Animal Models of Tuberculosis: An Overview

ANN WILLIAMS¹ and IAN M. ORME²
¹Health UK, Porton Down SP4 0JG, United Kingdom; ²Colorado State University, Fort Collins, CO 80523

ABSTRACT This article provides an overview of the animal models currently used in tuberculosis research, both for understanding the basic science of the disease process and also for practical issues such as testing new vaccine candidates and evaluating the activity of potential new drugs. Animals range in size, from zebrafish to cattle, and in degrees of similarity to the human disease from both an immunological and pathologic perspective. These models have provided a great wealth of information (impossible to obtain simply from observing infected humans), but we emphasize here that one must use care in interpreting or applying this information, and indeed the true art of animal modeling is in deciding what is pertinent information and what might not be. These ideas are discussed in the context of current approaches in vaccine and drug development, including a discussion of certain limitations the field is currently facing in such studies.

Animal models are an integral part of the scientific process, reflecting the physiological and anatomical similarities between many animal species and human beings. In the context of infectious diseases, multiple animal models have been used to extend our understanding of their pathophysiology and the host response to them. This is the backbone also of vaccine research, producing vaccines against once-dreaded multiple diseases that in previous centuries claimed the lives of many millions of people. Animal models are also invaluable in designing therapies, particularly drugs, with which to combat these diseases.

It is important to consider, however, that animals are not humans, even when the species used is genetically very close to Homo sapiens. Accordingly, care has to be used in interpreting data from a specific animal model and it cannot be absolutely guaranteed that the specific observation—for instance, disease pathology or a T-cell subset response—is identical to events happening in an infected human. In other words, while animal models can collectively provide a massive amount of information, not all of it will apply to human disease, and the true art of animal modeling is to understand what information is useful and pertinent, and what may be less useful and even potentially misleading.

Scientists who use animal models are acutely aware of the ethical issues in using other creatures that share our planet. Regulations and an ethical code of conduct ensure that suffering is avoided completely or at least kept to a minimum, and, in the case of chronic diseases such as tuberculosis, animals are carefully monitored and euthanized when they start to show clinical signs (such as weight loss) that are predictive of severe disease. Alternatively, disease burden can be measured at a fixed time point prior to the development of clinical signs. In addition, there is a consensus that, wherever possible, a minimum number of animals are used that can still provide a statistically valid result. These considerations have resulted in the current concept of “replacement, reduction, and refinement” for animal usage, first suggested in 1959 (1).
MAJOR ANIMAL MODELS OF TUBERCULOSIS

The use of animal models in tuberculosis research extends back to the 19th century when the identity of the tuberculosis bacillus—and the idea that tuberculosis just “ran in families” was incorrect and, instead, that the disease had an infectious origin—were only just being realized. Koch, after his seminal discovery, soon found that if he injected cultures of the Mycobacterium tuberculosis bacilli into mice, they developed lesions not unlike those seen in the lungs of patients, and it was soon discovered that the organism could also cause disease in multiple animals, including rabbits, guinea pigs, and rats. It became apparent that there were differences in apparent susceptibility between the species, with infected mice often outliving guinea pigs as an example, and that, in addition, the environment also played a role. The latter idea was first tested by E. L. Trudeau, who showed that rabbits kept in an environment that included sunlight and good nutrition (on “Rabbit Island”) lived longer than rabbits kept in a basement laboratory with poor nutrition.

Mice

As described elsewhere (77), studies in animal models, particularly the mouse, have provided a wealth of information about the functions of the immune response against M. tuberculosis. Early studies, including observations in athymic mice and cell transfer studies (2–5), first pointed to the central role of T lymphocytes in protective immunity, and over the past 3 decades or so have blossomed into a progressively deeper understanding of the complex immune response this organism elicits. This has happened in concert with refinement of the mouse model. An initial movement from the use of wild mice to inbred strains of mice can be traced back to the first MHC studies by Gorer, and these inbred strains started in appear in the tuberculosis field about 40 years ago. Advances in mainstream immunology have subsequently generated transgenic mice that have a variety of uses, as well as congenic lines (CD45, for example) that can be used to track cells of host origin after cell transfer. Mutations arising in inbred strains have also been useful; as an example, the FeJ mutant on the C3H background develops severely necrotic tuberculous lesions not seen in other inbred strains (6).

Perhaps the most important advance in the mouse model was the development of technology that allowed targeted disruption of individual genes in this animal, providing “gene knockout mice.” This provided a wealth of information regarding the role of various molecules in host immune systems that substantially advanced our knowledge of these. This included the central role of gamma interferon, long suspected but definitively proven in IFNγ-KO mice (7, 8), and the pleiotropic roles of tumor necrosis factor alpha (9). Mice lacking CD4 cells lost their ability to resist infection (10), whereas this was less pronounced in CD8-deficient mice until well into the chronic phase of the disease (11). In addition, the relative roles of myriad cytokines and chemokines were revealed by this new technology. In fact, given that disruption of many different genes mostly resulted in loss of resistance in mice to tuberculosis, either completely or temporarily, directly illustrated the highly complex and integrated nature of the overall host response.

Guinea Pigs

Guinea pigs are an extremely useful model of tuberculosis because they exhibit multiple similarities to human disease, especially lung necrosis, lymphadenopathy, and disease dissemination (12, 13). For this reason, they have been generally regarded as the gold standard for testing vaccine efficacy, their primary use in the field for decades (14, 15). A limitation of the model for many years was the lack of immunologic reagents that could be used to measure their immune response, but this situation has recently changed, principally because of the efforts of McMurray, who painstakingly developed assays for guinea pig cytokines and chemokines, and Ordway, who solved the serious problem of autofluorescence, allowing the application of flow cytometry to immune cells in this model.

In fact, early studies in tuberculosis primarily used the guinea pig model, both for vaccine testing and for developing skin test diagnostic reagents. This popularity reflected the susceptibility of this animal to tuberculosis, in contrast to other models such as the mouse and rat. This is not absolutely the case, however, and the general consensus—based on the use of laboratory strains—that “mice are resistant, guinea pigs are susceptible” has been recently challenged (16, 17), reflecting newer information arising from the study of clinical strains. In these newer studies, it has been observed that, first, certain newly emerging clinical isolates can grow even better in mice than in guinea pigs, and, second, guinea pigs can live for a significant period of time despite bearing large necrotic lesions in their lungs (18, 19).

In contrast to vaccines, the use of the guinea pig to test drugs against tuberculosis has been much more limited and confined mainly to the classical literature. This situation has undergone a renaissance recently, however, with the development of humane methods to administer
drugs over extended periods of therapy (20), with an important outcome of this being the first demonstration that an experimental new drug, bedaquiline, substantially reduced the duration of therapy when given with conventional drugs (21).

Non-Human Primates
Over the past decade or so, considerable effort has been put into the development of non-human primate (NHP) models of tuberculosis, and there are now several world-class facilities with this capability. The most commonly used NHP for tuberculosis (TB) studies is the macaque species (rhesus and cynomolgus), and several reviews have been written describing M. tuberculosis infection in these species (22–24). Because of genetic similarities with humans, reflected in our similar immune systems, there is general agreement that NHPs are the most important “gateway” for progression to efficacy testing in humans. Currently, there are various permutations of the macaque model with regard to species, route of challenge, and the primary readout of disease burden that is used to demonstrate vaccine efficacy. Considerable efforts are currently under way to better understand the impact of these different variables with the goal of more standardized models to allow reproducible vaccine efficacy testing. This will enable critically important studies where candidate vaccines are tested for efficacy in NHPs in study designs that bridge to early clinical studies of the same vaccines.

OTHER ANIMAL MODELS
Zebrafish
The zebrafish is a (very) small animal model of tuberculosis that has contributed useful information in terms of the pathogenesis of the disease. It has various advantages, including ease of use, and optical transparency in its larval stage, making it very amenable to imaging techniques (25).

The zebrafish cannot be infected with M. tuberculosis because of its much lower body temperature, but it can be productively infected with Mycobacterium marinum, a genetically related mycobacterium that can infect cold-blooded aquatic animals such as fish and frogs.

Like M. tuberculosis, M. marinum can be genetically manipulated, allowing the generation of mutant and transgenic strains. Zebrafish develop granulomatous-like structures when infected with M. marinum, which, as in many other animal models, can degenerate and become necrotic (26). As in mice and guinea pigs, the early granulomatous response is insufficient to contain the infection initially, and bacilli disseminate to other organs. In addition, the zebrafish expresses multiple elements of both innate and acquired immunity. In this model, however, acquired immunity takes several weeks to develop (27), whereas in mammals this occurs much faster.

A major strength of the model is that larval zebrafish are optically transparent and this allows serial observations using high-resolution microscopy (28). In addition, the animal can easily be modified genetically, and the effects of inactivation of genes using antisense oligonucleotides can persist for several days (29). A further useful property is that larval zebrafish have no phagocytes in their hindbrain cavity (HBC) and so, if bacilli are injected here, the kinetics of ingress of macrophages and neutrophils can be measured. If the bacilli are injected via the caudal vein, they encounter such cells straight away. If the HBC route is used, it can also be used to show cellular accumulations if host molecules are injected into this site, such as interleukin-8, CCL2, or leukotriene B4 (30). This is a unique and useful aspect of this model.

The zebrafish model provides data suggesting avoidance by M. marinum of the influx of neutrophils into lesions (30). This is different from observations in mice and in guinea pigs where early neutrophil influx has been implicated in early lesion necrosis (12, 31). Observations in humans (32, 33) do not support this either, suggesting a different mechanism of pathogenesis between M. marinum and M. tuberculosis, and draws into question interpretations drawn in fish (29) versus hosts possessing lungs (34).

Rabbits
Rabbits were used by Trudeau as a disease model at the end of the 19th century, but the model came to prominence in the 1930s because of the seminal studies by Lurie who described the disease process and developed inbred lines of rabbits of differing susceptibility (35, 36). This was followed by further studies performed in conjunction with Dannenberg that started to provide a framework for the first important models of the pathogenesis of the disease (37), as well as a distinction between protective cellular immunity, and potentially damaging mechanisms of “delayed type hypersensitivity” (38–40).

Rabbits can be infected with the laboratory strains H37Rv and Erdman, but they are quite resistant to these strains and, as a result, a larger dose of viable bacilli (>10³) is often needed to establish a productive infection (41). In contrast, infection of these animals with the
Ravenel strain of Mycobacterium bovis causes a rapid and severe disease process in which large areas of the lung become necrotic and then liquefy (42). This event has been regarded as a model of human disease cavity formation, but not everyone agrees (43), and it may reflect a different mechanism of pathogenesis separate from that caused by M. tuberculosis.

Although there may be useful attributes of the model, the larger size of rabbits makes the model more expensive and less tractable than mice or guinea pigs and it has never been seriously developed for vaccine or drug testing. Where it has provided some useful information is as a model of cerebral tuberculosis (44), although, here again, virulent M. bovis is needed to effectively establish this.

Rats
The rat model was first described in detail in seminal articles by Gray in the late 1950s in studies in which he compared the outcomes of infection in rats and mice (45–47). Thereafter, the literature on this model is very sparse until two 1973 studies by Lefford in which he demonstrated that thoracic duct lymphocytes (but not serum) conferred protection to recipient rats challenged with M. tuberculosis H37Rv, thus providing the first clues that immunity to tuberculosis was mediated by lymphocytes (48, 49). In recent years, however, very little has been done with this model, except as a potential diabetes/tuberculosis model (50).

Cattle
Cattle are natural hosts of M. bovis and are one of the target species (alongside wildlife reservoirs) for interventions aimed at controlling bovine TB. Highly relevant host-pathogen interaction studies can therefore be performed by experimental infection of cattle. The close similarity of M. bovis to M. tuberculosis means that there is potential synergy in the efforts to control human and bovine TB. An example of this is model and demonstrated efficacy (51). The relatively large size of these animals presents challenges in terms of housing at biosafety level 3 containment and quantities of drug or vaccine needed but there are advantages similar to those of the non-human primates in terms of frequent and sequential sample collection. A major advantage of the cattle model of M. bovis is the ability to conduct field trials, including natural transmission studies, where it is possible to evaluate the ability of a vaccine to prevent disease caused by a naturally acquired infection.

Mini Pigs
A mini pig model of M. tuberculosis infection has been described (52) as an alternative means of studying latent TB infection and the impact of treatments on lung lesion development. A key advantage of this model is related to the anatomy of the lungs which, because of the large size of these animals, and like humans, have intralobular septae that influence the development of granulomatous pathology. Similar to the cattle model for bovine TB, there are considerable practical and cost implications that prevent this model from being applied widely.

PRACTICAL APPLICATIONS OF ANIMAL MODELS

Host Response and Pathogenesis
For several decades, animal models have been used to try to predict the human immune response to tuberculosis infection, with the aim to improve the design of vaccines and diagnostic approaches. This basic research in animal models has provided fundamental information, the most important being that vaccines against TB have to induce protective T cells, rather than establish antibody responses as many other vaccines do. We now know that a favorable host response is driven by recognition systems, principally Toll-like receptors, which has helped in the development of new adjuvant formulations designed to maximize the T helper 1 (TH1) response, particularly to protein-based vaccine candidates. Animal studies have illustrated that the TH1 mechanism is not exclusive, however, and that, even within the CD4 subset, there are additional types of cells (interleukin-17-secreting cells, Foxp3+ regulatory T cells) that are clearly part of the overall equation, and that, moreover, additional subsets including NK, NKT cells, and CD8+ MAIT cells may be playing a contributory role (53, 54).

Animal models have also provided information about other aspects that could not be obtained by simply observing humans. One important example is the nature of the memory T-cell response (55), which is of course the primary target of new vaccines, discussed further below. Despite this, the field is still having difficulty on occasion in translating information from animals to humans, the most notorious being the search for “correlates of protection” (56). It would be invaluable to have biomarkers indicating that a response to a given vaccine gave very high confidence that the individual was likely to be completely protected, or that a drug regimen was completely sterilizing, but this has not been achieved. The most
famous example was the concept that if an interferon gamma (IFNγ) response was present, this indicated protection, which has not turned out to be useful.

All the animal models studied to date develop granulomatous inflammation resulting in the formation of a wall of tissue called a granuloma, which is designed to try to contain the infection (57). Our ability to form these structures probably arose from more primitive responses to particles, which gradually evolved into a more complex response centrally controlled by chemokines released by host immunity. The fine structure of the granuloma differs considerably between different animal models, ranging from relatively acellular granulomas in zebrafish to highly calcified lesions in cattle. Mice and guinea pigs differ considerably as well, and, while many reviews describe the mouse granuloma as disorganized and those in guinea pigs as organized, in fact the reverse is true (57). In mice, there are highly organized foci of CD4 cells as well as some small follicles of B cells, surrounded on the perimeter by CD8 cells. This type of pattern can be seen in human lungs, as well. In guinea pigs, however, the center of the granuloma becomes necrotic, as in humans, and this compresses the cellular response, leading to a disorganized cuff of randomly mixed CD4 and CD8 T cells.

Non-necrotic secondary lesions are a facet of the guinea pig model, and these tend to be pleural, supporting the idea that they get established by bacilli escaping from primary lesions and being carried down the peripheral lymphatics, an idea that challenges the classical view that secondary sites of infection are established by blood-borne bacteria. A further element of this is a cellular response in the lymphatic vessels themselves—lymphangitis—clearly visible in this animal model (58).

In all the models, there is clear evidence that granulomas are dynamic. Some become necrotic, whereas others show evidence of wound healing and/or fibrosis. Granuloma formation in zebrafish (29) can be observed after only a few days, whereas this takes much longer in mammals. This leads some (16) to suggest that the response in zebrafish is more akin to a primitive particle response rather than a true granuloma. While both zebrafish and mammalian models quickly develop an “epithelioid macrophage” field of cells, major aggregates of lymphocytes, seen in larger animal models, are less developed.

These differences may explain why macrophages in zebrafish entering sites of infection then leave and disseminate disease (59). Because this happens in a hemocoel, this is inevitable and is probably different from events in the lungs of larger models in which dendritic cells pick up bacilli in the interstitium and then enter lymphatic capillaries to carry them to draining lymph nodes, an essential event in triggering acquired immunity (60).

Animal models are invaluable for the study of pathological changes in response to infection because the kinetics and dynamics of the response can be described in relation to a single infection event. It could be argued, however, that this is not relevant for human TB infection, which occurs in the context of multiple exposures. More complex animal models are being considered or used that involve repeated exposure and exposure to natural aerosols. In the latter case, where guinea pigs were exposed to natural aerosols from TB patients, an unexpected spectrum of disease pathology was revealed that did not match the classical pattern of disease described following experimental infection (61). Although this altered pathology could have been explained by the differences in the M. tuberculosis strain that caused infection, such studies serve to illustrate that the findings from experimental studies should be interpreted with caution and should, wherever possible, be corroborated by findings in humans rather than become the accepted dogma.

**Assessment of Vaccines**

Vaccines have been tested in animal models for many decades and have enabled the progression of several candidates through to efficacy testing in humans. Despite some promising candidates being identified during the preclinical stages of testing, a new candidate that can facilitate the current BCG vaccine, or even replace it, has not been identified. Ongoing efficacy trials may yet yield a superior vaccine but, in the meantime, the research and development activities continue.

**Process and Capacity**

In a perfect world, a new candidate would be tested in a minimum of two different animal species, by at least two separate laboratories that themselves have no vested interest, using adequate vaccination to challenge intervals, with sufficient animals in each group to provide statistical power, and with relevant clinical isolates as the challenge inoculum. Once a reasonable number of vaccines have been identified as “active,” then head-to-head evaluations would help prioritize the candidates for final testing in NHPs.

A lack of resources and a limited degree of global cooperation have, until now, prevented this. A large number of candidates have been tested in mice, reflecting
the fact that many laboratories have that capability but, frequently, the candidate does not progress to being tested in more advanced or stringent models. There is a variety of reasons for this, but cost and capacity are the main ones. Even at the guinea pig level, very few facilities are equipped (and have the experience) to conduct appropriately controlled and powered efficacy testing using this model. The costs associated with caging or feed are much higher than studies in mice, and so the cost of even a relatively small study can become prohibitive. One solution to this is for vaccine developers to apply for their candidates to be tested by independent laboratories that have expertise in the animal models rather than in vaccine development. Such facilities are supported by NIH funding in the United States and by the European Union. These facilities provide increased capacity, an independent evaluation of efficacy, and the capability to conduct comparative evaluations where the efficacy of one candidate can be compared directly with another. Given the advantages of independent testing facilities, it would be beneficial to the field to mobilize resources to increase the capacity in terms of including more laboratories and extending the models that are offered, beyond mice and guinea pigs.

**Mechanism of Protection**

BCG is invariably used as a positive control in evaluating new candidates, but this imprints a built-in bias (62). It assumes that a new candidate will induce protective mechanisms that have similar kinetics to BCG, and will result in at least a minimum of 0.5 to 0.7 log_{10} reduction in lung CFU levels after low-dose aerosol infection. If the new candidate works differently, such as needing a finite time as well as boosting to achieve a stable, long-lived memory T-cell response, or, alternatively, if it does not reduce the bacterial load but still can establish granulomas that are highly stable and bacteriostatic, then, in our current screening protocols, it would be rejected as inactive (63). There is therefore a risk that new candidates or novel mechanisms of action have been lost because of this assumption. In this regard, a vaccine that induces an immune response but completely fails to provide any signal of efficacy can provide extremely useful information about mechanisms of protection. Unfortunately, such information is often not accessible because of the tendency for reporting only positive efficacy data. The importance of negative data is recognized by most researchers, yet it remains extremely difficult to get these data published and made available to others.

**Vaccine Testing Protocols**

Many factors must be considered when constructing models for vaccine testing. The first major consideration is the vaccine to challenge interval; in many protocols, the challenge infection is given at the peak of the effector T-cell response to the (positive control) BCG vaccine. Although a candidate may demonstrate protective immunity (assuming it has similar kinetics to BCG) under such a protocol, no information will be obtained about the induction of adequate memory immunity. For example, BCG establishes memory immune T cells in the lungs in about 10 to 15 weeks, but, in some protocols, the challenge infection is given only 6 to 8 weeks post-BCG.

The other key consideration is the infectious challenge, since the strain of *M. tuberculosis* used, the route, and dose of challenge can all have an impact on the vaccine-induced host response. This could be termed “bacterial immunogenicity,” which is the speed with which host immunity recognizes the presence of the challenge infection. It is probably related to the kinetics of bacterial production of immunodominant antigens such as ESAT6. Most laboratories rely on challenge studies that use either H37Rv or Erdman as the infectious agent, since these strains replicate well in most of the animal models, they are drug-sensitive, and they allow direct comparisons of data between laboratories. However, there is a growing concern that they do not reflect reality, given the increasing frequency of newly emerging clinical isolates that are more virulent and induce much broader T-cell responses (64).

There is evidence that strain fitness is an important consideration, and this has recently been proposed (18) as a factor in the failure of the MVA85A vaccine to boost BCG in a phase IIb efficacy trial. This arose from the observation that, while highly virulent Beijing strains from the United States and the Western Cape (where the trial was conducted) grew equally well in unvaccinated guinea pigs, BCG strongly inhibited strains from the Cape, whereas its activity against the U.S. strains was far more variable. This suggested the possibility (yet to be proven) that, while the Western Cape strains were highly virulent, they were also of relatively low fitness and it was not possible to see the effect of the MVA85A boost above that afforded by BCG alone (65). This is supported by data that suggest that host recognition of these infections is rapidly leading to more effective control by BCG (18).

The impact of vaccine on challenge interval and challenge strain is illustrated in Fig. 1. Following introduction of a vaccine, there is an effector immune response, which
eventually contracts and is replaced by an emerging memory immune response. This takes a finite time, and, in many protocols, boosting vaccines are given before memory has had a chance to become fully established (explaining why many boosting candidates fail). After the challenge infection is given, the speed of recognition of the infection depends on its native immunogenicity, coupled with its ability to grow, with “ΔCFU with time” reflecting its intrinsic virulence. Finally, what is not always taken into account is that an isolate can be virulent, but also vary in fitness, with high-fitness strains more likely to be only transiently affected by the immunity the vaccine has generated.

Assessment of New Drugs
Animal models have played a significant role in the assessment of new drugs against tuberculosis, beginning with the testing of streptomycin in guinea pigs by Feldman in the mid-1950s, followed by seminal studies in various animal species pioneered by Mitchison, Grosset, and others since then. In general, drugs found to be effective in these models tend to be equally effective in humans. In fact, most of the conventional drug regimens in use today were thoroughly tested in mice and other models.

Basic parameters needed to advance a potential tuberculosis drug are obtained from such models, including safety, maximum tolerated dose, and pharmacokinetic and pharmacodynamic determinations. This remains useful today, because the emergence of drug resistance has spurred the continued development of new drugs for tuberculosis.

The mouse models continue to advance and undergo refinements. The availability of IFNγ gene-disrupted mice provided a rapidly progressive infection model (66), which had the benefit of substantially reducing the amount of compound needed for evaluation, as well as a rapid readout of effectiveness. Similarly, the highly reproducible reactivation of disease seen in granulocyte-macrophage colony-stimulating factor knockout mice allowed the screening of compounds to prevent this (67). More recently, the C3HeB/FeJ mutant mouse model has been used to evaluate drugs in a model of severe lung necrosis (68).
As with certain vaccine candidates, some studies in different laboratories have yielded different results. One such example is metronidazole, which had some activity in a rabbit model, but consistently failed to have any activity in several other models (69–71). A subsequent reevaluation of the mouse model (72, 73) led to the conclusion that laboratory strains should not be used just by themselves, and that there was variability even between the H37Rv strains kept in different laboratories, emphasizing the point (also made above in the context of vaccines) that the efficacy of new drug regimens needs to be tested and reproduced by more than one laboratory before moving to clinical trials. A positive outcome of these problems was that they resulted in a collective discussion between the major screening laboratory groups, with an agreement to cooperate and converge toward a rationally derived set of standardized assays and protocols to use to evaluate compounds in the future.

LIMITATIONS

To What Extent Are Animal Models Predictive?

This question is particularly relevant, given the recent failure of the MVA85A trial in South Africa (74). The vaccine candidate had undergone extensive testing, showing activity in mice, guinea pigs, and NHP, as well as being safe and immunogenic in early clinical studies, and thus many were quick to conclude that the animal models are by definition “not predictive.” This was followed by a period of considerable introspection, particularly by those reflecting on the ongoing development of alternative vaccine candidates that have progressed to the clinic following testing in the same or similar animal models as MVA85A.

The outcome of such reflection, inevitably, is that there is clearly room for improvement and refinement of the animal models and the criteria that are used to prioritize the most promising vaccines. As recently discussed (75), MVA85A was tested in multiple models but always against laboratory strains, and not against more virulent Beijing strains. MVA85A was tested in infants, whereas the animal studies used adults with fully developed immune systems. The infant trial was additionally powered to observe a much greater improvement in efficacy over BCG alone than was detected in the animal studies where the difference between BCG boosted with MVA85A and BCG alone was only ever a small improvement. In an experimental setting where many variables are controlled, small improvements in efficacy can be observed with relatively small and tractable experiments. Prior to the trial, there was no benchmark against which to evaluate the performance of the animal models, and we did not know whether the small but significant improvements in efficacy in animals had any relevance to efficacy in humans. It appears, in retrospect, that a much stronger efficacy signal may be needed in animals. This could be a greater effect in the existing models or demonstration of efficacy in more complex and stringent animal models.

A commonly used protocol to measure whether a vaccine can boost BCG in animals is to demonstrate further reduction in CFU in the organs (compared with BCG alone) at an early time point postchallenge. This protocol has high statistical power to detect small differences and is therefore useful in screening and in comparing vaccines head-to-head. It could be argued that a more relevant and therefore predictive approach is the demonstration of increased long-term survival (Kaplan-Meier analysis). This is a much more stringent test of efficacy because the difference in mean survival times between the BCG and BCG-boost vaccine needs to be substantial. This is due to the low statistical power of TB vaccine survival studies, which, as discussed previously (62), require either very large group sizes or highly potent vaccines that induce close to 100% survival.

As discussed above, a parameter that has received very little attention is the fitness of the local strains against which the vaccine is being tested. Thus, we assume that if vaccine X works against H37Rv, then it will naturally work in the field. This could be a serious flaw since strain fitness could be an important variable. If it is low (circulating in a region with malnutrition and high rates of HIV) and BCG strongly protects against it, then the boost vaccine would be expected to offer only a small improvement. It would be impossible to observe a small, statistically significant boosting effect in a trial powered to observe a 60% improvement over BCG. In other words, it is feasible that the MVA85A candidate was tested in a boosting situation where boosting was not achievable. This argues strongly for a consideration of the target population for the vaccine when designing animal studies and this includes incorporation of relevant M. tuberculosis strains.

The complexity of human TB disease and the many variables that will influence the potency of a vaccine in different populations or environments can never be fully replicated in animal models. Different animal models can recapitulate certain features and be used to test hypotheses and make decisions in vaccine development, but it is important to not to assume that all information
gleaned from animal models is predictive. Human lungs develop necrosis similarly to guinea pigs, but, in terms of cellular organization, they are more similar to that seen in mice (76). NHPs are considered the closest to humans, but there are differences in their lung anatomy—the human lung, being much larger, has secondary lobes held together by intralobular septae, which are absent in the lungs of much smaller NHP. Having recognized the limitations in terms of predicting outcomes in humans, it is imperative that, moving forward, animal models are improved wherever possible and that there is greater scrutiny and stringency when evaluating whether a candidate vaccine shows protection in animals that could translate to efficacy in humans.

Ethical and Husbandry Issues

Animal studies, even those conducted with small animals, bear a cost that is both financial and ethical. The facilities needed to conduct these studies require a high level of sophistication, particularly when operating at biosafety level 3, which involves complex engineering and equipment to maintain operator protection. There are stringent rules and regulations covering both the animal welfare and biosafety aspects that must be considered in the logistics of the experiment design and conduct.

It can be difficult to achieve the level of funding that will cover the purchase of sufficient animals combined with their facility per diem costs, and this can lead to compromise in study design, for example, using shorter vaccine-to-challenge or vaccine-to-boost intervals than is ideal. While applying the principles of “replacement, reduction, and refinement” in animal studies, it is essential that this be balanced against the need to have sufficient statistical power (63). Experiments that cannot be interpreted are not ethically justified and there may be a case for increasing the number of animals to perform a robust study that has a high benefit to (ethical) cost ratio.

Most regulatory bodies (in the United States, the U.S. Department of Agriculture) have strict rules for the use of animals in research. This includes having predetermined endpoints for humane euthanasia. Tuberculosis is a progressive disease in most animal models, and so considerable care has to be taken (both by the scientific staff and the laboratory animal facility staff) to avoid, as much as possible, stress to the animal, to use care in terms of handling or restraint, and to monitor continuously for suffering. As an example, in the context of tuberculosis in guinea pigs, measurement of weight, observation of eye color (changes from bright pink to dark pink/red as the animal sickens), and measurement (using a pulse-Ox device) of arterial oxygen tension have proven useful to some degree, whereas measurement of body temperature is not (complicated by the body’s natural diurnal rhythm).

Animal husbandry issues are themselves important. Mice are usually housed in groups of five, and where possible this should be done for guinea pigs as well, given their highly social interactions. Enrichment is important as well (as an example, mice have hours of fun shredding used toilet rolls and making nests from them). Above all, the husbandry staff needs to be aware of the risks: some species such as NHPs are dangerous and can bite, whereas others, such as rabbits, have the potential to shed bacilli in their urine.

Concluding Remarks

Animal models have underpinned TB research historically and continue to provide essential information to identify antigen targets and vaccination strategies through pathogenesis and host-response studies. Various animal models have been used to establish proof of concept for vaccine immunogenicity, safety and efficacy (and therapeutic effect of drugs) and, in conjunction with human studies, have enabled progression of several candidates to clinical trials. Animal studies should always be justifiable and appropriately designed to obtain maximum benefit from the results, and there should be a greater commitment to sharing results, particularly negative data. As clinical trial data emerge, we are learning more about the predictive value of the animal models, and it is clear that improvements can be made, particularly with regard to consideration of the target human population for the vaccines.

REFERENCES


