁴ Both in systematic overviews, as well as in relation to the study of specific groups of disorders, the term was certainly used since the 1960s (e.g. Frenkel 1969; Leader 1969; McKinney 1969); see also the terminology and evidence in Leonelli et al. 2014.

¹⁵ For the emergence of the “laboratory revolution” and experimental medicine, see Cunningham & Williams 1992; Rheinberger & Hagner 1993, for the surge in the use of animals in medical research, see Bynum 1990.

which are used to study phenomena such as diseases, pathogenetic processes, toxicological effects, or behavioural traits targeted for interventions, and in which the phenomena studied are assumed to resemble those in humans, there is evidence that an implicit concept of “animal model” and related practices probably appeared in medical research in the second half of the 19th century (see e.g. Bynum 1990, Gradmann 2005b, Roelcke 2009, Kirk 2010). Due to the lack of an alternative terminological tool, I shall therefore tentatively employ the term “animal model” to also designate practices during these early phases of experimental research on human diseases, even if the historical actors themselves did not explicitly use the term.

It might be the topic of future research to map and analyse the thesaurus of terms which were used by historical actors in referring to “animal models” in this broader timeframe. A systematic reconstruction of the specific meanings and contexts of their use might indeed yield insights into the various and shifting epistemic preconditions and implications. However, in contrast to such an approach focusing on the conceptions that medical researchers used in their published work which is frequently idealised or normative (or both), the following analysis takes as its starting point a number of concrete cases of failure in the practice of experimental medicine in order to reconstruct the assumptions made by the researchers related to the use of their animal models, as well as the explanations, justifications, and conclusions drawn from the failures not only by the researchers themselves, but also by their peers and by contemporary state agencies and regulating bodies. This approach may yield a more clearly contoured picture of the preconditions, implications, and fallacies of experimental work with animals designed to clarify medical issues in humans than would a purely conceptual reconstruction. In this vein, the cautious considerations of Cohnheim, historically preceding the later failures, are used in the concluding section to further highlight the issues around the animal modelling of human diseases, and to provide some tentative insight into more general issues about preconditions and dynamics of those specific forms of knowledge which are at times regarded to be “strong”.

Three significant failures

The case of tuberculin

In November 1890 the bacteriologist and later Nobel laureate Robert Koch introduced a new substance for the therapy of tuberculosis, which he called tuberculin. Koch had been investigating the features and origins of human diseases in animal models with considerable success since the late 1870s (Gradmann 2005b).

A central precondition for Koch's research into human diseases in animals was the assumption of the transferability of knowledge between animals and humans. Without further reflection on the issue, Koch "proved" such transferability by demonstrating that a variety of human diseases might be reproduced (at least regarding criteria he perceived as crucial) in a great number of animals, such as mice, rats, rabbits, guinea pigs, etc. (Gradmann 2005b). As a result, he was convinced (and convinced others) that in the bacteriological laboratory, it was possible both to observe a disease without recourse to the diseased person, and to replace the analysis of clinical symptoms with that of the features of the disease won through the animal experiment.

Over the course of the 1880s, and following these assumptions, Koch had discovered the bacterial pathogens of tuberculosis and cholera. Parallel to these discoveries, he had enjoyed a meteoric career taking him from a country doctor's practice in the rural east of Germany to a full professorship and appointment as director of the Institute of Hygiene at Berlin University (Brock 1988; Gradmann 2005a). By the end of the 1880s, together with his French colleague and rival Louis Pasteur, Koch was perceived as the creator *of*, and protagonist *in* the new medical discipline of bacteriology, or microbiology (Gradmann 2005a; Geison 1995).

The new knowledge about human diseases caused by bacteria was constitutively linked to laboratory work on the pathogens, and to the reproduction, or production, of disease phenomena in the animal model (Gradmann 2005a, 105-134; Gradmann 2005b). The development of specific remedies against infectious diseases, then eagerly anticipated from all sides, also took place in the experimental animal, which Koch regarded as a kind of passive culture medium, or tool. It was inconceivable to him that the pathogen was not the sole cause of the disorder, but that factors associated with the host organism, such as its "constitution" (a term broadly employed in previous etiological theories), might also play a substantial role in creating the symptoms and influencing the course of a disease (Gradmann 2001).

To develop a drug against tuberculosis in humans, Koch used guinea pigs to investigate various methods of what was then called "inner disinfection" to combat the tubercle bacillus. In spring 1890 he came upon a substance that was able to stop the growth of the pathogen, which he called tuberculin. In August of that year, Koch proclaimed in public that:

"I can tell [...] this much, that guinea pigs, which are highly susceptible to the disease, no longer react to the inoculation with tubercle virus when treated with that substance, and that in guinea pigs [...], it is regularly

possible to bring the process of the disease to a complete standstill” (Koch 1890a, 659).¹⁶

For Koch it was clear after the successful experiments on the animal model that the step toward testing on humans should follow swiftly. First, he conducted a single self-experiment, and then began testing the new remedy on five healthy individuals starting in June 1890. Each of the subjects developed a high fever, pain in the limbs, and nausea, but all of these symptoms disappeared 24 hours after the injection. For Koch, this implied the harmlessness of his remedy. However, a decisive difference between his animal experiments and the testing on human beings apparently escaped his notice: All healthy human test subjects reacted strongly, while the compound made no impression at all on healthy guinea pigs.

The first *clinical* trial (that is, testing on patients, not on healthy subjects) began in mid-September. The immediate reactions to the injection, including faintness, pain in the limbs, and a rapidly climbing fever, but of short duration, mostly corresponded to what Koch had observed previously in the healthy subjects. Koch interpreted these signs as confirmation that the remedy was taking effect as he postulated. Only two months later, in November, Koch published his first article in a special issue of the journal *Deutsche Medizinische Wochenschrift*. Although barely 50 tuberculosis patients had received the new compound by this time, and the time of follow-up observation had been very short, in his article Koch was ready to declare it a harmless and effective medicine. He also claimed that even more serious forms of tuberculosis, such as lupus of the skin, or incipient consumption (*phthisis*) could be healed with the substance (Koch 1890b).

More sceptical observations by some doctors who had also tested tuberculin on patients were not included in the journal’s special issue. Instead, on the day after its publication, the famous surgeon Ernst von Bergmann, instigated by Koch, orchestrated a public demonstration of tuberculin injections in patients for an audience of high-ranking state officials. In the following weeks, specialised journals were full of reports about previously inconceivable cures. Daily, the German and the international press celebrated the success of the “Koch treatment”. Berlin soon became, as one of the leading newspapers noted, into a pilgrimage site for doctors and patients from all over the world (Elkeles 1990; Gradmann 2005a, 187-193).

However, after the first longer-term results of the injections were observed in the subsequent weeks, there was ever more critical commentary: Obviously, the fever episodes lasted longer than expected, and in some cases, they led to the death of the patients. In January, when relapses began occurring even in those patients on

¹⁶ The chronology of events up to the apparent failure is described in Gradmann 2001, with further details in Gradmann 2005a: 134-211.

whom the remedy's early fame had been based, the renowned pathologist Rudolf Virchow formulated a devastating critique. Virchow revealed that during his autopsies of the deceased patients, fresh tubercle bacteria had been found at the sites of injection (Virchow 1891). This suggested that Koch's tuberculin was not only ineffective, but rather that the substance itself might have fuelled a progression of the disease. These findings were soon confirmed by colleagues, resulting in the judgment by many physicians that the compound had failed as an effective treatment of tuberculosis, but rather had itself created clinical problems and theoretical questions (Gradmann 2005a, 197-211). For example, Friedrich Schultze from Bonn refrained from further application "since there are no criteria whatsoever on how the patient will fare if this unknown substance is applied" (quoted in Gradmann 2008, 155).

What were the conclusions which Koch himself and his followers like Paul Ehrlich or Emil Behring drew from the apparent failure of tuberculin?

Koch attributed the deterioration of clinical conditions and the fatalities to the impurities in early preparations of tuberculin. As a consequence, he launched a systematic investigation to exactly identify the efficacious component in the compound, so that he would "be able to apply it without any additional substances which might cause negative side effects" (Koch 1891, 673). The use of the animal model of tuberculosis as such as a potential cause of precarious therapeutic knowledge was obviously not one of the explanatory options for the serious problems which had occurred, an option in need of systematic analysis. On the contrary: Koch explicitly stressed that for his investigations directed at identifying the efficacious component of tuberculin, the "animal experiment and its correct evaluation" was crucial (ibid.). His previous experience had shown him that healthy animals were of no use for such a purpose, but that guinea pigs infected with tuberculosis were needed – although the dosages of tuberculin to be applied might diverge from those in humans (ibid.). And again, as in the approach leading to the original clinical application of the compound, Koch saw it as "being of the greatest interest" to get to know whether the now purified tuberculin would still show its efficacy in humans after it had been established in the animal (Koch 1891, 679). As in the procedure preceding the failure, experiments of the new agent in the animal model which Koch perceived as successful were followed by tests in a few healthy individuals (including Paul Guttman, Shibasaburo Kitasato, and August Wassermann, colleagues and researchers in Koch's laboratory). In a second step, the compound was tried in patients suffering from tuberculosis. Koch claimed that in both instances, in principle the effects were comparable to those observed in the first trials with the not yet purified compound. However, following his new observations, he stressed that

therapies should start with very low dosages, which should then gradually be increased (Koch 1891, 679-680).

Koch thus apparently did not believe that the grave empirical problems in the wake of the knowledge transfer from animals to humans were to be attributed to the method and location of knowledge production – the animal model – and the implied concept of disease (namely exclusively caused by features of the pathogen). Instead, the problems were ascribed to unintended byproducts and inadequate dosages in the process of applying the method. These byproducts were not perceived as the inherent result of the method, but instead claimed to be accessory and contingent in character. Remarkably, instead of *questioning* the method itself, Koch saw it as *essential* in his attempt to understand and clarify the problems which had occurred.

According to Koch's pupil Paul Ehrlich, himself a later Nobel laureate,¹⁷ the supposed failure was in part the effect of an exaggeration by pathological anatomists who were not qualified to judge the clinical value of the substance (Ehrlich 1891a, 18).¹⁸ Ehrlich conceded "that Koch's original mode of employing tuberculin against tuberculosis in man might in some cases prove hazardous" (Ehrlich 1891b, 920), but stressed that in the meantime, by systematic observation, the method had been advanced substantially. The implications of the patho-anatomical findings by Virchow had been overstretched; they focused on the early phase of applying the treatment, and the critical observations had only been made in those patients where the disease was in a very advanced state and who ultimately died (Ehrlich 1891a, 18-19; Ehrlich 1891b, 918). In addition, the negative conclusions had been "drawn from a relatively limited number of cases, very small as compared with those available for clinical purposes" (Ehrlich 1891b, 920). The observations during the early application "in no way invalidated the *principle* of the method of treatment, but at the most could only be directed against the *technique* and against the employment in very advanced cases" (Ehrlich 1891b, 918 [emphasis in the original]).¹⁹ Ehrlich contrasted the "exaggerated" conclusions of the pathologists with his own positive clinical findings from application of a regime starting with a low dose tuberculin, and gradually

¹⁷ On Ehrlich's biography and work in bacteriology, immunology, and pharmacology, see Parascandola 1981, Hüntelmann 2011.

¹⁸ Ehrlich formulated his comments and conclusions on „Koch's Method“ in the context of an invited lecture at the 7. International Congress of Hygiene and Demography in London (1891); the full text of the lecture was published in German (Ehrlich 1891a [1951]), a somewhat shorter version in English, „abstracted“ by Thomas W. Hime, in *The Lancet* (Ehrlich 1891b).

¹⁹ In the original German version of the text by Ehrlich, this whole sentence is emphasized by italics (Ehrlich 1891a, 14) – the only sentence marked in this way in the whole article, indicating its central importance for Ehrlich.

increasing dosages dependent on the clinical phenomena induced by the previous applications (Ehrlich 1891b, 919; see also Guttman & Ehrlich 1891).

Strikingly, in spite of his clinical experiences, which pointed to the importance of the individual condition of specific patients (Ehrlich 1891a, 17), Ehrlich's reaction to the failure neglected to mention the initial phase of Koch's research in the animal model which had focused exclusively on the biology of the infectious agent and completely bracketed out the impact of the host organism.²⁰ On the contrary, Ehrlich was convinced that Koch's methodological approach, namely testing therapeutic compounds experimentally in animals, was the only way to achieve a "rational mode of treating [human] disease in accordance with strictly systematic scientific principles", and in contrast to "pure empiricism" (Ehrlich 1891b, 918). He concluded that this methodology "must serve us as a standard in the further development of the art and science of medicine" (Ehrlich 1891b, 918; Ehrlich 1891a, 13).

Completely in tune with this thinking, Ehrlich and his colleague Behring closely followed Koch's approach in their own research aimed at finding therapeutic compounds to treat diphtheria and syphilis. For example, the necessary precondition for the development of Behring's serotherapy against diphtheria was the proof, in the animal model (in rabbits and mice), that an antitoxin serum against the disease could be transferred from one animal to another, and further, that its efficacy might be tested in the animal model.²¹ Ehrlich, in turn, used an animal model of diphtheria in guinea pigs in the classical study in which he outlined a method for the standardization of the antitoxin (Ehrlich 1897). There is no evidence that Behring, Ehrlich, or any other of Koch's followers developed any doubts about the general idea and the specific procedures to produce an animal model as the essential tool for research into human pathological conditions. By identifying the microbial pathogens of cholera and tuberculosis in animal models and thus creating a completely new understanding of these diseases, Koch had apparently attained such authority that his methodology had become untouchable.

The case of thalidomide

Beginning in 1959, physicians as well as parents in Germany, Brazil, Australia, and elsewhere suddenly encountered a wave of children born with severe abnormalities

²⁰ It may be noted here that at this time (around 1890/91), Ehrlich's theories on questions of immunity were only in their earliest stages, and that there existed no spelled out ideas about issues such as innate and acquired immunity, species specificity, etc.

²¹ Behring 1890; Behring & Kitasato 1890; on Behring's work related to the serotherapy against diphtheria, see Linton 2005, Simon 2007, Gradmann 2008.

(see e.g. the report in Wiedemann 1961).²² The defects included stunted arms and legs, as well as misshapen hands and feet: children's hands and feet appeared to grow directly from their torsos. In addition, damaged internal organs had been observed, including hearts and kidneys. Previously, this kind of condition had been described as extremely rare in the medical literature, but now such cases were rapidly on the rise in paediatric hospitals and practices. By the time the responsible drug, thalidomide, was identified as the culprit and withdrawn from the market in the early 1960s, some 10.000 children around the world had been born with such deformities. The substance obviously acted as a toxic agent in an early stage of embryonic development during which the formation and growth of bones and some internal organs were particularly vulnerable.

Thalidomide's initial development and testing had offered no evidence that it would be a potentially toxic agent for the human embryo – a result of the specific questions posed when it was put to trial. The compound had been synthesized in the early 1950s in the laboratories of Chemie Grünenthal, a rather small chemical and pharmaceutical company in Germany. Between 1954 and 1957, experiments were performed on animals to establish the essential properties and exclude toxicity. The animal studies included mice, rats, guinea pigs, rabbits, and dogs. Even when administered in high dosages, no adverse effects could be detected regarding blood counts, nor cardiac, pulmonary, and temperature regulation, nor regarding renal function. In order to model – that is, to simplify, standardize, and quantify – the intended sedative effect, sedation was operationalized as reduced motility. To measure this, the mice were set in a “tremble cage” (*Zitterkäfig*) (Kunz et al. 1956). Here, their movements, through an elaborate apparatus, caused the electrolysis of sulphuric acid, and the amount of emerging hydrogen was taken as parameter of the (reduced) motility of the animals. The potential loss of coordination was measured by the duration of the animals' ability to cling to a wooden stick by holding reflexes.

In a first publication of the results of the preclinical tests in a journal of pharmacology, the responsible scientists described “relatively strong sedative effects” and a “complete lack of toxicity”. They concluded that the data from animal experimentation justified trials in humans (Kunz et al. 1956, 429). In 1955, such clinical trials started at various German university hospitals. The ensuing publications, referring to several hundred patients, confirmed the sedative properties and claimed that only occasionally, minor side effects had occurred, such as headaches or obstipation (Jung 1956; Esser & Heinzler 1956; Stärk 1956).

²² For general reconstructions of the thalidomide case, see Kirk 1999; Daemrlich 2002; Großbölting & Lenhard-Schramm 2017; the following account on the introduction of the compound closely follows Daemrlich 2004,60-69.

From early 1957 onwards, the drug was marketed as a treatment for insomnia, anxiety, and morning sickness in pregnancies.²³ It appeared to offer a breakthrough in terms of side effects and safety. Patients reported that it induced a very pleasant sleep with few after-effects. Consequently, Grünenthal marketed the drug not only as a treatment for sleeping problems and against morning sickness, but also suggested its application as a break from stressful work life. By 1960, thalidomide was the top selling sedative in Germany. The drug was also sold under licensing agreements in more than 40 countries by the same year.

Beginning with the meeting of the German Association for Paediatric Medicine in September 1960, several physicians informally discussed the impression that thalidomide was causing the surge of malformations in newborns diagnosed in hospitals around the country. The Hamburg paediatrician and medical geneticist Widukind Lenz emerged as the most vocal proponent of this hypothesis. In 1961, he began compiling detailed reports of cases from his own patients as well as from other regional hospitals. Within a short time, he had collected almost forty cases of children that he linked to the drug. In a regional meeting of the Paediatric Association in November 1961, he reported his findings, a presentation which had considerable repercussions beyond the medical community as well (Kirk 1999, 155; Daemrich 2004, 62). At the same time, William McBride, an Australian physician, published the first article which postulated a causal connection between the malformations and thalidomide; Lenz followed with a short report in *The Lancet* in January 1962 (McBride 1961; Lenz 1962a und 1962b; for the chronology, see Lenz 1988).

During 1962, Lenz and other physicians accumulated further data linking thalidomide to birth defects. Subsequent publications turned from individual case reports to aggregated data sets (Daemrich 2004, 62; for examples, see Wiedemann 1961 and 1962; Dijkhuis et al. 1962; Lenz & Knapp 1962; Pfeiffer & Kosenow 1962). The evidence showed that the wave of birth defects followed the market introduction of the drug with a time lag of several months; the geographical distribution of the characteristic malformations coincided with the areas where the drug was marketed, and in countries where thalidomide was not available, there were no children born with the malformations (McBride 1961; Dijkhuis et al. 1962; Ward 1962; Wegerle 1962; Weicker & Hungerland 1962).

By that time, the evidence for a causal link between the intake of thalidomide and the occurrence of the birth defects was such that the drug was withdrawn from the market in the majority of the countries in which it had been for sale (Kirk 1999, 155-163). Another consequence of the statistical data was the turn to a second

²³ Before that, the compound had been marketed for a few months in the winter 1956/57 as an anti-influenza drug: Wimmelbücker 2017.

phase of animal experimentation: Since all available evidence pointed to the hypothesis of a toxic effect of the compound during the early phase of pregnancy, it was now clear that this specific pathogenic mechanism had not been looked at in the first stage of animal testing prior to the clinical trials, simply because no pregnant animals had been used. This was, however, in spite of ample contemporary evidence that a large number of compounds had been demonstrated to cause congenital malformations if administered to pregnant experimental animals (e.g. Woollam 1962b).

However, the new attempts to model the toxic effect in pregnant animals did not yield the results which had been envisaged: Almost none of the experiments in various strains of mice, in rats, or in cats and chicken yielded the characteristic malformations observed in humans (Goerttler 1962; Pliess 1962; Seller 1962; Somers 1962; Mauss & Stumpe 1963). Only a few strains of rabbits and hamsters were responsive to the toxic effect, as were certain strains of monkeys (Seller 1962; Somers 1962; Delahunt & Lassen 1964). These data actually demonstrated that different species showed a wide variety of responses to thalidomide. Even more, different strains of the same species of animals were also found to have highly variable sensitivity to the drug.

The interpretations of these findings on inter- and intraspecies variety were quite different: A small group of physicians and pharmacologists responded with very critical comments on the validity of further toxicity tests in animal models. For example, Mary Seller, a clinician scientist at the Paediatric Research Unit at Guy's Hospital Medical School in London drew the following conclusion in view of the considerable interspecies variation regarding toxicity:

“May I suggest that although experiments on pregnant animals in the trials of new drugs are essential, negative results in this respect may prove to be no indication that the drugs are safe for human use. [...] It seems, therefore, that the only way to investigate this type of danger in a new drug would be to administer it to pregnant women. This is clearly impossible. Consequently, the most satisfactory method at the present time, and in the light of the thalidomide experience, of dealing with drugs with an unknown effect in the pregnant woman, would appear to be not to administer them, except if absolutely life-saving” (Seller 1962, 249).

Similarly, the London endocrinologist Raymond Greene, referring both to the argument of Mary Seller and to his own animal experiments with thalidomide, confirmed that “the most careful tests of a new drug's effects on animals may tell us little of its effects on humans.” He arrived at the conclusion that “animal experiments cannot obviate the risk and may even prevent the use of excellent substances. We

must accept some risk or – perhaps the wiser course – do without new drugs” (Greene 1962, 452). This interpretation was also shared by David Woollam, a specialist in experimental teratology at Cambridge University (Woollam 1962a, 237).

Such evaluations acknowledged the precarious character of the knowledge produced in the various animal models. They admitted the fact that in any case of future drug testing, it would not be possible to anticipate which specific animal species might be suitable for predicting potential hazards in humans, or if testing in animal models might at all yield the results aimed for. However, as mentioned above, such a critical position was absolutely marginal in the debate.

The majority of physicians and pharmacologists took the broad variety of outcomes of the toxicity studies in pregnant animals as an argument for the need to implement more comprehensive procedures of animal testing. For example, the editors of the prestigious medical journal *The Lancet* proclaimed in November 1962 that

“it can now be said that many of the thalidomide malformations [...] would have been prevented had the drug been tested in pregnant females of a sufficient number of species of laboratory animals” (N.N. 1962, 1095)

At the same time, they acknowledged that there was

“an urgent need for basic embryological research on a number of mammalian species before they can be accepted as satisfactory test-animals in the examination of the teratogenic effects of drugs” (ibid.).

Apparently, the validity of the knowledge itself which had been produced in animal experimentation was not questioned. The problem was rather the insufficient scope and extent of the actual testing practice (see also Woollam 1962b). In other words: Diverting the focus of attention to the practicalities, and in particular to the extent of animal testing, served to reconfirm the reliability of the animal model as such as a source of valid knowledge.

Like the majority of physicians, the conclusion reached by professional bodies and state regulators concerned with the thalidomide disaster, in particular in the United States, was that longer periods of data gathering and more comprehensive clinical testing were necessary, as well as the extension of toxicity studies in animal models, with the regular inclusion of pregnant animals (Daemmrich 2002). In Germany, a federal law regulating the admission of drugs to the market had, after long controversies, only been introduced in August 1961, immediately before the thalidomide disaster became obvious. The new law, however, only formulated very weak and in part vague rules – due to the pressure from the pharmaceutical industry and political parties representing the former’s interest (Kirk 1999, 20-34; Lenhard-Schramm 2017, 136-146). It only required that the pharmaceutical company register

the new compound with the responsible state agency, without any necessity of preceding animal tests or clinical trials on the efficacy, efficiency, or side effects of the drugs – a requirement which was essential at the time according to the regulations of the US Food and Drug Administration, and also according to the Dutch pharmaceuticals act (Stapel 1988, 247-251; Daemmrich 2004). Instead, in Germany, the producer of a new compound was expected merely to submit a report of experiences (*Erfahrungsbericht*) on problematic side effects, compiled by physicians employed by the company (Gesetz 1961, 538).

Thus, at the time when thalidomide was brought to the market in 1956/1957, the only requirement was to register the compound, without any specific conditions for prior testing in animals or systematic clinical trials. However, once the evidence for the severe side-effects following the administration of thalidomide had solidified by the end of 1961 and the drug had been taken off the market in November of that year, debates about the obvious limitations of the new law rapidly intensified, and the increasing knowledge about the case amongst physicians, relatives of the victims, and the public led to enormous pressure on state agencies to revise and clarify the law. After intense negotiations, a decisive revision to the law was implemented in June 1964. Now, it made the presentation of extensive written documentation on preclinical and clinical trials by pharmaceutical companies an obligatory requirement before applying for admission to the market (Kirk 1999, 156-182; Lenhard-Schramm 2017, 151-162). For the first time, animal tests were mentioned as part of the obligatory preclinical trials. Producers had to confirm that the compounds had been tested “sufficiently and carefully according to the state of available scientific knowledge”, but without further specification (Zweites Gesetz 1964, 366). Remarkably, however, in the debates surrounding the implications of the thalidomide-disaster and the potential consequences for stricter regulations, there was no discussion broaching the critical statement of Mary Seller, that is, no systematic evaluation of the intrinsic limits and hazards of the animal model preceding any clinical trials. Accordingly, the revised law did not specify any rules for animal experimentation prior to clinical trials. The rationality of the use of animal models to develop new compounds for application in humans remained uncontested.

The case of the antibody TGN1412

On 13 May 2006, six healthy young men in a London hospital received an intravenous injection of a new monoclonal antibody. It was a first-in-human- or phase 1 clinical trial, that is, the first application in humans of a new drug, after extensive tests in various animal models. The compound was designed as an immunological

treatment against autoimmune disorders such as rheumatoid arthritis or multiple sclerosis. Within minutes after the administration, all six research subjects experienced massive headaches, nausea, and fever, and after a few hours, the volunteers developed a life-threatening cardiovascular shock and multi-organ failure. They survived only after being transferred to an intensive care unit. Two of the volunteers were in a coma for one week, one of them even for three weeks. One of the volunteers suffers from lifelong damage to his fingers and toes as a result of necrosis. Long-term effects of the drug-administration are still under investigation today.²⁴

The clinical phenomena experienced by the research subjects were interpreted as the result of the systemic release of pro-inflammatory cytokines, termed a cytokine-release syndrome (CRS) (Suntharalingam et al. 2006), in effect an overreaction of the immune system. As one of the immunologists responsible for the development of the new drug later reported in an article in *Nature Reviews Immunology*, “a shock wave went through the scientific community, the biotechnology industry and the regulatory authorities, who [all] asked why preclinical testing [in animals] had failed to warn of the impending catastrophe” (Hünig 2012: 317). Other authors talked of the “TGN1412 disaster” (e.g. Eastwood et al. 2010, 513; Tranter et al. 2013, 164).

What had been the exact reasoning and the intended effect before TGN1412 was administered to humans, and what went wrong? Before the phase 1 clinical trial, first experiments in rats had shown that the new antibody had two modes of action: It *stimulated* the immune response through its impact on the CD 28 receptor on *effector memory* T lymphocytes; in addition, through a parallel pathway which also involves

²⁴ The clinical course of the events is described in detail in Suntharalingam et al. 2006; an explanation for the events six years after the failure, as seen by one of the immunologists who developed the compound, may be found in Hünig 2012. A first account from the point of view of the philosophy of science is given in Lemoine 2017: It is a critique of what Lemoine calls the conventional “transportation view” of extrapolating knowledge from the animal model to humans, in contrast to which he proposes an alternative “chimera view” (Lemoine 2017), following previous analyses by Huber & Keuck 2013, and Baetu 2014. However, in his reconstruction of the “transportation view”, Lemoine directly follows the accounts of the scientists involved (thus e.g. concluding that their approach is inductive) which he then juxtaposes with his own theoretical analysis (identifying “in fact” a deductive approach in the case), a procedure which may be seen as the result of an asymmetrical standard of judgment; he thereby ignores previous insights from the STS tradition which have broadly demonstrated the gamut of explicit and implicit “pre-empirical” presuppositions entering into empirical laboratory research. Lemoine also ignores that Thomas Hünig, the leading scientist involved in the TGN1412 case, himself acknowledged a kind of “chimera” procedure when he explained that the three preclinical “test systems” each addressed different aspects of the mechanisms triggered by TGN1412, which only if seen together may yield a comprehensive view of the relevant combined dynamics initiated by the compound (Pallarey & Hünig 2010, 510; Hünig 2012, 318).

the CD 28 receptor, it *moderated* the immune response by triggering a *regulatory* type of T cell (Tacke et al. 1997; Lin & Hünig 2003). Thus, after these initial animal experiments, it was clear that the CD 28 receptor was part of a quite complex system of stimulation and moderation to maintain a balanced regulation of the immune system. In fact, it was also known that this system is embedded in a much broader, far more complex apparatus of immune-regulation which was not fully understood even many years after the disaster (the unresolved problem of complexity was acknowledged *post factum* e.g. in Pallardy & Hünig 2010, 511).

Further tests in animals had shown that the antibody's second property – the selective downregulation of the immune response without a general suppression of the immune system – was particularly promising. This mechanism led Thomas Hünig, the leading scientist involved, and his colleagues to expect that the newly discovered antibody might be used to treat various autoimmune disorders (Beyersdorf et al. 2006), that is, disorders which are characterized by an inadequate overreaction of the immune-system against cells of the own organism. Trials with animal models of rheumatoid arthritis and multiple sclerosis in mice and rats had confirmed this hypothesis: The rodents recovered from their conditions and did not show any side effects (Beyersdorf et al. 2005). In a further step, experiments in macaque monkeys had shown similar therapeutic effects, and again, no serious adverse reactions, such as an anaphylactic shock or immunosuppression (TeGenero 2005; replicated after the disaster in Duff 2006). These experiments on monkeys were of particular relevance, because the knowledge available suggested that the properties of the binding site and the affinity for the TGN1412 antibody were almost identical in humans and this primate species, whereas there were significant differences between humans and rodents regarding these parameters (Hanke 2006).

The requirements formulated in the then relevant guidelines from regulatory agencies (following the European Clinical Trials Directive 2001/20/EC) recommended carrying out toxicology and safety pharmacology analyses in two “relevant animal species” (one rodent and one non-rodent) to identify the target organs – a requirement fulfilled by the preclinical testing of TGN1412: Since biological agents and especially therapeutic monoclonal antibodies generally exhibit exclusive species specificity for the target antigen, the most difficult task for non-clinical safety studies is to find a “relevant species”. In the preceding years of therapeutic monoclonal antibody development, a relevant animal species had usually been defined as a species showing antibody binding to the animal homologous target and also comparable pharmacological effects. In the case of TGN1412, the cynomolgus (macaque) monkey was considered a potential candidate for a relevant animal model on the basis of the data on the sequence homology of CD28 as target structure. In

addition, also the data on binding affinity and pharmacodynamic downstream effects in the nonhuman primates were taken into account (Schneider et al. 2006, 494). After an official review of the data from the animal experiments, the responsible state regulatory bodies in both Germany and Britain agreed to phase 1 trials. It was only the sluggishness of the German regulatory bureaucracy which led to the decision to choose London for the study.

What were the explanations for the failure of the new compound in the phase 1 trial? The core problem was that the cytokine storm, that is, the massive immune stimulation, had not been expected after the tests in the animal models – because there, no such stimulation had occurred. In a first reaction only two weeks after the failure, the editors of *The Lancet* formulated what they perceived as the range of potential causes: “[...] it is unclear whether there was a fault with the quality of the drug, contamination, a deviation from the protocol, or whether this was an unpredicted adverse event“ (Editorial 2006, 960). The possibility that the reliance on positive testing in the animal model as the last and decisive step before the application of the compound in humans might be the core issue – the argument brought forward for example by Sellar and Greene in the same journal in the context of the thalidomide case – was not addressed as such.

However, a few months later, Marcel Kenter and Adam Cohen, representatives of Dutch institutions involved in the planning and regulation of first-in-human trials, articulated a somewhat more critical point of view: They argued that until the 1980s, “small molecules with fairly well characterised, classic, pharmacological mechanisms“ had been tested in drug trials, and pre-clinical testing in animal models had been adequate for this kind of substances. In contrast, since the 1990s, the advent of new, „increasingly potent and selective compounds for human-receptor systems led to situations in which predictability from animal data was diminishing“ (Kenter & Cohen, 2006, 1387). Similarly, but with focus on the recent disaster, representatives of the German regulatory authority Paul Ehrlich Institute diagnosed that „the adverse effects associated with the TGN1412 phase 1 trial indicate that the predictive value of animal models requires re-evaluation“. However, beyond this initial and short critical comment on the validity of animal models in general, they continued, „that, in certain cases, standard clinical protocols [defining the preconditions of clinical trials] may need refinement or redesign“ (Schneider et.al. 2006, 493). In their further deliberations, they spelled out such „refinements or redesign“, that is, modifications for the practicalities related to testing in animal models, without returning to the more basic issues related to this methodological tool.

In view of the experience with TGN1412, Richard Horton, the editor-in-chief of *The Lancet*, invited authors to submit reports specifically of phase 1 trials to the

journal, if they met two of the following three criteria: Did the trial test a novel substance for a novel indication in a specific disease; was there a strong or unexpected beneficial or adverse response to the medicine; and did the study throw light on a novel mechanism of action (Horton 2006)? This call for papers was a corollary of the apparently new awareness that until this date, comparatively few reports about phase 1 trials had been published (in contrast to phase 2 + 3 trials immediately preceding market introduction), with the implication that only scarce public information was available about the concrete experiences and hazards involved in the transfer of knowledge from animal models to humans (see e.g. Decullier et al. 2009). Indeed, as another author had observed in an article in *Science* published a few weeks before Horton's appeal, phase 1 studies were routinely rejected by high-impact-factor medical journals (Kaiser 2006, 1853). Significantly, however, in the years following Horton's appeal, the number of publications on phase 1 trials increased only modestly, and the cases reported almost completely avoided the reconstruction and analysis of „unexpected“ and „adverse“ responses which Horton had listed to be of particular interest (Horton 2006, 827).

Back to the case of TGN1412: In the years after the disaster, it could be clarified that the missing effect of massive immune-stimulation in the animal models, and thus the absence of severe adverse reactions, had different reasons in the rodents, and in the monkeys. In humans, the casualties had been shown to follow an accumulation of stimulating T cells, which is driven by multiple exposures to infectious agents. But it turned out that this accumulation does not occur in rodents housed under clean laboratory conditions (Hünig 2012, 317). As a consequence, the rodents had a relatively low proportion of stimulating T cells in comparison with the number of moderating regulatory T cells, leading to the overall result of a strongly moderating impact in the experiments with mice and rats.

Thus, it may be concluded that in the case of the rodents, the artificial conditions of the laboratory in which they were grown led to the production of a kind of knowledge which was both pre-structured and limited by these conditions. This knowledge from the animal model deviated in a decisive way from the requirements of strong, valid knowledge for a treatment's application in humans – but that aspect became clear only in hindsight.

And what about the monkeys? Given the supposedly known identity of the antibody binding site and of the binding affinity to the stimulating effector T cells in humans and macaques, the mechanism of the intervention appeared to be clear, and was part of the unquestioned assumptions for the downregulation of the immune response observed in the monkeys, and the intended down-regulation in humans (as

documented in the brochure by the responsible biotech company: TeGenero 2005). Only several years after the disaster, after a number of other potential explanations had been ruled out, was a surprisingly simple explanation for the difference between monkeys and humans discovered: The knowledge about the CD 28 molecule on effector T cells of the macaques referred to precursor T cells, that is, to an early developmental stage of this type of T cells (Eastwood et al. 2010; Pallardy & Hünig 2010). Contrary to the previous assumptions, it turned out that these precursor-T cells lose the decisive CD 28-binding site for the antibody during their development to mature effector T cells – in contrast to the equivalent human precursor T cells, which do not lose the binding site. Again, this crucial difference was not only unknown before the phase 1 trial, more importantly, there was no awareness at all that there might be a relevant problem on this level.

Thus, with regard to the experiments in the macaques, the assumed high validity of the knowledge produced in the animal model relied on the unquestioned presupposition that humans and primates shared a basic immunological structure for the biological mechanism targeted – a presupposition which in retrospect turned out to be not valid. As Hünig formulated a number of years after the disaster, this non-identity of a central target structure was „a fact not predictable for investigators and regulators alike, during preclinical development at the time [...] using a wrong model for hazard identification will always yield wrong results“ (Pallardy & Hünig 2010, 511). A group of pharmacologists at the UK National Institute of Biological Standards and Control, a government regulatory agency, even talked more generally of „the poor predictive value of standard preclinical safety tests and animal models“ (Stebbing et al 2013, 75). Thus, in view of mounting evidence for previously unknown differences in species in terms of basic immunological mechanisms, one might assume that the dramatic case of TGN1412 should have led to the insight that the knowledge gained in animal models is intrinsically precarious, and that therefore, the animal model should be fundamentally questioned as a core method to establish the safety and efficacy of drugs.

However, on the level of the responsible regulatory authority, there was apparently also no recognition of such basic implications. In July, 2007, as an immediate consequence of the TGN1412 disaster (that is, before the later clarification of the likely causes described above), the Committee for Medicinal Products for Human Use of the European Medicines Agency (EMA) enacted a new guideline for “first-in human” trials of new drugs: It required, amongst other points, more information on the mode of action of the new compound which was to be tested in humans, and on the structure of the molecular target. In addition, variations in the expression of the target structure in the general population were to be clarified before

first-in-human use, and a “demonstration of relevance” was required of the specific animal models used to produce preclinical data on the pharmacological and toxicological properties of the new drug. Finally, a strictly sequential procedure was prescribed for the concrete application, rather than the previous practice of simultaneous administration to small cohorts (EMA 2007). After the likely causes of the TGN1412 disaster had been clarified by 2012, a revision of these guidelines was apparently not deemed to be necessary. Instead, a revision regarding practicalities of the administration in humans was merely initiated after the next disaster had occurred in early 2016: In the context of the phase 1 trial of the compound BIA 10-2474, one participant died and four were left with long-term neurological symptoms.²⁵

Again, as in the previous examples of tuberculin and thalidomide, the failure of the phase 1 trial of TGN1412 pointed to some fundamental inherent weaknesses of the knowledge produced in animal models in general, such as inter-species differences, or only partly known intra-species developmental processes that imply significant changes in the expression and activity of basic immunological structures. In addition, this case made clear that the artificial environments in which the laboratory animals were grown, and the related lack of exposure to, for example, infections may well be a further critical point influencing the development of the immune system with its inherent memory function, and a potential cause for different immunological effects in laboratory animals as opposed to “free range” humans. As in the earlier cases of tuberculin and thalidomide, these various weaknesses of the knowledge produced in animals became visible only in the aftermath of the failure.

One would have expected that the fundamental issues raised by the failure, e.g. about the precarious similarity between the animal models chosen for preclinical testing and the human organism, or the inherently incomplete knowledge about complex and essential immunological mechanisms, were on the table and ready for critical analysis. However, in their reactions to the failure, not only the scientists involved, but also the regulatory authorities focused their attention and conclusions not on the animal model as such, but on technical details of the case-specific forms of animal models, and beyond that, on the practicalities of the application in human volunteers.

The origins of the idea of an animal model

²⁵ Before the application in humans, BIA 10-2474 had been tested in mice, dogs, monkeys and rats (Butler & Callaway 2016; Randerson 2016). The revised version of the EMA-guidelines was published in 2017. It formulated modified rules on sentinel dosing, definition of maximum exposure, and decisions on dose escalation steps (EMA 2017; see also Breithaupt-Grögler et al. 2019).

The examples of failures described above²⁶ indicate that there appears to be almost a blind spot, or inability to acknowledge and discuss potential intrinsic limitations of the animal model. Significantly, however, such limitations and the related issues were definitively seen at the very outset of the animal model's career. This will be the focus of the following section.

Conventionally, the origins of the animal model in the broader sense (as defined above) are localized in the second half of the 19th century (e.g. Canguilhem 1977; Bynum 1990). In this context the bacteriologist Robert Koch is usually regarded as the one who, in the years around 1880, established the use of animals in the laboratory as the most efficient way to obtain knowledge about human diseases (Gradmann 2005b, 78, 84). Yet Koch was by no means the first to employ experimentation in animals to clarify pathological processes in humans. Some medical researchers interested in human pathology, such as Francois Magendie, Ludwig Traube, and Rudolf Virchow, had used systematic experiments in animals already in the first half of the 19th century, however, with the primary aim to study the effects of specific interventions on healthy organisms, such as cutting the vagal nerve, or administering chemicals, followed by an evaluation of the ensuing effects in post mortem investigations of the animals (see e.g. Schmiedebach 1993). Others, such as Claude Bernard, used similar procedures with the aim to elucidate physiological processes, by drawing inferences from the resulting losses of function to the proper physiology of living organisms. However, all of them did not make any systematic attempt to recreate, or model human diseases in living animals to then study disease mechanisms and potential forms of therapeutic interventions.²⁷

On a theoretical and programmatic level, the Göttingen anatomist and physiologist Jakob Henle argued for the need to introduce systematic experimental work into research on human diseases in the years around 1850. In this context, he also postulated that diseases in humans were “experiments of nature” which should be systematically investigated by the physiologist interested in human pathology – a further step approaching the idea of human diseases as objects of experimental inquiry. In the correspondence with Henle in the mid-1850s, the physiologist Carl Ludwig developed the idea of an experimental pathology which would attempt to artificially produce disease conditions (*Krankheitsbilder*) in the laboratory which

²⁶ Further examples could easily be added, such as e.g. the BIA 10-2474 incident in France in January 2016 mentioned above (see Butler & Callaway 2016, Randerson 2016).

²⁷ One exception might be Bernard's description of his production of „artificial diabetes“ in animals which, however, again occurred in the context of his interest in the „normal“ metabolism of glucose and glycogen (e.g. Bernard 1865, 304-305; see also Grmek 1965); in contrast to the case of Cohnheim described below, Bernard was apparently not at all interested in the general issues around the possibilities and limitations of such „artificial“ diseases in animals.

resemble those in humans, but without turning this program into practice (see Roelcke 2013, 26-29).

Indeed, it was one of Koch's mentors, the pathologist Julius Cohnheim of Breslau, who in the 1860s started practical laboratory work to reproduce human disease processes in animals for further experimental research in such artificially created conditions. Cohnheim's methodology had considerable influence on Koch's early bacteriology. Koch visited Cohnheim's laboratory in Breslau first in 1876, and once again in 1877, in order to demonstrate his own findings on anthrax, and to study the methodology of animal experimentation (for these visits, see Gradmann 2005b, 78). In the following section I shall examine more closely Cohnheim's considerations about the possibilities and limits of knowledge production by means of the animal model.²⁸

The point of departure for Cohnheim's ideas was the broader interest among mid-19th century pathologists in inflammation as a fundamental phenomenon in many disease processes. In this context, Cohnheim decided around 1860 to dedicate his dissertation to the topic of inflammation (Maulitz 1978). A central conclusion of this work was that some fundamental questions in the aetiology of inflammatory processes could well be approached by substituting animals for human subjects (Cohnheim 1914 [1867]).

In the following years, Cohnheim and his colleague Bernhard Fränkel developed a new technique to transfer a specific human disease, in this case tuberculosis, to animals. Initially, Cohnheim's interest here was not the study of the disease itself. Rather, he pursued the question whether by transmitting tissues or substances from one body to another "it is at all possible to generate a condition in guinea pigs that corresponds to tuberculosis in human beings"? (Cohnheim & Fränkel 1868, 265).

It is of particular interest to note that Cohnheim explained his choice of experimental animal with two quite pragmatic considerations: First, he asserted that guinea pigs, in contrast to other animal species such as rabbits, are much more rarely affected by undesired concomitant parasitic diseases; and second, in the guinea pig, the area on the animal's abdomen where the inoculation was to take place was easily accessible, which would allow the consequences of inoculation to be observed without any problems (ibid., 264-265).

Several years later, in 1882, Cohnheim elaborated on the issue of producing human disease in animals in his *Lectures on General Pathology*. Here, he defined disease as a deviation from the regular process of life through changes in the normal conditions of life. For him, it followed that the causes of diseases were changed

²⁸ The following is described in more detail in Roelcke 2017: 70-77.

conditions of life, and therefore factors located outside of the organism. He clearly acknowledged that an adequate investigation of the plethora of factors affecting human beings from the outside would theoretically require including a broad spectrum of sciences. Relevant as reference disciplines that deal with the external living conditions of humans, he continued, were natural science disciplines such as chemistry, botany, zoology; but in addition, Cohnheim stressed, the social sciences were needed as well (Cohnheim 1882: 10).

Cohnheim found the broad spectrum of potential causes of diseases so extensive and complex, however, that he chose to restrict his own interest to 'physiological pathology', a subspecialty of pathology which focused not on the first causes of diseases (that is, aetiology), but on the "functions of organs under pathological conditions", triggered by external causes (ibid., 11-12). Thus, it was the result of a conscious and pragmatic decision taken by Cohnheim to reduce his scientific attention entirely to the body interior and to the internal dynamics of disease formation, because only these processes were accessible to the laboratory sciences. At the same time, he was very aware that by doing so, he excluded many relevant factors from his attention in order to arrive at a simplified and easily to utilize model of disease mechanisms (ibid., 12).

Thus, for Cohnheim, the animal model enabled pathologists to produce a kind of knowledge with a validity similar to that produced in the natural sciences by making use of experiments, in contrast to the previously dominant methods, in particular systematic observation and deduction. Yet, as mentioned before, he was very conscious of the fact that the methodological scope of the natural sciences was not sufficient to elucidate all the relevant dimensions of human disease. There would always be aspects of disease which would be inaccessible to experimentation in the animal model, namely

"all of those processes which are peculiar to human beings and neither occur spontaneously in animals nor can be artificially produced – and there are quite a lot of these, even apart from mental processes, many of them quite important" (ibid., 14-15).

With this and his previous statements, Cohnheim expressly addressed four limitations to the animal model which might be of general relevance:

1. Human disease produced in the animal is of an artificial character. This artificial character creates features of the constructed disease in the animal that deserve special attention. In addition, this implies that the existing understanding of a specific disease condition prior to the modelling process will be the framework for the construction of the target in question, and thus pre-structure it.

2. The knowledge of disease processes produced in animals is selective, not comprehensive: In particular, it does not cover the “first causes” of diseases, that is, their aetiology, but only the “dynamics of disease formation” within the body (their pathogenesis), and this is a consequence of a decision taken by the researcher.
3. Of necessity, the choice of animals is driven not only by the “nature” of the condition which is to be modelled, but also by pragmatic reasons, that is, the feasibility of the actual work with the chosen animal. With this, Cohnheim brought up the question as to whether focusing on the utility of an animal in the context of experimental practice might, intentionally or not, have an impact on the researcher’s attention to the targeted pathological processes. In effect, the feasibility of experimental work might indeed override considerations of what might be termed “similarity” between the human disease, and the artificially induced condition in the animal.
4. A range of important disease processes “which are peculiar to human beings” are not accessible by the animal model. This, at first glance, sounds rather vague, but taken together with Cohnheim’s reference to the importance of the social sciences, it makes clear that he had some awareness of the impact of social relations as well as social institutions on the causation and course of diseases in general, even those conventionally called somatic diseases.

Conclusions

The reconstruction of Cohnheim’s path to the animal model, and his reflections on its use and limitations, enable a sharpened and potentially more critical view of this basic methodological tool of medical research. Such a sharpened view may be used to look at the broader history of animal models of human disease, and the various evaluations of their validity which have been employed in specific historical contexts.

As I have shown, such a history arguably might start with Cohnheim’s work and acknowledgment of the artificial nature of the animal model, of the conscious selection, that is, reduction of features of this artificial disease, and of the precarious validity of the (selective) knowledge produced by utilizing it. Further, Cohnheim was very aware of the pragmatic reasons for the choice of animals used, that is, a basic contingency inherent in this artificial knowledge, as well as various limitations resulting e. g. from significant inter-species variation in basic biological functions. Finally, he accepted that certain important dimensions of human diseases, in particular the socio-psychological dimension, are generally not accessible by the methods of the natural sciences as applied in the animal model. With the limitations

spelled out by Cohnheim in mind, we may realize that the explanations for the various failures described above, and the ensuing conclusions formulated by the scientists involved, their colleagues, and regulatory agencies addressed only in part the range of aspects which might contribute to the precarious validity of any knowledge derived from animal models. Remarkably, in particular the social dimension intrinsic to the development, symptomatology, and further course of diseases, which Cohnheim clearly acknowledged, was completely absent in any of the deliberations of later medical researchers and regulators, although e.g. the methodological standards of placebo-controlled and double-blind studies in clinical trials imply that the relevance of the underlying psychosocial processes is not contested in the practice of phase 3 clinical trials.

The generation after Cohnheim, headed by Robert Koch, crucially re-evaluated the animal model, amounting to what might be called an epistemological upgrade. After Koch had succeeded in using the animal model to identify the pathogens of wound infection, cholera, and tuberculosis, for which he had found enormous resonance and acclaim, the epistemological status of the animal model was elevated: the previously helpful and important methodological tool for specific, but limited questions now became the obligatory and de facto unique method for the production of strong knowledge on human diseases.

Following this epistemological upgrade, and concomitant with the privileged status of experimental methods in medicine starting around the end of the 19th century, the concept of the animal model moved centre stage in medical research and became a constitutive element of the production of valid knowledge on human diseases. In a way, one might speak of a culture of biomedical research in human disease which took the value of this methodological tool as self-evident, and more or less beyond question for more than a century. This configuration of epistemological privilege was supported by ethical considerations which argued that potentially hazardous experimental interventions in humans should, if at all possible, be substituted by analogous interventions in animals.

Although regular and in part spectacular failures occurred in the transition from the animal model to phase 1 trials in humans, these events led to little more than temporary irritation about this process of knowledge transfer. Even when individual scientists brought up the question as to the precarious validity of the animal model, and the hazards associated with the knowledge derived from it, the overwhelming majority of the biomedical scientists involved, as well as the responsible state or transnational regulatory authorities, reacted on a completely different level: They focused their attention on issues related to the selection of suitable animals, problems with the practical execution of the experiments in animals, or on

procedures related to the first administration in humans, such as proper recruitment of research subjects, or the sequential application of new compounds. By focusing the attention on these aspects, the limitations and the ensuing weaknesses of the animal model as such were more or less bracketed out of any systematic scrutiny, and instead, as an indirect result, the supposedly strong validity of the knowledge derived from the animal model was reconfirmed.

Similarly, the high attrition rate of new therapeutic compounds entering phase 1 trials has not really led medical researchers to question the utility of the animal model. As mentioned previously, a number of recent empirical studies have shown that over a period of at least ten years in the early 2000s, the knowledge produced in the animal model was unsuitable for application in humans in approximately nine out of ten cases (e.g. Hackam & Redelmeier 2006; van der Worp et al. 2010; N.N. [editorial] 2012; Cummings et al. 2014; Hay et al. 2014; Perrin 2014). In spite of such data, the authors of one major study published in 2014 who had found such remarkable attrition rates in their conclusion urged for “more predictive animal models” to increase success rates. Further strategies they recommended were “improvements in communication between sponsors and regulators”, “greater flexibility with surrogate endpoints [of clinical trials]”, and “improved technologies for assessing patient risk-to-benefit” (Hay et al. 2014, 50) – but the more basic idea to focus on the validity of the knowledge actually obtained from the animal model was conspicuously absent.

If corroboration in empirical practice is taken as the core criterion for the “strong” character of any knowledge, the repeated failures related to the transfer of knowledge from the animal model to human research subjects, not to mention the high attrition rates of therapeutic compounds entering phase 1 trials cast doubt on evaluations which do not address the validity of knowledge from animal models as such. In view of the empirical evidence delineated here, the knowledge produced in animal models of human diseases appears to be of a rather precarious, or weak character. The lack of attempts of the scientists concerned, as well as of state regulatory authorities, to look into the causes of such failures in terms of the validity of the animal model is an interesting phenomenon. Indeed, the almost indestructible belief in the epistemic value of the animal model and the fact that its use persisted despite the obvious failures needs some explanatory reflection. One possible explanation might assume that the scientists concerned had a realistic appreciation of the empirical data on high attrition rates and sometimes disastrous failures. In this case, the research community would openly accept the poor outcome of animal model-based drug research as inherent to the methodological approach itself, but would continue its practice in spite of this awareness, since the success rate of ca.

10% justified the price. However, almost all explanations and justifications presented by the historical actors in the cases described here illustrate that such an awareness and, based on this, an open risk-benefit calculation may not be found.

An alternative explanation might argue that the historical actors attributed a high priority to „risk substitution“, namely the possibility to follow up research hypotheses about potential pathomechanisms and therapeutic interventions in animals instead of humans, and thus to reduce the risks of severe side effects in humans. In this view, the ethical dimension of using animal models to study human diseases may either have made the poor success rates more acceptable, or even have blurred systematic and critical analyses of the epistemological shortcomings and costs (which, in fact, in themselves include human suffering). However, the historical evidence presented here indicates that such ethical considerations, if formulated at all, were completely marginal in the deliberations of the researchers concerned with the failures. But this does, of course, not exclude the possibility that by looking at more cases, and potentially making use of further sources (such as private letters or diaries), a different picture would emerge about the historical actors' systematic weighing up of the epistemological and ethical dimension of their practices.

A third potential explanation might refer to Thomas Kuhn's concept of scientific paradigms: The persisting belief in the epistemic value of the animal model might then be interpreted as a reasoning which accepts the presuppositions and explicit as well as implicit assumptions of a shared set of general rules and practices, given by a specific and successful example of actual scientific practice, which in turn define how to approach, to define, and to solve open questions within a research community. Within this set of shared rules, practices, and assumptions which constitute the ramification for a period of „normal science“, there is no perceived need to question fundamental assumptions (Kuhn 1970, 10-11). This explanation and interpretation appears to be most in tune with the cases described here: Starting with Koch, the historical actors simply *assumed* and accepted the strong validity of the knowledge obtained from the animal model, without providing adequate justification for such an assumption, even in the face of massive failures.²⁹ The basic presuppositions underlying the methodological approach, such as the artificial and selective “nature” of the disease modelled in the animal, the inherent exclusion of empirically known determinants for the causation and course of human diseases (such as psycho-social factors), and the impact of contingent interferences related to the practicalities of

²⁹ „Justification“ was a core feature of knowledge, as opposed to belief, in Plato's *Theaetetos*, and the starting point of a discourse on the gradation of forms of knowing, as lined out in Epple 2019.

dealing with animals in the laboratory, were – if at all – only sporadically and marginally mentioned in the cases analysed here, but not addressed in an obligatory and systematic fashion. The historical key to understanding this phenomenon may be related to the “epistemological upgrade” of the animal model, which, in the narrative outlined here, is historically situated between the deliberations of Cohnheim starting in the 1860s and the practices, assumptions, and justifications of Koch and his followers such as Ehrlich and Behring in the 1890s. More specifically, Koch’s success in the early 1880s in proving that tuberculosis in humans was linked to a bacterial agent by means of an animal model may be seen as establishing this paradigm.³⁰ Further historical research is needed to elucidate the specific mechanisms and dynamics in the late 19th century which created the kind of unquestionable epistemological authority that was attributed to the animal model for more than a century.

References

- Ankeny, Rachel A. 2010. „Historiographic Reflections on Model Organisms: Or How the Mureaucracy may be Limiting our Understanding of Contemporary Genetics and Genomics.“ *History and Philosophy of the Life Sciences* 32: 91-104.
- Baetu, Tudor M. 2016. „The ‚Big Picture‘: The Problem of Extrapolation in Basic Research.“ *British Journal for the Philosophy of Science* 67: 941-967.
- Behring, E[mil]. 1890. „Untersuchungen ueber das Zustandekommen der Diphtherie-Immunität bei Thieren.“ *Deutsche Medicinische Wochenschrift* 50:1145–1148.
- Behring [, Emil] and [Shibasaburo] Kitasato. 1890. „Über das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren.“ *Deutsche Medicinische Wochenschrift* 49: 1113-1114.
- Bernard, Claude. 1865. *Introduction à l’étude de la médecine expérimentale*. Paris: J. B. Baillière.
- Beyersdorf, Niklas, Stefanie Gaupp, Karen Balbach et al. 2005. “Selective targeting of regulatory T cells with CD28 superagonists allows effective therapy of experimental autoimmune encephalomyelitis.” *Journal of Experimental Medicine* 202: 445-455.
- Beyersdorf, Niklas, Thomas Hanke, Thomas Kerkau and Thomas Hünig. 2006. “CD28 superagonists put a break on autoimmunity by preferentially activating CD4+CD25+ regulatory T cells.” *Autoimmunity Reviews* 5: 40-45.

³⁰ For a reconstruction of Koch’s work on the bacteriological origins of tuberculosis, see Gradmann 2001.

Breithaupt-Grögler, Kerstin, Tim Hardman, Jan de Hoon et al. 2019. "The New First-in-Human EMA Guideline: Disruptive or Constructive?" *Frontiers in Pharmacology* 10: 398.

Brock, Thomas D. 1988. *Robert Koch: A Life in Medicine and Bacteriology*. Madison/Wisconsin: Science Tech Publishers.

Butler, Declan and Ewen Callaway. 2016. "Scientists in the dark after French clinical trial proves fatal." *Nature* 529: 263-264.

Bynum, William. 1990. „'C'est un malade': Animal Models and Concepts of Human Diseases". *Journal of the History of Medicine and Allied Sciences* 45: 397-413.

Canguilhem, Georges. 1965. *La connaissance de la vie*, Deuxième édition revue et augmentée. Paris: J. Vrin.

Canguilhem, Georges. 1977. „L'effet de la bactériologie dans la fin des 'théories médicales' au XIXe siècle." In *Idéologie et Rationalité dans l'Histoire des Sciences de la Vie*, by Georges Canguilhem. Paris: Vrin.

Cohnheim, Julius. 1914 [1867]. *Ueber Entzündung und Eiterung* Leipzig: J. A. Barth (reprinted in *Sudhoffs Klassiker der Medizin*, vol. 23).

Cohnheim, Julius. 1882. *Vorlesungen über allgemeine Pathologie*, 2nd edition, vol. 1. Berlin: Hirschwald.

Cohnheim, J[ulius], and B[ernhard] Fränkel. 1885 [1868]. „Experimentelle Untersuchungen über die Uebertragbarkeit der Tuberkulose auf Thiere." In Julius Cohnheim, *Gesammelte Abhandlungen*, 264-278. Berlin: Hirschwald.

Creager, Angela N. H. 2002. *The Life of a Virus: Tobacco Mosaic Virus as an Experimental Model, 1930–1965*. Chicago: University of Chicago Press.

Cummings, Jeffrey L., Travis Morstorf, Kate Zhong. 2014. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimer's Research & Therapy* 6: 37.

Cunningham, Andrew and Perry Williams (eds.). 1992. *The Laboratory Revolution in Medicine*. Cambridge: Cambridge University Press.

Daemmrich, Arthur. 2002. „A tale of two experts: Thalidomide and political engagement in the United States and West Germany." *Social History of Medicine*, 15: 137-158.

Daemmrich, Arthur. 2004. *Pharmacopolitics: Drug Regulation in the United States and Germany*. Chapel Hill: University of North Carolina Press.

Davidson, M. K., J. R. Lindsey, and J. K. Davis, 1987. „Requirements and selection of an animal model." *Israel Journal of Medical Sciences* 23: 551–555.

Decullier, Evelyne, An-Wen Chan, Francois Chapuis. 2009. "Inadequate dissemination of phase I trials: a retrospective cohort study." *PLoS Medicine* 6: 202-209.

Delahunt, C. S., and L. S. Lassen. 1964. „Thalidomide syndrome in monkeys.“ *Science* 146: 1300-1305.

Dijkhuis, H. J., J. H. Hagenbeek, B. V. Bekker et al. 1962. "Thalidomide and congenital abnormalities." *The Lancet* (September 1): 452.

Dombrowski, Frank, P[eter] Bannasch, and U[rich] Pfeifer. 1997. "Hepatocellular neoplasms induced by low-number pancreatic islet transplants in streptozotocin diabetic rats." *American Journal of Pathology* 150: 1071-1087.

Duff, G. W. 2006. *Expert Scientific Group on Phase One Clinical Trials Final Report*. Stationery Office: Norwich, UK.

[Editorial]. 2006. "SuperMAB on trial." *The Lancet* 367: 960.

Ehrlich, P[aul]. 1891a [1957]. „Über neuere Erfahrungen in der Behandlung der Tuberkulose nach Koch, insbesondere der Lungenschwindsucht.“ In Paul Ehrlich, *Gesammelte Arbeiten/Collected Papers, 2. Band: Immunitätslehre und Krebsforschung*, edited by F[red] Himmelweit. 13-20. Berlin: Springer/London: Pergamon Press.

Ehrlich, P[aul]. 1891b. Recent Experiences in the Treatment of Tuberculosis (With Special Reference to Pulmonary Consumption) by Koch's Method. *The Lancet* October 24: 917-920.

Ehrlich, Paul. 1897 [1957]. „Die Wertbemessung des Diphtherieheilserums und deren theoretische Grundlagen.“ In Paul Ehrlich, *Gesammelte Arbeiten/Collected Papers, 2. Band: Immunitätslehre und Krebsforschung*, edited by F[red] Himmelweit. 86-106, English transl. 107-125. Berlin: Springer/London: Pergamon Press.

Elkeles, Barbara. 1990. „Der Tuberkulinrausch von 1890.“ *Deutsche Medizinische Wochenschrift* 115: 1729-1732.

EMA (= European Medicines Agency). 2007. „Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products (Doc. Ref. EMEA/CHMP/SW/28367/07).“ 31 October 2020.
< https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-strategies-identify-mitigate-risks-first-human-clinical-trials-investigational-medicinal_en.pdf >

EMA (= European Medicines Agency). 2017. „Guideline on strategies to identify and mitigate risks for first-in human clinical trials with investigational medicinal products. Revision 1 (Doc. Ref. EMEA/CHMP/SWP/28367/07 Rev. 1).“ 31. October 2020.
< https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-strategies-identify-mitigate-risks-first-human-early-clinical-trials-investigational_en.pdf >

Epple, Moritz. 2019. „The *Theaetetus* Problem: Some Remarks Concerning a History of Weak Knowledge”. In *Weak Knowledge: Forms, Functions, and Dynamics*, edited by Moritz Epple, Annette Imhausen, and Falk Müller, 19-40. Frankfurt am Main/New York: Campus.

Esser, H. and Heinzler, F. 1956. „Klinische Erfahrungen mit einem Sedativum und Hypnotikum.” *Therapie der Gegenwart* 95: 374-376.

Freeman, Matthew, and Daniel St Johnston. 2008. “Wherefore DMM [Disease Models and Mechanisms]”. *Disease Models and Mechanisms* 1, 6-7.

Frenkel, J. K. 1969. „Choice of animal models for the study of disease processes in man: Introduction.” *Federation Proceedings* 28: 160-161.

Gad, Shayne C. 2007. “Current animal studies.” In *Animal models in Toxicology*, edited by Shayne C. Gad. 2-18. Boca Raton etc.: Taylor & Francis.

Geison, Gerald L. 1995. *The Private Science of Louis Pasteur*. Princeton, NJ: Princeton University Press.

Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz). 1961. *Bundesgesetzblatt* no. 33 (19. Mai 1961): 533-546.

Goerttler, Klaus. 1962. „Der ‚teratologische Grundversuch‘ am bebrüteten Hühnchenkeim, seine Möglichkeiten und Grenzen.” *Klinische Wochenschrift* 40: 809-812.

Gradmann, Christoph. 2001. „Robert Koch and the Pressures of Scientific Research: Tuberculosis and Tuberculin.” *Medical History* 45: 1-32.

Gradmann, Christoph. 2005a. *Krankheit im Labor: Robert Koch und die medizinische Bakteriologie*. Göttingen: Wallstein.

Gradmann, Christoph. 2005b. „Das Maß der Krankheit: Das pathologische Tierexperiment in der medizinischen Bakteriologie Robert Kochs.” In *Maß und Eigensinn: Studien im Anschluss an Georges Canguilhem*, edited by Cornelius Borck, Volker Hess, and Henning Schmidgen, 71-90. München: Fink.

Gradmann, Christoph. 2008. „Locating Therapeutic Vaccines in Nineteenth-Century History.” *Science in Context* 21: 145-160.

Greene, Raymond. 1962. „Thalidomide and congenital abnormalities.” *The Lancet* II (September 1): 452.

Grmek, Mirko 1965/66. Examen critique de la genèse d’une grande découverte: La „pique diabétique“ de Claude Bernard. *Clio Medica* 1: 341-350.

Großbölting, Thomas and Niklas Lenhard-Schramm (eds.). 2017. *Contergan – Hintergründe und Folgen eines Arzneimittel-Skandals*. Göttingen: Vandenhoeck & Ruprecht.

Guttman, P[aul], and P[aul] Ehrlich. 1891. "Die Wirksamkeit kleiner Tuberkulindosen gegen Lungenschwindsucht." In Paul Ehrlich, *Gesammelte Arbeiten, 2. Band: Immunitätslehre und Krebsforschung*, edited by F[red] Himmelweit. 7-12. Berlin: Springer.

Hackam DG, Redelmeier DA. 2006. Translation of research evidence from animals to humans. *Journal of the American Medical Association* 296: 1731–1732.

Hackel, Donald B. 1980. "Introduction to the Workshop." *American Journal of Pathology* 101, iss. 3S (= Suppl. to December): S1-S2.

Hanke T[homas]. 2006). „Lessons from TGN1412.“ *The Lancet* 368:1569-1570.

Hau, Jann. 2014. „Animal Models.“ In *Handbook of Laboratory Animal Science, 3rd edition, volume II: Animal Models*, edited by Jann Hau and Steven J. Schapiro. 1-10. Boca Raton, Florida: CRC Press / Taylor and Francis.

Hay, Michael, David W. Thomas, John L. Craighead, Celia Economides, and Jesse Rosenthal. 2014. "Clinical development success rates for investigational drugs." *Nature Biotechnology* 32: 40-51.

Holmes, Frederic L. 1993. "The Old Martyr of Science: The Frog in Experimental Physiology." *Journal of the History of Biology* 26: 311-328.

Horton, Richard. 2006. "Phase I clinical trials. A call for papers." *The Lancet* 368: 827.

Huber, Lara, and Lara K. Keuck. 2013. "Mutant mice: Experimental organisms as materialised models in biomedicine." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 385-391.

Hünig, Thomas. 2012. „The storm has cleared: lessons from the CD28 superagonist TGN1412 trial.“ *Nature Reviews Immunology* 12: 317-318.

Jung, Hermann. 1956. "Klinische Erfahrungen mit einem neuen Sedativum." *Arzneimittelforschung* 6: 430-432.

Kaiser, Jocelyn. 2006. "A Cure for Medicine's Ailments?" *Science* 311: 1852-1854.

Kenter, M[arcel] J. H., and A[dam] F. Cohen. 2016. "Establishing risk of human experimentation with drugs: lessons from TGN1412." *The Lancet* 368: 1387-1391.

King, C. T. G., and F. J. Kendrick. 1962. "Teratogenic effects of Thalidomide in the Sprague Dawley Rat." *The Lancet*, November 24: 1116.

Kirk, Beate. 1999. *Der Contergan-Fall: eine unvermeidbare Arzneimittelkatastrophe? Zur Geschichte des Arzneistoffs Thalidomid*. Stuttgart: Wissenschaftliche Verlagsgesellschaft.

Koch, Robert. 1890a [1912]. „Über bakteriologische Forschung.“ In Robert Koch, *Gesammelte Werke*, vol. 1, edited by J. Schwalbe, 650-660. Leipzig: Georg Thieme.

Koch, Robert. 1890b [1912]. „Weitere Mitteilungen über ein Heilmittel gegen die Tuberkulose.“ In Robert Koch, *Gesammelte Werke*, vol. 1, edited by J. Schwalbe, 661-668. Leipzig: Georg Thieme.

Koch, Robert. 1891 [1912]. „Weitere Mitteilung über das Tuberkulin.“ In Robert Koch, *Gesammelte Werke*, vol. 1, edited by J. Schwalbe, 673-682. Leipzig: Georg Thieme.

Kohler, Robert E. 1994. *Lords of the Fly: Drosophila Genetics and the Experimental Life*. Chicago: University of Chicago Press.

Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions. 2nd Edition, Enlarged*. Chicago: The University of Chicago Press.

Kunz, W[ilhelm], H[erbert] Keller, and H[einrich] Mückter. 1956. „N-Phthalyl-glutaminsäure-imid: Experimentelle Untersuchungen an einem neuen synthetischen Produkt mit sedativen Eigenschaften.“ *Arzneimittel-Forschung* 6: 426-430.

Leader, R[obert] W. 1969. Living models for the study of disease processes in man: Discovery and exploitation of animal model diseases. *Federation Proceedings* 28: 1804-1809.

Leader, R[obert] and G. A. Padgett. 1980. “The genesis and validation of animal models.” *American Journal of Pathology* 101, iss. 3S (= Suppl. to December): S11-S17.

Lederer, Susan. 2004. “Research without borders: The origins of the Declaration of Helsinki.” In *Twentieth Century Ethics of Human Subject Research: Historical Perspectives on Values, Practices, and Regulations*, edited by Volker Roelcke and Giovanni Maio. 199-218. Stuttgart: Steiner.

Lemoine, Mael. 2017. “Animal extrapolation in preclinical studies: An analysis of the tragical case of TGN1412.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 61: 35-45.

Lenhard-Schramm, Niklas. 2017. “Contergan und das Arzneimittelrecht.” In: *Contergan. Hintergründe und Folgen eines Arzneimittel-Skandals*, edited by Thomas Großbölting and Niklas Lenhard-Schramm. 135-165. Göttingen: Vandenhoeck & Ruprecht.

Leonelli, Sabina. 2007. „Growing Weed, Producing Knowledge. An Epistemic History of *Arabidopsis thaliana*.“ *History and Philosophy of the Life Sciences* 29: 55-87.

Leonelli, Sabina, Rachel A. Ankeny, Nicole C. Nelson, and Edmund Ramsden. 2014. “Making Organisms Model Human Behavior: Situated Models in North-American Alcohol Research, 1950-onwards.” *Science in Context* 27: 485-509.

Lenz, W[idukind]. 1962a. “Thalidomide and congenital abnormalities.” *The Lancet* 279 (7219): 45.

- Lenz, Widukind. 1962b. "Entstehung von Missbildungen durch Medikamente." *Ärztliche Mitteilungen* 9 (3 March 1962): 494.
- Lenz, W[idukind]. 1988. "A short history of thalidomide embryopathy." *Teratology* 38: 203-215.
- Lenz, W[idukind], and K. Knapp. 1962. "Die Thalidomid-Embryopathie." *Deutsche Medizinische Wochenschrift* 87: 1232-1242.
- Levy, Arnon, and Adrian Curry. 2015. "Model Organisms are (Not) Theoretical Models." *British Journal of the Philosophy of Science* 66: 327-348.
- Lin, Chia Huey, and Thomas Hünig. 2003. "Efficient expansion of regulatory T cells in vitro and in vivo with a CD28 superagonist." *European Journal of Immunology* 33: 626-638.
- Linton, Derek S. 2005. *Emil von Behring: Infectious disease, Immunology, Serum Therapy*. Philadelphia: American Philosophical Society.
- Löwy, Ilana, and Jean-Paul Gaudillière. 1998. "Disciplining cancer: Mice and the practice of genetic purity." In *The invisible industrialist: Manufactures and the production of scientific knowledge*, edited by Jean-Paul Gaudillière and Ilana Löwy. 209-249. London: Palgrave Macmillan.
- Maulitz, Russell C. 1978. "Rudolf Virchow, Julius Cohnheim, and the program of Pathology." *Bulletin of the History of Medicine* 52: 162-182.
- Maus, Hans-Jörg, and Klaus Stumpe. 1963. "Tierexperimentelle Untersuchungen zur Frage der Thalidomid-Embryopathie." *Klinische Wochenschrift* 41: 21-25.
- McKinney Jr. WT, Bunney Jr WE. 1969. "Animal model of depression. I. Review of evidence, implications for research." *Archives of General Psychiatry* 21: 240-248.
- McBride, William G. 1961. "Thalidomide and congenital abnormalities." *The Lancet* II: 1358.
- N. N. 1962. "[Annotations:] Thalidomide and After." *The Lancet*, November 24: 1095.
- N. N. 2012. "Editorial: A decade in drug discovery." *Nature Reviews Drug Discovery* 11: 3.
- Pallardy, M[arc] and T[homas] Hünig. 2010. "Primate testing of TGN1412: right target, wrong cell." *British Journal of Pharmacology* 161: 509-511.
- Parascandola, John. 1981. "The theoretical basis of Paul Ehrlich's chemotherapy." *Journal of the History of Medicine and Allied Sciences* 36: 19-43.
- Perrin, Steve. 2014. "Make mouse studies work." *Nature* 507: 423-425.

Pfeiffer, R[udolf] and W[ilhelm] Kosenow. 1962. "Zur Frage einer exogenen Verursachung von schweren Extremitätenmissbildungen." *Münchener Medizinische Wochenschrift* 104: 68-74.

Pliess, G. 1962. "Thalidomide and congenital abnormalities." *The Lancet* I (May 26): 1128-1129.

Randerson, James. 2016. "Fatal French clinical trial failed to check data before raising drug dose." *Nature*: doi:10.1038/nature.2016.21190.

Rheinberger, Hans-Jörg. 1997. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford: Stanford University Press.

Rheinberger, Hans-Jörg. 2005. "Überlegungen zum Begriff des Modellorganismus in der biologischen und medizinischen Forschung." In *Modelle des Denkens*, edited by Berlin-Brandenburgische Akademie der Wissenschaften. 69-74. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften.

Rheinberger, Hans Jörg. 2006. "Vers la fin des organismes modèles?" In *Les organismes modèles dans la recherche médicale*. edited by Gabriel Gachelin. 275-277. Paris: Presses Universitaires de France.

Rheinberger, Hans-Jörg, and Michael Hagner (eds.). 1993. *Die Experimentalisierung des Lebens: Experimentalsysteme in den biologischen Wissenschaften 1850/1950*. Berlin: Akademie.

Roelcke, Volker. 2009. „Tiermodell und Menschenbild. Konfigurationen der epistemologischen und ethischen Mensch-Tier-Grenzziehung in der Humanmedizin zwischen 1880 und 1945.“ In *Kulturgeschichte des Menschenversuchs im 20. Jahrhundert*, edited by Birgit Griesbeck, Marcus Krause, Nicolas Pethes, and Katja Sabisch, 16-47. Frankfurt/Main: Suhrkamp.

Roelcke, Volker. 2013. „Repräsentation – Reduktion – Standardisierung: Zur Formierung des ‚Tiermodells‘ menschlicher Krankheit in der experimentellen Medizin des 19. Jahrhunderts.“ In *Tier – Experiment – Literatur, 1880-2010*, edited by Roland Borgards and Nicolas Pethes, 15-36. Würzburg: Königshausen & Neumann.

Roelcke, Volker. 2017. "The use and abuse of medical research ethics: The German *Richtlinien* / guidelines for human subject research as an instrument for the protection of research subjects – and of medical science, ca. 1931–1961/64." In *From Clinic to Concentration Camp: Reassessing Nazi Medical and Racial Research 1933-1945*, edited by Paul Weindling, 33-56. London: Routledge.

Schmiedebach, Heinz-Peter. 1993. „Pathologie bei Virchow und Traube: Experimentalstrategien in unterschiedlichen Kontexten.“ In *Die Experimentalisierung des Lebens: Experimentalsysteme in den biologischen Wissenschaften 1850/1950*, edited by Hans-Jörg Rheinberger and Michael Hagner, 116-134. Berlin: Akademie.

Schneider, Christian K., Ulrich Kalinke, and Johannes Löwer. 2006. „TGN1412 – a regulator’s perspective.“ *Nature Biotechnology* 24: 493-496.

- Seller, Mary J. 1962. "Thalidomide and congenital abnormalities." *The Lancet* (August 4): 249.
- Shanks, Niall, Ray Greek, and Jean Greek. 2009. "Are animal models predictive for humans?" *Philosophy, Ethics, and Humanities in Medicine* 4. doi:10.1186/1747-5341-4-2.
- Simon, Jonathan. 2007. „Emil Behring's Medical Culture: From Disinfection to Serotherapy.“ *Medical History* 51: 201-218.
- Somers, G. F. 1962. "Thalidomide and congenital abnormalities." *The Lancet* (April 28): 912-913.
- Stärk, G. 1956. "Klinische Erfahrungen mit dem Sedativum K 17 in der Lungenheilstätte und der allgemeinen Praxis." *Praxis* 45: 966-968.
- Stapel Ute. 1988. *Die Arzneimittelgesetze 1961 und 1976*. Stuttgart: Deutscher Apotheker Verlag.
- Stebbing, R., D. Eastwood, S. Poole, and R. Thorpe. 2013. "After TGN 1412. Recent developments in cytokine release assays." *Journal of Immunotoxicology* 10: 75-82.
- Suntharalingam, Ghanesh, Meghan R. Perry, Stephen Ward, Stephen J. Brett, Andrew Castello-Cortes, Michael D. Brunner, and Nicki Panoskaltis. 2006. "Cytokine Storm in a Phase 1 Trial of the Anti-CD28 Monoclonal Antibody TGN1412." *New England Journal of Medicine* 355: 1018-1028.
- Tacke, M., G. Hanke, T. Hanke and T. Hünig. 1997. „CD28-mediated induction of proliferation in resting T cells in vitro and in vivo without engagement of the T cell receptor: evidence for functionally distinct forms of CD28.“ *European Journal of Immunology* 27: 239- 247.
- TeGenero AG (2005). TGN1412 investigator's brochure. Available at: http://www.circare.org/foia5/tgn1412_investigatorbrochure.pdf (accessed 26 October 2020).
- Tranter, Elizabeth, Gary Peters, Malcolm Boyce and Steve Warrington. 2013. "Giving monoclonal antibodies to healthy volunteers in phase 1 trials: is it safe?" *British Journal of Clinical Pharmacology* 76: 164-172.
- Van der Worp, H. Bart, David W. Howells, Emily S. Sena et al. 2010. Can Animal Models of Disease Reliably Inform Human Studies? *PloS Medicine* 7: 00245. doi:10.1371/journal.pmed.1000245.
- Virchow, Rudolf. 1891. "Ueber die Wirkung des Koch'schen Mittels auf innere Organe Tuberculöser." *Berliner klinische Wochenschrift* 28: 49-52.
- Ward, S. P. 1962. "Thalidomide and congenital abnormalities." *British Medical Journal* 8 Sept (5305): 646-647.

Wegerle, Hans. 1962. "Beitrag zur Häufung von kongenitalen Missbildungen der Gliedmaßen und ihre mögliche Ätiologie." *Medizinische Klinik* 57: 877-883.

Weicker, Heinz und Heinz Hungerland. 1962. "Thalidomid Embryopathie I: Vorkommen inner- und außerhalb Deutschlands." *Deutsche Medizinische Wochenschrift* 87: 992-1002.

Weindling, Paul. 2001. „The origins of informed consent: The international scientific commission on medical war crimes and the Nuremberg code." *Bulletin of the History of Medicine* 75: 37-71.

Wiedemann, Hans-Rudolf. 1961. "Hinweis auf eine derzeitige Häufung hypo- und aplastischer Fehlbildungen der Gliedmaßen." *Die Medizinische Welt* 12: 1863-1866.

Wiedemann, Hans-Rudolf. 1962. "Derzeitiges Wissen über Exogenese von Missbildungen im Sinne von Embryopathien beim Menschen." *Die Medizinische Welt* 13: 1343-1349.

Wimmelbücker, Ludger. 2017. "Grippex 1956-1961: Ein anderer Blick auf die Geschichte thalidomidhaltiger Medikamente in der Bundesrepublik Deutschland." In *Contergan – Hintergründe und Folgen eines Arzneimittel-Skandals*, edited by Thomas Großbölting and Niklas Lenhard-Schramm, 167-203. Göttingen: Vandenhoeck & Ruprecht.

Woollam D[avid] H. M. 1962a. "Thalidomide disaster considered as an experiment in mammalian teratology." *British Medical Journal* II (5299, July 28): 236-237.

Woollam, D[avid] H. M. 1962b. "Drugs and Foetal Abnormalities." *British Medical Journal* II (iss. 5301, Aug. 11): 406.

Zweites Gesetz zur Änderung des Arzneimittelgesetzes vom 23. Juni 1964. 1964. *Bundesgesetzblatt* no. 30 (27. Juni 1964): 365-369.