Published in final edited form as:

Tuberculosis (Edinb). 2012 May ; 92(3): 283–288. doi:10.1016/j.tube.2011.12.003.

BCG: Myths, realities, and the need for alternative vaccine strategies

Helen McShane^{a,*}, William R Jacobs^b, Paul E Fine^c, Steven G Reed^d, David N. McMurray^e, Marcel Behr^f, Ann Williams^g, and Ian M Orme^{h,*}

^aJenner Institute, Nuffield Department of Medicine, University of Oxford, UK

^bDept of Microbiology & Immunology, Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, New York, USA

^cDepartment of Immunology and Infection, London School of Hygiene and Tropical Medicine, UK

^dInfectious Disease Research Institute, Seattle, Washington USA

^eDepartment of Microbial and Molecular Pathogenesis, Texas A&M Health Science Center, College Station, Texas USA

^fMcGill University Health Centre, Montreal Canada

^gHealth Protection Agency, Porton Down, Salisbury, UK

^hDept of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins Colorado USA

This document is a synopsis of a workshop held on August 2nd, 2011, at the Bill and Melinda Gates Foundation in Seattle, WA. The purpose of the meeting was to re-evaluate exactly what BCG does, what it doesn't do, and perhaps more importantly, what we think it does but it doesn't. The basic conclusions arising from the meeting reflected the fact that many questions can still be raised about BCG, and it is important for the field to fully understand the strengths and limitations of BCG to ensure that lessons are learned in the design and evaluation of new vaccine strategies.

1. BCG

The high percentage of global coverage [from the 1990's, there has been about 85% coverage globally], and protection conferred against disseminated disease, mean that in any development of new vaccine strategies, either in terms of modifications or additions to BCG, its use as part of a prime boost strategy, or the implications of replacing BCG with an alternative, requires consideration of BCG itself. The workshop briefly reviewed the history

*Meeting co-Chairs

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of BCG from inception to the current circumstances of its use today, and the various implications that this knowledge base has for new tuberculosis vaccines.

It was recognized that some level of protection is manifestly possible under some circumstances, and that considerable previous experience has accumulated from earlier BCG trials. However, many questions remain, including the issue of the heterogeneity of BCG efficacy and whether the effectiveness of any new vaccine approach will show similar or analogous heterogeneity in efficacy. Despite many years of study, the mechanism of action of BCG is not fully understood, particularly with respect to the induction of memory immunity and the durability of that response and the various mechanisms by the host and pathogen that can interfere with its potential efficacy. The workshop also noted the importance and complexity of defining “what exactly” it is one wants to protect against [infection, disease, pulmonary, non-pulmonary, primary, reinfection, reactivation...], issues that have plagued discussions of the effectiveness of BCG in the past and which have obvious implications for designing and evaluating new vaccines in the future. There is no doubt that BCG has some protective effect against severe forms of extra-pulmonary disease in children, estimated in a 2006 meta-analysis at 73% for meningitis, and 77% for military disease. However, despite the fact that the reduction of meningeal tuberculosis in children is often emphasized, the reality is that prevention of pulmonary tuberculosis, particularly in adults, who are responsible for most transmission, must be the primary goal of an improved vaccination regimen. Any new vaccine must be proven to be “better” than BCG, by a combination of efficacy in relevant animal models and safety. Any improved vaccination regimen must also be affordable, particularly in TB endemic countries.

Heterogeneity of BCG

It has been recognized for decades that BCG has tremendous variability in efficacy, and this is coupled with a publication bias towards positive results; for instance, a study from a given country may report “BCG decreases general morbidity”, but nothing gets reported if this not seen in a separate country. In both controlled trials and observational trials the efficacy of BCG has varied widely, from excellent protection in British schoolchildren and Native American Indians, to studies in which protection has been negligible, such as in Georgia, Chingleput, Colombia and Malawi. The variability also helps established dogmas, for instance the general opinion that “BCG does not protect adults from pulmonary tuberculosis” currently pervades, but in certain trials in the past one can argue that vaccination was indeed protective [including in infants]. Similarly, while certain analyses put the durability of protective effect at around 10-years with BCG, others suggest it can in some circumstances / populations be much longer (e.g. in Native American Indians).

There are many variables within the clinical trials which may account for some of the variability in efficacy, for example study design, the strain and formulation of BCG used, host genetics, the prevalence of environmental mycobacteria, the fitness/virulence of the *M. tuberculosis* strains prevalent in the trial site area at the time of the trial, the nutritional state of vaccinees including their Vitamin-D status, exposure to other pathogens such as helminths, impact of other vaccinations, and so on. The impact of many of these variables remains poorly understood and currently we have insufficient evidence to design new

vaccine strategies which avoid being impacted by the same variables. The influence of the strain of BCG and exposure to environmental mycobacteria are discussed further below.

BCG strain variation

BCG has undergone considerable evolution, and cannot be regarded as simply a generic organism/vaccine; in fact, it is a family of genetically similar strains. The workshop discussed the evolution of BCG since the 1920's, in which the organism has become increasingly attenuated. Much of Europe has and does use BCG Danish, but other strains (e.g. Russia, Poland, Romania, and Pasteur) have also been used. In various animal models, the Danish and Pasteur strains are most widely used, but other strains have been used (e.g. Japan, Glaxo, Prague). There is no consensus on which strain to use in preclinical animal models. In head-to-head preclinical animal studies comparing the efficacy of different strains and substrains against virulent *M. tuberculosis* challenge, it has been difficult to identify clear differences in protective efficacy between them. Furthermore, the same strain of BCG has been found efficacious in some clinical trials and not others. It is possible that more subtle differences could be detected in preclinical models, using protocols that are sufficiently powered to discern those differences. However, the conclusion based on the evidence to date is that the loss of genes in different BCG strains does not confer any difference in protective efficacy.

Exposure to environmental mycobacteria

The effects of exposure to environmental mycobacteria on both interfering with the efficacy of BCG and on conferring any degree of protection against *M. tuberculosis* is still far from being fully understood. This issue originally arose from observations that the protection elicited by BCG was highest in infants and in individuals with no prior tuberculin sensitivity, has been sporadically addressed, both in animal models and in clinical studies. However, there is no real consensus as to the mechanisms that may be involved. Studies on the effects of exposure to environmental mycobacteria [EM] on BCG efficacy have provided two possible mechanisms, which indeed may well be complementary. In the first, the protective effects of BCG are blocked, as prior immunity induced by exposure to EM inhibits BCG replication. In the second, the immunity induced by exposure to EM cannot be enhanced by BCG vaccination, and thus the actual relative contribution of BCG induced immunity is masked. It is not clear whether in humans immunity induced by EM confers any significant degree of protection, nor whether the type of cellular immunity induced by EM is comparable to that induced by BCG. A further issue is that EM is not a single discrete entity, but a complex collection of different strains of EM, which vary according to geographical location. Furthermore the route, frequency and nature (killed or live bacteria) of human exposure is uncertain. This hugely complex issue is extremely difficult to model in animals. Further work in animals and humans is essential to understand whether such interference has relevance to the new vaccine strategies currently being developed.

2. The immune response to BCG

The workshop discussed the long held concept that BCG simply establishes memory immunity, and this is all that is needed protect the individual over the long term against

exposure to *M.tuberculosis*. It was felt however that there still is a fundamental lack of knowledge regarding the actual mechanism of action of BCG at the immunological level. The actual demonstration [in mice] that BCG generated memory immune CD4 T cells was made nearly 25-years ago. There is now considerable emphasis on detecting “multifunctional T cells” in vaccine models, but the point was made that [a] finding T cells making more than one cytokine is hardly surprising, [b] the kinetics of secretion of these cytokines is not actually parallel, and [c] which of the subsets based on relative secretion are actually critical and which are not remains completely unknown. Moreover, it was agreed that there is no solid data at all as to why the protective window of BCG clearly wanes in some but not all populations [there are multiple potential reasons], nor is there any data on the stability/longevity of the induced memory T cell subsets, or what happens to these subsets beyond the point of initial exposure to *M. tuberculosis* [something which, in endemic areas, will happen repeatedly]. In addition, we have little or no information from endemic areas in areas such as mixed infections, and secondary infections.

New data was presented at the workshop discussing the now established knowledge that there are two broad subsets of memory T cells, both at the CD4 and CD8 level. These consist of effector memory T cells, which BCG generates in large numbers in the lungs, and central memory T cells, that BCG appears to induce in only relatively low numbers. Whether the failure of BCG to generate any appreciable levels of central memory has any relevance to protection is still very unclear, but this issue further emphasizes the lack of understanding regarding the immunological action of BCG. Likewise, very little is known about the durability of the memory T cell response induced by BCG. Data was presented from a secondary infection model [mice infected then cured with chemotherapy] that showed that a very potent memory response was induced by a secondary infection, but this then decayed, possibly because of T cell exhaustion. Whether such events occur in humans as well is unknown.

At the clinical level, measurement of the immune response is necessary to assess vaccine “take”, and to measure responses believed to be important for, and/or correlated with, subsequent protection. In the latter case, the field has considered single measures of cytokines, as well as more complex cytokine profiles [so called polyfunctional T cells], and is currently reviewing the use of the more functional growth inhibition assays. However, to date, there is evidence both for and against all of these parameters as being important in protective efficacy, in both preclinical and clinical studies. Many of these aspects are known to be important in protection [activated CD4 T cells, IFN γ , TNF, etc] but it is unlikely that any single, simple measure of cellular immune function will be useful as a direct correlate of protection. It is possible that an immune correlate will be vaccine-specific and may also be disease-stage specific, for instance, a correlate in mycobacteria-naïve, HIV-uninfected babies may be very different to one then observed in HIV-infected *M.tuberculosis*- infected adults.

The durability of immunity induced by BCG is poorly understood. Estimates range from ~10-12-years in certain analyses, but others [in Brazil and in Alaskan natives] have suggested this is far longer.

One factor not addressed in such studies relates to the longevity of the memory T cells induced by BCG, and whether they are subject to different rates of attrition in different geographical areas [e.g. prevalence of tuberculosis, low population density, level of exposure to environmental mycobacteria]. Theoretical models describe the amplification of protective immunity and potential boosting of this by further vaccination. In several trials with new generation BCG booster vaccines, some durability of cellular immunity has been demonstrated. However the relevance of this for protective immunity is unclear and can only be resolved by efficacy trials. Again further studies in animals and humans addressing durability of immunity in different settings including TB endemic countries, is needed.

3. The use of animal models

Animal models to evaluate safety, immunogenicity and efficacy are critical to the development of an effective TB vaccine for human use. The workshop discussed the format of challenge models and the data derived from these models which has helped shaped our views on the mechanisms of action of BCG.

In these animal models, BCG is used as the “gold standard” positive control and the study design “favors” this by providing a clear and measurable difference between saline and BCG. Under these conditions any new vaccine has to be better than BCG. However there are significant limitations with this approach. The immunity induced by the new vaccine regimen may be different in form, kinetics, and effectiveness when compared directly to the immune response to BCG. In this regard, it was noted that of 204 vaccines tested at HPA, only seven regimens were demonstrated to be better than BCG. Whether or not any of these regimens would be better in humans, where BCG works less well and less reproducibly, is unclear. The idea was discussed, therefore, whether alternative models should be used where BCG is only transiently or completely ineffective? One such possibility could to simply use one of the well-characterized W-Beijing strains [such as HN878 or SA161] in which the effect of BCG was recently shown to be transient at best.

The potential protective effect of BCG has been shown in multiple animal models where it slows but does *not* prevent disease, and the workshop discussed the limitations to the currently used study designs which attempt to demonstrate the potency of new vaccine strategies. To extend this point further, much of our screening emphasizes the importance of CD4 TH1 responses. While this is obviously a major factor, the role of other cells, such as CD8, NKT, and TH17 cells as examples are probably not still being adequately addressed, plus one can argue that the “standard assays” favor the detection of the CD4 response and hence may be missing the input of other cells if these have different kinetics.

Read-outs of efficacy and statistical power

As animal screening models have evolved and have been subjected to more stringency it has become clearer that relevant indicators of disease burden are difficult to measure in a statistically robust manner and Study designs are frequently used that do not have the statistical power to distinguish between anything other than the positive and negative control groups. Retrospective analysis of the experiments performed in the guinea pig model showed that analysis of survival curves using the log-rank test has the power to demonstrate

statistically significant improved protection over BCG *only* if the test vaccine gives close to 100% survival [over a time-frame where BCG gives circa 60%] or if the numbers of animals used per group is ~ 200. The former scenario is unlikely, the latter is impractical. Analysis of bacterial load or disease pathology as secondary endpoints provides no further information because each of these measures is fundamentally related to survival. Measurement of bacterial load at fixed time-points, is a more robust read-out in terms of statistical power, relatively small differences in can be reliably detected using group sizes that are practically feasible. However, the reduction in bacterial load is rarely substantial, and the relevance of a 0.5 log₁₀ reduction in bacterial load is frequently questioned. While “sterilizing immunity” remains a “Holy Grail”, a vaccine that could substantially drive the bacterial load downwards would be a major advance. Bacterial load measurements provide a means of ranking or down-selecting candidate vaccines which is much more cost-effective and reliable than survival studies and if combined with appropriate immunization schedules can indicate the potential of a vaccine to protect against long-term disease. In the mouse model, the difference in bacterial load between the unvaccinated and BCG groups becomes smaller over time and some vaccines have been shown to have a sustained protective effect relative to BCG at these late time points (e.g. rBCG-Hly, H56). Investigation into the mechanism underlying these differences in protection kinetics would be valuable information which could have practical implications for vaccine design. In guinea pigs, where disease progresses more rapidly compared to mice, the bacterial load at a fixed time-point is more discriminatory at early time points and it may not be possible to identify vaccines which are able to direct the immune system against the more chronic or ‘latent’ stages of disease. This emphasizes that we should avoid ‘tailoring’ the animal models to individual vaccines but also that we should not expect any one animal model to provide efficacy data for all vaccine types.

Study designs to demonstrate improvements over BCG

This question is difficult to answer because the field has to decide what is actually meant. It is *feasible* to demonstrate that a vaccine can offer marginal improvement compared to a BCG-induced effect that is already quite potent, but in reality this is observed infrequently and may not be relevant to clinical settings where the effect of BCG is much less than in the animal models. The idea was discussed, therefore, whether alternative models should be used where BCG is only transiently or completely ineffective? This idea is attractive in terms of the statistical power of the studies – if the difference between the BCG and test vaccine groups is increased, the group size needed to reliably demonstrate this becomes lower. One approach that can be employed is to prolong the vaccination-to-challenge intervals. In mice the potency of BCG reduces with longer vaccination to challenge intervals and the protective effect can be restored with a booster vaccine. A similar effect can be observed in guinea pigs but intervals of around 11 months are needed before the potency of BCG is reduced. These study designs have obvious disadvantages in terms of resources and cost but the additional expense would be warranted if the outcome is a robust demonstration of a test vaccine which provides protection where BCG fails. For example, out of 26 studies comparing BCG with BCG plus a boosting vaccine, only 2 showed evidence of improved protection, with 22 showing equivalent protection to BCG. The risk is that candidates may be halted in their development because of a failure to demonstrate a boosting effect or

equally, that candidates progress without clear evidence of protective efficacy greater than BCG alone.

Advanced animal models

Non-human primates are phylogenetically similar to humans and therefore may be expected to provide data which is most likely to predict the outcome of clinical trials. Selection of the most appropriate measures of ‘protection’ is even more critical in these species because failure to robustly quantify a vaccine-induced effect [either positive or negative] has serious ethical and financial consequences. Read-outs which are commonly used in smaller animal species can be impractical in macaques and the lessons learned from mice and guinea pigs about the power of survival studies should be noted by those performing similar, time-to-event studies in macaques. Bacterial load in organs is technically difficult to measure in larger animals and measurement of gross pathology is the most commonly used primary read-out of disease burden. This is a semi-quantitative measure which is difficult to analyze statistically and to compare between different investigators. Advanced imaging of organs or whole animals using combinations of CT and PET scanning, combined with stereology software to objectively measure the extent of disease is a very powerful means of improving pathology analysis and offers great promise in refining this read-out, particularly if technology permits such analysis ‘in-life’. The greatest advantage of macaque studies over any other animal model is the ability to perform longitudinal measures using robust assays that are the same as those performed in humans. This opens the possibility to evaluate candidate biomarkers of disease and, crucially for the TB vaccine field to identify and then validate putative correlates of protection. Studies conducted in parallel, involving vaccination of both humans and macaques with the same vaccine and identical immune measures, but with subsequent challenge of the macaques with *M.tuberculosis* could provide a wealth of information about the relationship between vaccine-induced immune responses and the outcome of infection which are directly translatable to humans. Despite the power and potentially fruitful outcome of such studies, the current literature suggests that they are rarely conducted and the full potential of these models clearly remains to be realized.

Use of clinical isolates in vaccine candidate testing

The workshop raised the issue of virulence of the newly emerging clinical *M. tuberculosis* isolates, and the impact that this may have on pre-clinical screening. Data was presented which demonstrated that many outbreak strains exhibited high levels of virulence and severe levels of lung damage in the relevant guinea pig low dose aerosol model, far in excess of that seen using the laboratory strains. These strains seem to exert their virulence in part through the induction of FoxP3+ regulatory T cells, and further work is needed to evaluate the mechanism of action of these strains. In the mouse models, BCG protects far less well against these highly virulent isolates, and protection is often only transient at best. Work to evaluate the efficacy of the new vaccine regimens being developed against these strains is essential. The question then arose as to whether BCG protects against such clinical isolates to the same extent as the laboratory strains H37Rv and Erdman usually used to screen vaccine candidates. Studies in mice intravenously immunized with BCG and then challenged with high doses of various strains by the same route showed that W-Beijing strains might be more virulent than another family [Canetti] and the vaccinated mice were far less resistant to

the W-Beijing strain compared to H37Rv. A range of BCG efficacy has been observed following aerosol challenge with Beijing and non-Beijing strains, both in terms of protection and kinetics and it has been demonstrated that an early (30 day CFU) protective effect of BCG against two high virulence Beijing strains rapidly waned. Loss of effector immunity in this study was shown to be associated with the emergence of regulatory T cells. The impact of these findings to clinical trials of new vaccines clearly needs to be considered and further studies are needed to understand the mechanisms by which these strains cause severe disease and appear refractory to the immune response induced by BCG. The few studies that have been conducted in animal models have only focused on a few examples, and preferentially those within the W-Beijing family. The recommendation from the group was that if we are going to understand the overall epidemic better, then we need to compare multiple families of outbreaks, as well as strains from different geographic areas [“W-Beijing strains are very troublesome in South Africa but cause no problems in Montreal”, as an example underlying this].

A further complication arising from observations in the field reflects the fact that relapse of infection occurring in individuals previously treated with chemotherapy is now often associated with a different strain. Moreover, cases are increasingly being seen where the infection is mixed. In both cases, we do not understand the T cell response [amplified, contracted, exhausted?], or the influence of vaccination. And yet, such events could have very serious outcomes on the spread of tuberculosis. As one example, if an individual harbors both a low fitness but MDR/XDR strain of *M.tuberculosis* plus a drug-sensitive high fitness/high virulence strain, will the damage caused by the latter [necrosis, regulatory T cells] promote the spread of the former?

When animal modeling is taken as a whole it provides a critical overall profile of safety, immunogenicity and efficacy. The mouse model gives information about immunogenicity and some insight into mechanisms of action, the guinea pig is more stringent, discriminative, and exhibits similar necrotic pathology to that in humans, whereas non-human primate models give more detailed and relevant immunogenicity data, as well as safety data. The majority of the new vaccine candidates have been evaluated in these models and the published data suggests that some of these may be protective in humans. For those that have advanced to human efficacy trials, it will be possible to judge whether the animal data was predictive for humans. However, improved animal models are required to improve predictions and our ability to down-select which vaccines should progress in clinical trials and efficacy trials. We need improved models of BCG boosting with adequate vaccine to challenge intervals, validated models to test vaccines targeting latent infection, ways to screen unconventional antigens, innovative new delivery systems. To aid these developments, the field needs improved visibility of ‘negative’ or commercially sensitive data which tends not to reach the public domain.

4. Implications for new vaccines

Despite substantial progress by the field in the past decade, to date we do not know if any of the current strategies will prove efficacious in humans and there is a need to continue to investigate alternative vaccines and alternative vaccine strategies, such as mucosal boosting;

further development of novel adjuvants, and live attenuated mycobacterial strains other than BCG. As a working model it is proposed that the vaccine should be able to induce long lived memory T cells, with both effector memory and central memory phenotypes represented. After exposure, the memory T cell population should remain stable with little attrition or exhaustion. Protection should remain potent despite extreme vaccination to challenge intervals. The immunity established should be free of the effects of interference by EM, and react fast enough to nullify the effects of regulatory T cell emergence. Multiple laboratories should be able to demonstrate efficacy [and potential superiority to BCG] against a wide range of geographically representative high virulence clinical isolates.

5. Immune correlates and mechanisms of protection

An understanding of the mechanisms underlying protection and which aspects of immunity correlate with protection would be a major advance for the field and one which iteratively would greatly facilitate vaccine development and refinement. All animal models should be used to identify potential correlates, and all clinical efficacy trials should store samples from all participants for immune correlate analysis once the study is unblinded.

6. Current trials

One question that arises is whether BCG Danish 1331 given at birth actually protects against TB? BCG 1331 used in the Chingleput RCT had *no* efficacy, and the rest of the available information for 1331 comes from observational data. For this reason, is it appropriate to now question the dogma that efficacy trials with new TB vaccines/regimens cannot have a saline control group? Under controlled conditions, with good health-care professionals, and access to drugs, should we now contemplate a saline or no vaccine control group to determine whether BCG is truly protective in the population being tested? This would then mean any efficacy trial had 3 arms: no vaccine, BCG alone and new vaccine, but such information would be invaluable to our understanding of the efficacy of the current strains of BCG in clinical use in the 21st Century.

There are currently fourteen new vaccines which have been evaluated in early clinical trials and two currently being evaluated in efficacy trials. However, current tools for down-selection, i.e. determining which vaccines should enter clinical testing and, more importantly, which should enter efficacy testing are far from perfect. As noted above, there is no immunological correlate with which to evaluate these vaccines prior to efficacy testing, and moreover the debate continues as to the extent to which efficacy in the current animal models predicts efficacy in humans. Currently, expensive and time consuming efficacy trials are the only way to evaluate the efficacy in humans of a new tuberculosis vaccine.

It is clear that there are limited resources available to fund and to do such efficacy trials. There is an urgent need for a better understanding of the vaccine induced immune response in humans, and biomarkers of protection that could be productively used to rationally down select vaccine candidates. In parallel, a better understanding of which animal models best predict efficacy in humans is needed.

A human challenge model, as is used for malaria and influenza vaccine development, may be invaluable to the field, but such a model, such as the BCG challenge model currently under evaluation, needs validating against *M.tuberculosis* challenge in a relevant animal model (cattle and NHPs are under consideration). Ultimately, such a model will need validating against human Phase IIb/III efficacy trial data.

7. Primary Recommendations of the workshop

Basic and pre-clinical studies

- Before a new vaccine candidate is tested at a clinical trial site, it needs to be shown in at least two animal models systems, and by independent laboratories with no vested interest, to be superior to BCG alone in sufficiently powered short and/or long term assay protocols against both a laboratory and [as a bare minimum] at least one clinical isolate of *M.tuberculosis* that is prevalent within the trial site region.
- The field should make a concerted effort to pool data sets [published and unpublished] in order to objectively evaluate the statistical power and value of read-outs, and come to a consensus that the study designs being used are sufficient to demonstrate a measurable effect.
- Reach a consensus on a set of ‘recognized’ read-outs for each model [following similar principles as clinical trials] which should be applied as a minimum in all studies. This will improve our ability to compare data between sites and between vaccines, and provide an evidence base to assist in the development of the ‘down-selection’ criteria which are currently being proposed by product development partnerships [TBVI and Aeras]. This in turn would enable development of well-characterized models with clearly defined parameters and limitations that will allow vaccine developers and the animal modelers to select the most appropriate route of pre-clinical testing, thus in turn ensuring that maximum value is gained from animal studies
- Where possible, every effort should be made to conduct parallel studies in humans and animals in order to increase the relevance of the animal model data. Consideration should be given as to whether more relevant ways of assessing vaccine impact are feasible in experimental settings; for example, study designs which demonstrate that vaccination can prevent spread of disease, or prevent infection by natural transmission.
- The animal model programs should be encouraged to break the current ground rules and try to develop innovative new study protocols that allow recognition of vaccines whose mechanisms/kinetics differ to BCG, and could be detected under conditions where BCG fails.
- Further studies should be encouraged to understand the immunological mechanisms underlying the masking and blocking hypotheses with regard to the effect of environmental mycobacteria on TB vaccine efficacy. It is important to

determine if sensitivity to these organisms will influence the efficacy of the new vaccines.

- The field should further carefully examine the whole issue of “homologous” versus “heterologous” protection and the degree to which this could influence the effort to develop *M. tuberculosis*-based vaccines. This should include continued exploration of the use of attenuated mutants of *M. tuberculosis*, such as double auxotrophs, and those with immune enhancing mutations such as phoP, secA2, as well as new candidates that appear to induce central memory T cells responses such as IKE-plus. Such studies should also determine if these candidates could replace BCG in prime boost regimens.
- The field should develop new models to reflect new events in the field, specifically the influence of mixed infections [where the drug resistance and the fitness of the two strains may differ], and the host response to secondary exposure.
- The field should continue to evaluate new vaccine candidates in models of BCG priming, and emphasize complementing rather than replacing good quality BCG products. Furthermore, the field still completely lacks any attempt to standardize BCG to achieve homogeneous responses, including manufacturing guidelines.

Clinical issues

- The field should continue to identify and validate new correlates of protection. The lack of any clear correlate of protection is a major limitation for vaccine development. The focus to date has been primarily on cytokine responses [blood levels or following in vitro stimulation with antigen] and more recently in terms of gene signatures. Some advances are being made in terms of metabolomics, but this is still in its infancy.
- The field should attempt to re-examine data from clinical trials which seem to show long-lived protection [e.g., the Aronson trial] to see if hypotheses for the basis for persistent protection can be derived.
- The field needs to discuss the provocative idea that new trials should include a control arm in which BCG is withheld.
- Continue to pursue recent gene expression studies which can potentially identify different populations of subjects with gene expression profiles associated with disease susceptibility, in order to understand more about mechanisms underlying disease susceptibility and understand more about population heterogeneity.
- The field should try to further investigate the degree to which BCG can protect against infection. An endpoint of infection (as defined by QGIT or TSPot) should be part of all new TB vaccine efficacy trials as the incidence of infection is always higher than the incidence of disease, and this is potentially a very important outcome measure.