1 Introduction

Tuberculosis (TB) is the world’s most deadly infectious disease, with an estimated 9.6 million new cases and 1.5 million deaths annually (WHO, 2016). Incidence of infection in endemic countries remains very high despite good coverage with BCG (Bacillus Calmette-Guérin) vaccine, the only currently available vaccine (Mahomed et al., 2006; Moyo et al., 2010). There is a desperate need for a more efficacious vaccine. New candidate TB vaccines are currently tested for safety, immunogenicity and efficacy using preclinical animal models such as mice, guinea pigs, cattle and non-human primates (NHPs). Mice are the most widely used species due to the potential for screening a high number of candidates at low cost and the availability of gene knockout strains to characterize the immune response (Apt and Kramnik, 2009). However, research in species other than mice is becoming more commonplace with increasing availability of immunological reagents (McShane and Williams, 2014). Guinea pigs have emerged as a useful model, replicating many aspects of Mtb infection in humans such as granuloma formation, dissemination and caseating necrosis (Clark et al., 2015). They are also considered a more stringent model in discriminating the efficacy of different vaccines due to the variety of human-like pulmonary and extrapulmonary lesions observed (Basaraba, 2008; McMurray et al., 1996). Despite the utility of small animals in early screens, larger animal models such as cattle and NHPs are considered more relevant to human TB. NHPs are naturally susceptible to infection with Mtb, and develop the most human-like disease with latency and reactivation (Flynn et al., 2015). BCG confers some level of protection in NHPs, which can be quantified through a variety of clinical and nonclinical parameters (Sharpe et al., 2010). In the absence of a surrogate marker of protection from TB disease, animal Mtb infection models remain an essential pre-requisite for novel vaccine candidates progressing to clinical trials.

To evaluate the protective efficacy of a candidate TB vaccine, animals must be infected with virulent Mtb in a challenge experiment following vaccination. While M. bovis challenge studies in cattle are classified as “mild” in severity by the UK Home Office due to a lack of clinical symptoms, Mtb challenge experiments in mice, guinea pigs and NHPs are generally defined as “moderate”. Moderate severity indicates that the animals are likely to experience “short term moderate pain, suffering or distress... or moderate impairment of the well-being or general condition”. 

As TB disease progresses, animals may experience loss of body

---

weight, fever and respiratory distress and if left untreated will eventually die of pulmonary insufficiency (Gupta and Katoch, 2009). As this is unethical, humane euthanasia at predefined clinical endpoints, which will be discussed later in this report, is now enforced by UK Home Office legislation.

In addition to welfare concerns, the many differences between animal models and human TB bring into question the predictive value of such studies. “Protection” in animal models, as determined by the outcome of *Mtb* challenge experiments, is on a continuous spectrum and usually defined in terms of a relative improvement in a disease-related readout such as bacterial load, pathology score or long-term survival. A vaccine is considered to provide protection even if there is a measurable bacterial load or pathology in the organs or if some animals do not survive (Elias et al. 2005; McShane and Williams, 2014; Vordermeier et al., 2009). In humans, however, efficacy is binary and defined as the prevention of TB disease using clinical endpoints; any individual developing disease, however minimal, is not protected (McShane and Williams, 2014). Clearly, an artificial aerosol challenge is very different from natural transmission in humans, and the laboratory strains of *Mtb* commonly used (such as H37Rv) are genetically dissimilar to clinical isolates (Niemann and Supply, 2014), with much higher challenge doses employed (McShane and Williams, 2014). This issue has recently been addressed with new advances in ultra-low dose challenge, as discussed below. In addition to these fundamental differences in the model itself, animals are genetically distinct from humans, with several discrepancies in both innate and adaptive immunity between mice and humans (Mestas and Hughes, 2004). The widely used Balb/c and C57BL/6 mouse strains do not exhibit caseating granuloma formation following *Mtb* infection (Orme and Basaraba, 2014) and manifest a chronic phase of disease unlike latent *Mtb* infection in humans (Rhoades et al., 1997). Furthermore, responses in genetically diverse humans will be considerably more variable than in an inbred laboratory animal strain. Although outbred mice give a diversified picture of TB, which may be more representative of human disease, larger group sizes are required to offset the increase in variability (Niazi et al., 2015).

The predictive value of animal challenge models in determining TB vaccine efficacy in humans is uncertain, and will remain unclear until a successful vaccine is developed. Furthermore, there is evidence from studies of other diseases that animal models can fail to reliably predict safety in humans (Suntharalingam et al., 2006; McKenzie et al., 1995). Other difficulties include the large numbers of animals required, and the nature and slow growth of mycobacteria making experiments long and costly with the need for highly specialized Category 3 animal facilities. Given the scientific and logistical limitations of these models as well as the ethical concerns, it is imperative that potential alternatives are pursued. The principles of the 3Rs (replacement, reduction and refinement) were first proposed by Russell and Burch in the 1950s, with the aim of ensuring ethical use of animals in research (Balls, 2009). This framework, now formalized in national and international legislation, provides the basis of our discussion on the use of animals in TB vaccine research.

2 Replacement

The National Centre for the 3Rs (NC3Rs) in the UK describes replacement as “methods that avoid or replace the use of animals defined as ‘protected’ under the Animals (Scientific Procedures) Act 1986, amended 2012 (ASPA) in an experiment where they would have otherwise been used”\(^2\). Protected animals in the UK refer to all living vertebrates except humans. Alternatives include using humans, *in vitro* cell culture models, computational/mathematical modelling, or less sentient animals. All of these have been reported in the context of TB vaccine development.

2.1 Use of humans

Given the differences between human and animal manifestations of TB disease, one may argue that a more appropriate focus would be the target species. Controlled human challenge models have been successfully implemented for other pathogens, including those responsible for malaria and typhoid (Marwick, 1998; Sauerwein et al., 2011), and are a valuable tool for assessing vaccine efficacy. However, the safety and ethical barriers to challenging humans with live virulent mycobacteria have thus far limited the development of an *in vivo* challenge model for TB.

Already licensed for use in humans, BCG represents a potential surrogate for *Mtb* challenge, and is a safe replicating mycobacterium that causes a contained, short-term infection in immunocompetent individuals. A BCG challenge model has recently been described in which participants were challenged with intradermal (ID) BCG. Skin biopsies of the challenge site were taken 2 weeks later. BCG load was quantified by culture and quantitative polymerase chain reaction (qPCR) (Harris et al., 2014; Minassian et al., 2012). The model demonstrated the ability to detect differences in anti-mycobacterial immunity induced by BCG and MVA85A (modified vaccinia Ankara 85A, a new-generation vaccine against tuberculosis) vaccination, with a significant inverse correlation between immune signatures, particularly IFN-γ and IL-17 pathways, and BCG load detected by qPCR (Harris et al., 2014). A dose escalation study and comparison of BCG SSI and BCG TICE has also been reported (Minhinick et al., 2016).

One criticism of intradermal challenge is that it does not mimic the natural route of infection, and to that end a clinical trial evaluating the safety and feasibility of an aerosol BCG human challenge model is currently ongoing (NCT02709278).

2.2 In vitro assays

The development of vaccines against other pathogens has been greatly expedited by the identification of a biomarker or immune correlate of protection (Thakur et al., 2012). Such indicators, for example antibody titer or cytokine level, may
be measured using an in vitro assay, allowing the use of human blood or cell samples. Frustratingly, there are currently no validated correlates to reliably assess the efficacy of candidate TB vaccines. Most TB vaccine studies to date have used quantification of antigen-specific IFN-γ by ELISpot and/or intracellular cytokine staining as the primary immunological read-out, though it remains unclear whether this measure correlates with protection (Elias et al., 2005; Mittrücker et al., 2007). Although one study of BCG vaccinated infants in South Africa found no difference in frequency and extended cytokine profiles of Mtb specific cells between protected and non-protected infants (Kagina et al., 2010), a more recent trial indicated an association between the BCG antigen-specific IFN-γ ELISpot response and reduced risk of TB disease (Fletcher et al., 2016). This latter study also found a negative correlation between levels of Mtb antigen-specific IgG and risk of disease, suggesting that protective immunity may not be restricted to the T cell compartment (Fletcher et al., 2016).

An alternative to measuring predefined individual parameters is the use of mycobacterial growth inhibition assays (MGIAs), which take into account a range of immune mechanisms and their additive effects and interactions. These systems measure the ability of human or animal cells to inhibit growth of mycobacteria following in vitro infection. Using samples taken pre- and post-vaccination, functional efficacy may be assessed without the requirement for in vivo Mtb challenge or natural infection in animals. Several such MGIAs have successfully discriminated BCG vaccinated from non-vaccinated human volunteers using both whole blood or peripheral blood mononuclear cells (PBMC) (Cheng et al., 1988; Cheon et al., 2002; Fletcher et al., 2013; Hoft et al., 2002; Kampmann et al., 2004; Worku and Hoft, 2000).

Animal models provide an opportunity to test novel vaccine candidates, and an in vitro assay using blood or cells from vaccinated animals offers a potential surrogate of protective efficacy that may negate the need for in vivo challenge during early selection of vaccine candidates. MGIAs have been described using cells from mice (Cowley and Elkins, 2003; Kolibab et al., 2009; Marsay et al., 2013; Parra et al., 2009; Sada-Ovalle et al., 2008), cattle (Carpenter et al., 1997; Denis et al., 2004) and NHPs (Tanner et al., submitted). Importantly, both Parra et al. (2009) and Marsay et al. (2013) showed that differences in mycobacterial growth inhibition between groups were consistent with different levels of protection in experimentally-matched mice challenged in vivo, thus demonstrating the utility of animal MGIAs for biological validation. Preliminary work applying a whole blood MGIAs in cynomolgus macaques has demonstrated a correlation between mycobacterial growth inhibition following vaccination and protection from BCG challenge as measured by lymph node CFU (own unpublished data).

MGIAs also allow efficacy against different strains (including hypervirulent strains) to be tested in parallel in cells from the same animal, rather than limiting to one laboratory strain, which may be unrepresentative of clinical strains affecting humans, as for in vivo challenge.

2.3 In silico modelling

The availability of genome sequences for Mtb and other mycobacterial species, mice and humans together with relatively recent developments in computer algorithms have facilitated the use of in silico bioinformatics methods for the identification of new TB vaccine candidates.

Comparative analyses of mycobacterial genomes have allowed the identification of 16 genomic regions of Mtb which are absent in one or more strains of BCG, known as regions of difference (RD) (Behr et al., 1999; de Jonge et al., 2005). RD proteins have been generated using recombinant methods or overlapping synthetic peptides (Mustafá, 2005) followed by testing in immune assays to identify those suitable for vaccine development. Wang et al. analyzed RD proteins in silico for their ability to bind to a range of HLA class I alleles and showed that a significant proportion were high-affinity binders, representing promising epitopes for inclusion in experimental TB vaccine candidates (Wang et al., 2010).

In 2011, a study by Tang et al. used novel computational search tools to identify new Mtb antigens activating polyfunctional CD8+ T cells which were then validated in human-based assays (Tang et al., 2011). A further study scanned multiple published databases of Mtb gene expression to select the proteins most highly expressed in all phases of infection. The proteins were evaluated for the presence of B and T cell promiscuous epitopes and population coverage in terms of allelle presentation. Sequence alignments were then used to determine identical epitopes on M. smegmatis, and two M. smegmatis-derived experimental vaccines were tested in mice to assess humoral immunogenicity and cross-reactivity with Mtb (Rodriguez et al., 2011).

More recently, Monterrubio-López et al. (2015) identified potential vaccine targets using NERVE (New Enhanced Reverse Vaccinology Environment) prediction analysis of the Mtb H37Rv proteome. Proteins were further down-selected based on VaxJen-predicted antigenicity and amino acid sequence alignments, with 6 novel candidates finally selected. Bowman et al. described the incorporation of the machine learning approach support vector machine (SVM) classification, which resulted in superior accuracy in discriminating protective antigens from non-antigens (Bowman et al., 2011). In addition to the 3Rs benefits, reverse vaccinology offers several advantages over conventional methods including speed, reduced cost and ability to identify all the putative protective antigens rather than just the most abundant (Bowman et al., 2011).

2.4 Less sentient animals

As opposed to absolute replacement of animal models with in vitro or inanimate systems, another 3Rs approach involves the replacement of more sentient vertebrates with animals thought to have a lower potential for pain perception. The amoeba Dictyostelium discoideum and the fruit fly Drosophila melanogaster, though useful in understanding host-pathogen interactions and innate immune responses during mycobacterial infection, have limited applicability for the study of vaccines due to their lack of adaptive immunity (Dionne et al., 2003; Hagedorn et
al., 2007, 2009). Zebrasfish, however, have an immune system similar to that of humans with a fully developed adaptive arm in adults, and represent a popular model organism for various pathogens (Meijer and Spaink, 2011). It has been suggested that the course of mycobacterial infection in zebrafish has some parallels to that of human TB, with high-dose infection leading to progressive disease resembling acute TB, and low-dose infection leading to spontaneous latency with reactivation following immunosuppression (Parikka et al., 2012; Swaim et al., 2006). Importantly, many virulence factors, host genes and immune cell types involved in human Mtb pathogenesis have conserved functions in the zebrafish-M. marinum model (Cronan and Tobin, 2014). Zebrafish have already proven useful in elucidating the early events of a mycobacterial infection, the role of the innate immune system in resistance and understanding the mechanisms of granuloma formation and its role in controlling infection, with the limitation that zebrafish do not have lungs (Clay et al., 2007; Cronan and Tobin, 2014; Swaim et al., 2006). It is also a potentially promising model for aiding the development of TB therapeutics and vaccines for preventing reactivation of latent TB. In a study by Oksanen et al., both BCG and a DNA-based vaccine protected fish from mycobacterial infection, reducing mortality and bacterial burden following infection with a lethal dose (Oksanen et al., 2013).

3 Reduction

Reduction is defined as “methods that minimise the number of animals used per experiment or study, either by enabling researchers to obtain comparable levels of information from fewer animals, or to obtain more information from the same number of animals, thereby avoiding further animal use”\(^1\). Examples include reducing replication by increased data sharing, improved experimental design and technologies enabling longitudinal studies in the same animals.

3.1 Reducing replication

Publication bias arises when negative or non-confirmatory findings are suppressed, either by researchers themselves choosing not to submit for publication or lack of journal acceptance (Song et al., 2013). Emphasis is frequently placed on impact rather than quality of research or reproducibility, leading to what has been described as a “crisis of false positives” in biomedical research where many published results are false or exaggerated, with an estimated 85% of resources wasted (Macleod et al., 2014). Importantly, failure to share negative findings results in needless repetition of animal experiments. One potential solution is the prospective registration of preclinical studies similar to that of clinical human trials such as the BMJ AllTrials campaign, aiming for “all trials registered, all results reported”\(^3\). In recent years, some measures have been taken to encourage the publication of negative findings by provision of such repositories as the BioMedicine Journal of Negative Results\(^4\) and a recent PLOS ONE collection of negative, null and inconclusive results. Furthermore, BMC Research Notes was produced with the specific objective of publishing repeat studies and negative results\(^5\).

3.2 Experimental design

The NC3Rs state that “appropriate experimental design and statistical analysis techniques are key means of minimising the use of animals in research”\(^6\). One critical consideration is sample size, which should not be so large as to use an unnecessary number of animals. However, under-powering an experiment with too few animals to provide a biologically meaningful result is equally wasteful (Festing and Altman, 2002). Rigorous statistical calculation such as power analysis should be performed to identify an appropriate sample size; methods for this have been described by Dell et al. (2002). However, Williams et al. (2009) highlighted the issues involved in powering TB challenge experiments based on survival using guinea pigs and larger animals. The authors note that while survival studies can be extremely informative in establishing that a new candidate vaccine can confer equivalent protection to BCG and that this protection is sustained, demonstrating significant improvement over BCG is more challenging. Due to the binomial nature of survival data, statistical power is extremely low, and a simulation exercise revealed that for a substantial increase in mean survival time (from 150 to 250 days), a prohibitively large group size of 74 would be required to reach a significant p-value of 0.05 (Williams et al., 2009). This provides further support for the use of predefined fixed endpoints, as alternative measures such as bacterial load in target organs offer superior statistical and discriminative power (Williams et al., 2009).

Minimizing variation (for example by controlling for confounding variables such as age, weight and genetics) also improves power, allowing the same effect to be detectable with a smaller number of animals (Festing and Altman, 2002). Experiments should be unbiased with random allocation of animals to treatment groups and blinding of researchers, and it is encouraging that many of these factors are now taken into consideration when granting ethical permission to conduct animal trials. Furthermore, multiple questions may be answered, and therefore numbers of experiments and animals reduced, by applying adequately powered factorial designs (Festing and Altman, 2002). The NC3Rs recently launched an online Experimental Design Assistant (EDA) to guide researchers in the design of experiments and ensure the minimum number of animals is used to achieve the scientific objectives\(^1\).

3.3 In vivo imaging

In vivo imaging techniques enable longitudinal studies of the same animals through the course of infection, reducing the number of groups required for assessment at sequential timepoints and therefore variation. As opposed to CFU quantifica-

---

3 http://www.alltrials.net/
4 http://jnrbm.biomedcentral.com/
5 http://bmcresearchnotes.biomedcentral.com/
6
tion in the lungs, which necessitates euthanasia, bioluminescent or fluorescently-tagged mycobacteria can be tracked in live animals for real-time assessment of vaccine efficacy (Zelmer et al., 2012). Such non-invasive techniques also fall into the “refinement” category.

Zhang et al. used an autoluminescent strain of *Mtb* as a surrogate marker to replace CFU counts. Relative light units (RLU) *in vivo* paralleled CFU counts *in vitro* during the active phase of bacterial growth, and the ability of a recombinant BCG vaccine to limit bacterial growth was demonstrated. Although the modest sensitivity of the system necessitates a greater bacterial burden, leading to more widespread dissemination of infection, the authors suggest methods by which the degree of luminescence may be improved (Zhang et al., 2012).

*In vivo* imaging may also be used to visualize clinical symptoms of TB disease, again reducing the need for endpoint measures such as CFU. Lewinsohn et al. (2006) described the use of computed tomography (CT) scanning in macaques, demonstrating a strong correlation with pathohistologic findings at necropsy.

CT and MRI scanning have since been applied in a number of NHP TB drug and vaccine studies to determine number, structure and distribution of pulmonary lesions across the lung nodes following *Mtb* challenge (Lin et al., 2013; Rayner et al., 2013; Sharpe et al., 2016). Importantly, whereas routine readouts such as bacterial burden and gross pathology necessitate the use of high doses of *Mtb*, sensitive imaging techniques permit much lower challenge doses of *Mtb*, as discussed in the following section (Rayner et al., 2013; Sharpe et al., 2016).

4 Refinement

Refinement refers to “methods that minimise the pain, suffering, distress or lasting harm that may be experienced by the animals”1. This applies to all aspects of animal use, including housing and husbandry. Wolfensohn et al. (2015) recently described a quantitative system for assessment of lifetime experience, assigning a combined welfare score (CWAS) to various permutations of the macaque model of TB. This system scored four parameters (physical, psychological, environmental and procedural [experimental and/or clinical]), which contributed to a level of combined severity. This measure will be discussed in the context of TB challenge models; improvements in more general areas such as importation and living conditions are beyond the scope of this review.

4.1 BCG challenge

Replacing virulent *Mtb* with attenuated BCG would not only reduce the severity of animal challenge experiments, but also offer the opportunity to conduct challenge experiments in humans as a tool for assessing and prioritizing candidate vaccines at an early stage of development as described.

As described above, a BCG challenge model for use in humans has recently been developed (Harris et al., 2014; Minassian et al., 2012; Minhinnick et al., 2016). BCG challenge in animal models similarly represents an alternative to the use of virulent *Mtb*, reducing the severity of pathology; with the caveat that reduced pathology makes for a less realistic and perhaps less sensitive model. The impact on lifetime experience that could be achieved through the use of a BCG (rather than *Mtb*) challenge model in macaques was evaluated through comparison of the CWAS during the post-challenge phase, which was shown to be considerably greater (i.e., improved welfare) at 2 weeks post-BCG challenge compared with 16, 26 or 52 weeks post-*Mtb* challenge (Wolfensohn et al., 2015). Using a BCG challenge model in cynomolgus macaques, significantly lower levels of BCG were detected in the axillary lymph nodes draining the site of challenge in BCG-vaccinated compared with naïve animals. Furthermore, higher *ex vivo* PPD-specific IFN-γ ELISpot responses and enhanced *in vitro* mycobacterial growth inhibition were associated with lower CFU counts in the draining lymph node, suggesting utility of this model in identifying correlates of immunity (own unpublished data). A similar model was employed in cattle, demonstrating that BCG vaccinated animals had lower BCG CFU counts than naïve animals following intranodal challenge with BCG (Villarreal-Ramos et al., 2014). It has been shown that BCG-vaccinated mice later challenged with intradermal BCG had reduced mycobacterial growth, and this protection was predictive of BCG efficacy against aerosol *Mtb* challenge (Minassian et al., 2011).

4.2 Ultra-low-dose challenge

High doses of *Mtb* are typically required to induce meaningful changes in clinical parameters and pathology that permit the measurement of vaccine efficacy. However, a lower challenge dose would not only more closely resemble natural infection, but also reduce disease burden and therefore symptoms. Whereas NHP challenge models typically use inoculum sizes of between 50 and 3000 CFU of *Mtb* (Langermans et al., 2001; Lewinsohn et al., 2006; Lin et al., 2012; Verreck et al., 2009), an ultra-low dose challenge recently described by Sharpe et al. (2016) exposed macaques to less than 10 CFU. Macaques did not exhibit abnormal behaviors or marked clinical signs, unlike with normal high dose challenge. Furthermore, comparison of the CWAS score for unvaccinated animals during the first 16 weeks after challenge with high and low dose *Mtb* shows that there was a beneficial effect on welfare of using a reduced dose (Wolfensohn et al., 2015). However, such a model does require more sensitive approaches to evaluate disease burden such as FHG PET-CT *in vivo* imaging, though these offer the added advantage of serial assessment as described above. Using these methods, the authors were able to discriminate between rhesus and cynomolgus macaques in terms of disease burden and progression, reflecting previously described differences in disease...
outcome (Sharpe et al., 2016). Concerns that ultra-low doses would lead to increased variability or fail to reliably infect all of the challenged animals did not appear to be founded (Sharpe et al., 2016).

4.3 Humane end-points

The NC3Rs defines humane end-points as “clear, predictable and irreversible criteria which substitute for more severe experimental outcomes such as advanced pathology or death”1. Unfortunately, such criteria often remain poorly defined, and the long duration typical of animal experiments involving TB and other chronic progressive infections provides greater potential for ambiguity. If no measures are taken to treat \textit{Mtb}-infected mice, they will succumb to infection and die before their average life-span (Medina and North, 1998), and in a systematic review of endpoints implemented in 80 murine TB studies published in 2009, 47% of the studies were classified as “lethal” (not terminated before animals reached advanced stages of disease, which would rapidly progress towards spontaneous death if no other endpoints were applied). 66% of these were categorized at the highest severity level, meaning that animals were allowed to die spontaneously or reach a moribund state (Franco et al., 2012).

In addition to the welfare concern, survival may not necessarily be the most controlled or statistically powerful measure, as described above. In a TB vaccine study of long-term survival in rhesus macaques, lung lesion burden using MR imaging and stereology, but not survival time, was able to distinguish naïve and vaccinated NHPs (Sharpe et al., 2010). Such “in-life” imaging may represent a more humane readout of vaccine efficacy than survival.

In most TB candidate vaccine studies reported in recent years, animals were euthanized at a fixed time-point following \textit{Mtb} challenge and alternative measures of disease severity such as CFU counts in lungs and spleen were assessed (Gillis et al., 2014; Stylianou et al., 2015). Various measures have been described as a cut-off parameter for euthanasia including non-transient hypothermia and, more commonly, change in body weight. However, although weight itself is quantitative and objective, this measure is confounded by natural fluctuations and the definition of the outcome (Sharpe et al., 2016). Concerns that ultra-low doses would lead to increased variability or fail to reliably infect all of the challenged animals did not appear to be founded (Sharpe et al., 2016).

<table>
<thead>
<tr>
<th>Tab. 1: Summary of 3Rs approaches to tuberculosis vaccine research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3Rs area</strong></td>
</tr>
<tr>
<td><strong>Replacement</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Reduction</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
upper boundary varies considerably across studies, ranging from 10 to 30% weight loss (Franco et al., 2012; Williams et al., 2009). As described, survival studies remain informative in certain circumstances (for example in demonstrating sustained protection by a vaccine candidate), but even then “death” is now defined as the time at which moribund animals are humanely euthanized (Franco et al., 2012).

Wolfensohn et al. (2015) compared CWAS scores to quantify differences in lifetime experience when a survival endpoint (52-week post-challenge follow-up) was used as opposed to a fixed end-point (16 or 26 weeks post-challenge) for the evaluation of TB vaccine efficacy. There was a considerable reduction in welfare “cost” with decreasing time post-challenge. Although alternative read-outs such as bacterial load usually support survival data, further work is required to develop more accurate predictors of death or survival; indeed parameters such as blood glucose homeostasis have been shown to correlate with degree of severity in other disease models (Yu et al., 2011).

Alternatively, McShane and Williams suggest that the field may consider a movement from reduction of disease to prevention of Mtb infection as the endpoint of animal challenge experiments (McShane and Williams, 2014). Though larger numbers of animals would be required, such a model would eliminate many of the issues associated with TB disease end-points and better align with the definition of vaccine efficacy in humans.

5 Conclusion

Animal models have been, and continue to be, invaluable in the TB vaccine field. As well as enhancing our understanding of TB disease and immunology, they are essential for assessing immunogenicity, efficacy and safety of vaccine candidates prior to advancement into clinical trials. UK legislation dictates that research must be performed in accordance with the 3Rs framework to ensure ethical rigor. We have reviewed examples from the TB vaccine field of replacing, reducing and refining animal experiments, summarized in Table 1. Replacements include the use of humans for BCG challenge experiments, in vitro assays such as IFN-γ ELISpot and MGIAs, in silico techniques and the use of less sentient animals. The number of animals used may be reduced by avoiding data replication, improving experimental design, and technologies such as in vivo imaging enabling longitudinal studies in the same animals. Refinements specific to TB challenge models include the use of lower doses of Mtb or less pathogenic mycobacterial strains, and clearly-defined humane fixed end-points. Recent technological advances have opened the door to alternatives and improvements which must now be taken forward to ensure impact. The approaches described not only advance animal welfare, but also demonstrate the ability of projects with a 3Rs justification to drive forward the science in the development of improved TB vaccines.

References


Denis, M., Wedlock, D. N. and Buddle, B. M. (2004). Ability of T cell subsets and their soluble mediators to modulate the replication of Mycobacterium bovis in bovine macro--


Festing, M. F. and Altman, D. G. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J* 43, 244-258. https://doi.org/10.1093/ilar.43.4.244


**Conflict of interest**

The authors have no conflicts of interest or copyright issues to declare.

**Acknowledgements**

We are grateful to UFAW (Universities Federation for Animal Welfare) for a grant supporting this work. HMcS is a Wellcome Trust Senior Clinical Research Fellow.

**Correspondence to**

Rachel Tanner, PhD
The Jenner Institute
Nuffield Department of Clinical Medicine
University of Oxford
Oxford, UK
Phone: +44 1865 617617
e-mail: rachel.tanner@ndm.ox.ac.uk