SYSTEMATIC REVIEW

Practical considerations for a TB controlled human infection model (TB-CHIM); the case for TB-CHIM in Africa, a systematic review of the literature and report of 2 workshop discussions in UK and Malawi [version 2; peer review: 2 approved, 1 approved with reservations]

Stephen B. Gordon, Simon Sichone, Anthony E. Chirwa, Phoebe Hazenberg, Zacharia Kafuko, Daniela M. Ferreira, JoAnne Flynn, Sarah Fortune, Shobana Balasingam, Giancarlo A. Biagini, Helen McShane, Henry C Mwandumba, Kondwani Jambo, Keertan Dedha, Nimisha Raj Sharma, Brian D. Robertson, Naomi F Walker, Ben Morton, TB Controlled Human Infection Model Development Group

Abstract
Background: Tuberculosis (TB) remains a major challenge in many domains including diagnosis, pathogenesis, prevention, treatment, drug resistance and long-term protection of the public health by vaccination. A controlled human infection model (CHIM) could potentially facilitate breakthroughs in each of these domains but has so far been considered impossible owing to technical and safety concerns.
Methods: A systematic review of mycobacterial human challenge
studies was carried out to evaluate progress to date, best possible ways forward and challenges to be overcome. We searched MEDLINE (1946 to current) and CINAHL (1984 to current) databases; and Google Scholar to search citations in selected manuscripts. The final search was conducted 3rd February 2022. Inclusion criteria: adults ≥18 years old; administration of live mycobacteria; and interventional trials or cohort studies with immune and/or microbiological endpoints. Exclusion criteria: animal studies; studies with no primary data; no administration of live mycobacteria; retrospective cohort studies; case-series; and case-reports. Relevant tools (Cochrane Collaboration for RCTs and Newcastle-Ottawa Scale for non-randomised studies) were used to assess risk of bias and present a narrative synthesis of our findings.

**Results:** The search identified 1,388 titles for review; of these 90 were reviewed for inclusion; and 27 were included. Of these, 15 were randomised controlled trials and 12 were prospective cohort studies. We focussed on administration route, challenge agent and dose administered for data extraction. Overall, BCG studies including fluorescent BCG show the most immediate utility, and genetically modified *Mycobacteria tuberculosis* is the most tantalising prospect of discovery breakthrough.

**Conclusions:** The TB-CHIM development group met in 2019 and 2022 to consider the results of the systematic review, to hear presentations from many of the senior authors whose work had been reviewed and to consider best ways forward. This paper reports both the systematic review and the deliberations.

**Registration:** PROSPERO (CRD42022302785; 21 January 2022).

**Keywords**
Human infection studies, tuberculosis, TB, BCG, bacille Calmette Guerin, vaccine, immunology, genetically modified organism
Corresponding authors: Stephen B. Gordon (sgordon@mlw.mw), Ben Morton (ben.morton@lstmed.ac.uk)

Author roles: Gordon SB: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Sichone S: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing; Chirwa AE: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Hazenberg P: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing; Kafuko Z: Conceptualization, Investigation, Resources, Supervision, Writing – Review & Editing; Ferreira DM: Conceptualization, Funding Acquisition, Investigation, Resources, Supervision, Validation, Writing – Review & Editing; Flynn J: Conceptualization, Investigation, Supervision, Validation, Writing – Review & Editing; Fortune S: Conceptualization, Investigation, Supervision, Validation, Writing – Review & Editing; Biagini GA: Conceptualization, Investigation, Methodology, Resources, Validation, Writing – Review & Editing; McShane H: Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Writing – Review & Editing; Mwandumba HC: Conceptualization, Investigation, Validation, Writing – Review & Editing; Jambo K: Conceptualization, Investigation, Validation, Writing – Review & Editing; Dedha K: Conceptualization, Investigation, Supervision, Writing – Review & Editing; Raj Sharma N: Conceptualization, Investigation, Validation, Writing – Review & Editing; Robertson BD: Conceptualization, Investigation, Methodology, Resources, Writing – Review & Editing; Walker NF: Conceptualization, Investigation, Methodology, Writing – Review & Editing; Morton B: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by Wellcome [211433, https://doi.org/10.35802/211433; to SG]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Gordon SB et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gordon SB, Sichone S, Chirwa AE et al. Practical considerations for a TB controlled human infection model (TB-CHIM); the case for TB-CHIM in Africa, a systematic review of the literature and report of 2 workshop discussions in UK and Malawi [version 2; peer review: 2 approved, 1 approved with reservations] Wellcome Open Research 2023, 8:71
https://doi.org/10.12688/wellcomeopenres.18767.2

First published: 10 Feb 2023, 8:71 https://doi.org/10.12688/wellcomeopenres.18767.1
Introduction

Need for a tuberculosis (TB) vaccine and challenges in development

More than 1.5 million people died of tuberculosis (TB) in 2021, of whom 214,000 were individuals living with HIV. An estimated 9.9 million people worldwide became ill with TB, with the greatest burden of this disease in Africa and Southeast Asia, and with 47% of TB-affected households suffering catastrophic costs. The COVID pandemic has reversed years of progress in global targets to end TB, making the search for an effective vaccine even more urgent.

The Bacille Calmette-Guérin (BCG) vaccine was the first, and is still a century later, the only vaccine approved for humans that protects against TB. BCG is an attenuated Mycobacterium bovis strain with considerable sequence homology with M. tuberculosis. BCG is used both as a vaccine and as immunotherapy.

BCG vaccine efficacy is highly variable in adults (0–80%) being most effective at high latitudes and particularly poor in tropical and subtropical regions.

There are several new TB vaccines in development as listed in Table 1, targeting initial infection, disease, or drug resistant infections. Vaccine trial design is very challenging because initial infection is common in early life. This sensitisation is often partially immunising but not always sterilising, leaving both reinfecion or recrudescent disease a risk for some years after initial infection and making vaccine trial clinical endpoints very difficult to define. In a recent vaccine trial, the M72/AS01E candidate vaccine demonstrated 49.7% efficacy against microbiologically proven pulmonary TB disease at 3 year follow up in a phase 2b clinical trial of more than 3,500 BCG-vaccinated subjects recruited with evidence of prior TB sensitisation (positive interferon gamma release assay (IGRA)). Now, a very large study is required for full evaluation of the clinical impact for this vaccine and the cost has been estimated at GBP 400 million.

Given the wide choice of potential vaccines, uncertainty about vaccine trial endpoints, and the huge cost of vaccine trials against disease, precise tools are required to select the most promising new vaccine candidates for clinical evaluation. As it is very difficult to detect viable organisms in subjects with TB, an alternative vaccine trial design might be to rely on immunological endpoints demonstrating infection—but these endpoints may

<table>
<thead>
<tr>
<th>Latest Phase</th>
<th>Vaccine candidate</th>
<th>Vaccine platform</th>
<th>Target population</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Ad5 Ag85A</td>
<td>Viral vector</td>
<td>Booster vaccine for adults with latent TB (NCT02337270)</td>
</tr>
<tr>
<td>IIa</td>
<td>AEC/BC02</td>
<td>Mycobacterial – whole cell or extract</td>
<td>Booster vaccine for adults with latent TB (NCT05284812)</td>
</tr>
<tr>
<td></td>
<td>Anhui Zhifei Longcom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTBVAC</td>
<td>Biofabri, TBVI</td>
<td>Mycobacterial – live</td>
<td>BCG replacement vaccine for infants (NCT03536117), with a Phase III planned (NCT04975178)</td>
</tr>
<tr>
<td></td>
<td>University of Zaragoza</td>
<td></td>
<td>Adults with latent TB (NCT02933281)</td>
</tr>
<tr>
<td>ID93 + GLA-SE</td>
<td>IDRI, Wellcome Trust</td>
<td>Protein/adjuvant</td>
<td>Booster vaccine for adults with latent TB (NCT02465216)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BCG vaccinated healthcare workers (NCT03806686)</td>
</tr>
<tr>
<td>TB/FLU-04L</td>
<td>RIBSP</td>
<td>Viral vector</td>
<td>Booster vaccine for adults with latent TB (NCT02501421)</td>
</tr>
<tr>
<td>GamTBvac</td>
<td>MoH, Russian Federation</td>
<td>Protein/adjuvant</td>
<td>BCG vaccinated adults (NCT03878004), Phase III planned (NCT04975737)</td>
</tr>
</tbody>
</table>

be similar in patients experiencing sterilising immunity or even subclinical disease. Controlled human infection models (CHIM), also called Human Infection Studies (HIS) might offer a further alternative to down-select vaccine candidates prior to phase 2b/3 efficacy trials. In CHIM trials, vaccines are tested for protection against experimental infection managed in volunteer subjects. An optimised TB-CHIM would help with vaccine selection, and identification of immune correlates. It would have potential to predict the efficacy of new drug regimens, and even offer discovery opportunities regarding pathogenesis and correlates of immunity.

Potential for a TB-CHIM

CHIM have been successfully employed for many decades to accelerate vaccine development in enteric infections, and to select appropriate clinical trials in vaccines against other infections. Very recently, CHIM studies of SARS-CoV-2 have contributed substantially to knowledge of pathogenesis, infectivity, diagnostic precision of lateral flow tests and post-exposure protection against disease. Notable examples of recent vaccines that are being rolled out after clinical development accelerated by CHIM are the RTS,S/AS01 E malaria vaccine and the typhoid conjugate vaccine.

There are significant challenges, however, for a Mycobacterium tuberculosis (M. tb) or TB CHIM (TB-CHIM). First, and most importantly, is the current concern for safety using wild-type M. tb as the infection cannot be reliably eradicated from an infected subject; treatment is prolonged (6 months) and toxic with potential for serious adverse events; and there is also...
potential risk of immunopathology for which treatment options may be inadequate or complicated. There is a small risk of recurrent infection (~12% cases) after treatment, highest in the first year and subsequently receding. There are reasons for optimism, however, that a conditionally replicating wild-type mycobacterium, including genetically inserted suicide switches to ensure complete sterility post-infection could become available and transform the study of human TB infection. In the meantime, there are sufficient studies of BCG to show that this organism induces anti-mycobacterial immunity and could serve as a model of wild-type infection, useful for vaccine testing provided that the vaccine mechanism was not unique to M. tb. Further, there are modified BCG models, including those with included fluorophores, that will allow diverse endpoint measurements to alleviate the current difficulty in precise microbiological detection in the site of infection and remote sites. Opinions and limitations of the BCG models tested are discussed later in this review.

Non-human primate models of virulent M. tb disease are critical in vaccine research and pathogenesis discovery. The macaque is susceptible to severe TB and intravenous BCG has been shown to be protective, with protection associated with greatly increased numbers of antigen specific CD4 T cells in the lung and bronchoalveolar lavage. Repeated limiting (low) dose (RLD) infection models have increased the sensitivity of this model with less severe disease that has allowed comparison of protective effects from less effective vaccines. The RLD approach may be of great benefit in developing safe human challenge models, and the responses observed in NHP will inform future human TB-CHIM design and aims. Ultimately, for a vaccine or therapeutic to reach licensure, early phase testing in humans is required. An optimised TB-CHIM design, informed by work in non-human primates has the potential to rapidly accelerate product development.

Need for a TB-CHIM in Africa
There is a reasonable and strong drive to deliver CHIMs in infection endemic settings so that vaccines, drugs and therapeutics are directly tailored to the populations that need them most. This drive is particularly marked in infections where prior exposure and the intensity of community infection define the immunological context in which novel vaccines would be used. In addition, there is political momentum to ensure that the delivery of these needed vaccines to low and middle-income countries (LMIC) is not hindered by economic or regulatory factors arising in the country of invention. Recently, there has been considerable progress in delivering CHIM studies in Africa—examples include falciparum malaria, pneumococcus, N. lactamica and schistosomiasis.

Exposure to both M. tb and non-tuberculous mycobacteria (NTM) are very common in Africa, often occurring at a very young age. Recurrent TB exposure occurs throughout life owing to late diagnosis of infectious cases who circulate in the community. Exposures that increase susceptibility to mycobacterial infection are common. These include but are not limited to malnutrition, smoke exposure, HIV, and potentially, other infections. In Africa, a TB-CHIM could be transformative for testing of vaccines, early-stage drug efficacy studies and scientific discovery as the CHIM would recruit relevant populations with prior mycobacterial experience. Further, this is important strategically to facilitate endemic sites with tools to take a leading position in the clinical development of novel products for TB. As with other CHIM studies, a TB-CHIM would use a consistent infection in well-characterized subjects that could be quantitatively monitored over time, dramatically reducing the number of experimental subjects needed in any study as well as allowing much shorter trials.

We reviewed the existing published TB-CHIM options by systematic review and conducted two workshop discussions—one in Europe and one in Africa—to determine if and how a TB-CHIM in Malawi could be developed. We have included a reflexivity statement describing how we have promoted equity in our research partnership (see Extended data). The review and workshop discussions are reported here.

Systematic review of human challenge studies using M. tuberculosis

Methods
All attendees at the workshops have been included as authors (including as part of the collaborative group) of the manuscript; all have reviewed the manuscript and given explicit consent for their inclusion. Our PROSPERO registered (CRD42022302785; 21 January 2022) systematic review is reported using the PRISMA 2020 checklist (see Extended data) and synthesis without meta-analysis (SwIM) reporting guidelines.

Eligibility criteria
Inclusion criteria: adult humans ≥18 years of age; administration of live Mycobacterium tuberculosis with either wild-type, or genetically modified organism (GMO); Bacillus Calmette-Guerin (BCG) interventional trials or prospective cohort studies with immune and/or microbiological end points.

Exclusion criteria: animal studies; publications with no primary data; interventional studies with no administration of live bacillus (e.g., viral vector vaccination trials or purified protein derivative challenge studies); chemotherapy studies in patients with cancer; and case-series, case reports and retrospective cohort studies. The studies were grouped for narrative synthesis according to the study methodology (randomised controlled trial (RCT) versus non-randomised designs); challenge agent (GMO organism and BCG); administration route (intradermal, lung and oral); reporting of adverse and serious adverse events; confirmation of infection (classical culture, molecular diagnostic, and no confirmation); and immune response measurement (diverse methodologies reported so narrative synthesis applied).

Information sources
MEDLINE (RRID:SCR_002185) (1946 to current) and EBSCO CINAHL (RRID:SCR_022707) (1984 to current) databases were systematically searched. Table 2 describes the search terms and the search strategy. Following application of inclusion and exclusion criteria and removal of duplicates, we used Google Scholar (RRID:SCR_008878) to search citations in selected manuscripts (see CONSORT diagram, Figure 1). The final search was conducted on 3rd February 2022.
Selection process
After removal of duplicates, two authors independently assessed article titles, abstracts, and full manuscripts in turn to make study selection decisions. BM assessed all article titles, abstracts, and full manuscripts with secondary independent review by AC, SS and PH. Discrepancies were resolved by a third independent reviewer (DF) where these occurred. In addition, SG read all selected papers.

Data collection
The data extraction tool was piloted by BM, AC, SS and PH independently to ensure uniformity of data collection and to refine outcome variables. Subsequently, two authors independently reviewed each manuscript and collected data. Authors then met to finalize data collection and resolve any discrepancies. No automated tools were used. The completed raw data collection tool is included in Table S1 as Extended data. Collected data items included: study design; setting (country); participant characteristics (number and basic demographics); challenge agent and dose; administration route; adverse events; follow up period; infection confirmation (site and method) and immunological response measurement assays. We applied relevant risk of bias assessment (Cochrane Collaboration for RCTs and Newcastle-Ottawa Scale for non-randomised studies) tools for included studies after data extraction. Risk of bias assessment was conducted by two independent reviewers with discrepancies resolved as previously. Given the heterogenous nature of the studies with disparate primary and secondary outcomes, we...
focused predominantly on methodological aspects of controlled human infection model delivery.

Synthesis methods
We present a narrative synthesis based on study characteristics to describe how TB-CHIM studies have previously been delivered. Meta-analysis has not been conducted due to the heterogeneity of studies, study inclusion criteria and our stated aim to focus on methodological aspects of TB-CHIM delivery.

Results
Study selection. A total of 27 studies met inclusion criteria (Figure 1). These included 15 RCTs\(^{16-30}\); one of which was a trial protocol\(^{30}\); and 12 non-randomised interventional studies\(^{17,51-61}\). Other articles of potentially of relevance, but without primary data, did not meet eligibility criteria\(^{62-71}\). Only three studies were conducted in high-burden settings as shown in Table 2 and Table 3.

Study characteristics. Table 3 (RCTs, one (18 participants) from South Africa) and Table 4 (non-randomised studies, one from Brazil (6 participants) and one from South Africa (106 participants)) describe characteristics for studies selected. Primary outcome measures for the studies were diverse and as our focus was on CHIM methodology, we focused on administration route, challenge agent and dose administered in data extraction. Figure 2 describes risk of bias assessments for selected studies. Three RCTs had a low risk of bias across all assessment domains\(^{16,41,44}\) and three non-interventional studies had an overall low risk of bias\(^{17,52,55}\).

We focused on several aspects of CHIM study design to inform our workshops discussions. These were participant selection in mycobacterial challenge studies, challenge agent administration route and sampling post-challenge, challenge agent strains and doses for intradermal and mucosal routes, microbiological confirmation and immunological assays to confirm infection as well as reported adverse events.

### Table 3. Study characteristics for included randomised controlled trials.

*Trial protocol, results not published at time of writing. Interventional group receive escalating doses of BCG (1 x 10^5 maximum, McShane personal communication) and intradermal saline; control group receive aerosolised saline and intradermal BCG SSI at a dose of 1 x 10^5 CFU. X: not reported; a: mean; b: median; ID: intradermal; PO: per oral; BCG: Bacillus Calmette-Guerin; MTBVAC: live attenuated strain Mycobacterium tuberculosis; SSI: Statens Serum Institut; AERAS-422: recombinant BCG with overexpression antigens Ag85A, Ag85B, and Rv3407 and expressing mutant perfringolysin; GMO: genetically modified organism; VPM1002: recombinant BCG expressing listeriolysin, lacking urease C gene and containing a hygromycin resistance marker; rBCG30: recombinant BCG overexpressing antigen Ag85b; TICE: US brand name for intravesical BCG; CFU: colony forming units; TB: tuberculosis.

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Country</th>
<th>n</th>
<th>Males</th>
<th>Average Age</th>
<th>Route</th>
<th>Intervention</th>
<th>Control</th>
<th>Follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoft</td>
<td>1999</td>
<td>US</td>
<td>66</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (Connaught &amp; TICE)</td>
<td>Saline</td>
<td>90</td>
</tr>
<tr>
<td>Hoft</td>
<td>1999</td>
<td>US</td>
<td>48</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (Connaught &amp; TICE)</td>
<td>NA</td>
<td>56</td>
</tr>
<tr>
<td>Hoft</td>
<td>2000</td>
<td>US</td>
<td>48</td>
<td>X</td>
<td>X</td>
<td>PO &amp; ID</td>
<td>BCG (Connaught)</td>
<td>PBS</td>
<td>365</td>
</tr>
<tr>
<td>Hoft</td>
<td>2008</td>
<td>US</td>
<td>35</td>
<td>13</td>
<td>29</td>
<td>ID</td>
<td>rBCG30 (GMO BCG)</td>
<td>BCG (TICE)</td>
<td>252</td>
</tr>
<tr>
<td>Wardhana</td>
<td>2011</td>
<td>Indonesia</td>
<td>34</td>
<td>8</td>
<td>65</td>
<td>ID</td>
<td>BCG (Pasteur)</td>
<td>BCG solvent</td>
<td>90</td>
</tr>
<tr>
<td>Grode</td>
<td>2013</td>
<td>Germany</td>
<td>80</td>
<td>80</td>
<td>33(^{a})</td>
<td>ID</td>
<td>VPM1002 (GMO BCG)</td>
<td>BCG (SSI)</td>
<td>180</td>
</tr>
<tr>
<td>Harris</td>
<td>2014</td>
<td>UK</td>
<td>49</td>
<td>21</td>
<td>23(^{b})</td>
<td>ID</td>
<td>BGC (SSI)</td>
<td>NA</td>
<td>42</td>
</tr>
<tr>
<td>Leentjens</td>
<td>2015</td>
<td>Netherlands</td>
<td>40</td>
<td>40</td>
<td>21(^{b})</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>Saline</td>
<td>28</td>
</tr>
<tr>
<td>Spertini</td>
<td>2015</td>
<td>Switzerland</td>
<td>36</td>
<td>14</td>
<td>27(^{a})</td>
<td>ID</td>
<td>MTBVAC (GMO wild TB)</td>
<td>BCG (SSI)</td>
<td>365</td>
</tr>
<tr>
<td>Hoft</td>
<td>2016</td>
<td>US</td>
<td>24</td>
<td>14</td>
<td>29(^{a})</td>
<td>ID</td>
<td>AERAS-422 (GMO BCG)</td>
<td>BCG (TICE)</td>
<td>182</td>
</tr>
<tr>
<td>Blazevic</td>
<td>2017</td>
<td>US</td>
<td>86</td>
<td>X</td>
<td>X</td>
<td>PO &amp; ID</td>
<td>BCG (SSI &amp; Connaught)</td>
<td>NA</td>
<td>180</td>
</tr>
<tr>
<td>Arts</td>
<td>2018</td>
<td>Netherlands</td>
<td>30</td>
<td>30</td>
<td>X</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>BCG solvent</td>
<td>118</td>
</tr>
<tr>
<td>Tameris</td>
<td>2019</td>
<td>RSA</td>
<td>18</td>
<td>3</td>
<td>29(^{a})</td>
<td>ID</td>
<td>MTBVAC (GMO wild TB)</td>
<td>BCG (SSI)</td>
<td>180</td>
</tr>
<tr>
<td>Giamarellos-Bourboulis</td>
<td>2020</td>
<td>Greece</td>
<td>150</td>
<td>67</td>
<td>80(^{a})</td>
<td>ID</td>
<td>BCG (Bulgaria)</td>
<td>Saline</td>
<td>365</td>
</tr>
<tr>
<td>TBO41 Trial</td>
<td>2020</td>
<td>UK</td>
<td>60</td>
<td>X</td>
<td>X</td>
<td>Aerosol</td>
<td>BCG (SSI and Bulgaria)</td>
<td>Saline</td>
<td>168</td>
</tr>
</tbody>
</table>
Table 4. Study characteristics for included non-randomised studies. *Matsumiya used samples taken from the same participants as Harris et al.16. X: not reported; a: mean; b: median; ID: intradermal; Bronch: bronchoscopic installation; PO: per oral; Neb: nebulised; BCG: Bacillus Calmette-Guerin; SSI: Statens Serum Institut; TICE: US brand name for intravesical BCG; CFU: colony forming units.

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Country</th>
<th>n</th>
<th>Males</th>
<th>Average age</th>
<th>Route</th>
<th>Intervention</th>
<th>Follow up (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosenthal</td>
<td>1968</td>
<td>US</td>
<td>43</td>
<td>X</td>
<td>X</td>
<td>Neb</td>
<td>X</td>
<td>90</td>
</tr>
<tr>
<td>Ravn</td>
<td>1997</td>
<td>Denmark</td>
<td>20</td>
<td>12</td>
<td>22x</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>365</td>
</tr>
<tr>
<td>Hoft</td>
<td>1999</td>
<td>US</td>
<td>69</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (Connaught &amp; TICE)</td>
<td>1095</td>
</tr>
<tr>
<td>Monteiro-Maia</td>
<td>2006</td>
<td>Brazil</td>
<td>6</td>
<td>1</td>
<td>35x</td>
<td>PO</td>
<td>BCG (Moreau)</td>
<td>180</td>
</tr>
<tr>
<td>Schreiber</td>
<td>2010</td>
<td>UK</td>
<td>7</td>
<td>5</td>
<td>27x</td>
<td>PO</td>
<td>BCG (Moreau)</td>
<td>365</td>
</tr>
<tr>
<td>Minassian</td>
<td>2012</td>
<td>UK</td>
<td>40</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>168</td>
</tr>
<tr>
<td>Matsumiya</td>
<td>2015</td>
<td>UK</td>
<td>24</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>14</td>
</tr>
<tr>
<td>Boer</td>
<td>2015</td>
<td>Netherlands</td>
<td>12</td>
<td>6</td>
<td>24x</td>
<td>ID</td>
<td>BCG (SSI)</td>
<td>378</td>
</tr>
<tr>
<td>Minhinnick</td>
<td>2016</td>
<td>UK</td>
<td>41</td>
<td>13</td>
<td>27x</td>
<td>ID</td>
<td>BCG (SSI &amp; TICE)</td>
<td>28</td>
</tr>
<tr>
<td>Blazevic</td>
<td>2017</td>
<td>US</td>
<td>5</td>
<td>X</td>
<td>X</td>
<td>ID</td>
<td>BCG (TICE)</td>
<td>85</td>
</tr>
<tr>
<td>Davids</td>
<td>2019</td>
<td>South Africa</td>
<td>106</td>
<td>43</td>
<td>28x</td>
<td>Bronch</td>
<td>BCG (SSI)</td>
<td>180</td>
</tr>
<tr>
<td>Koeken</td>
<td>2020</td>
<td>Netherlands</td>
<td>22</td>
<td>22</td>
<td>X</td>
<td>ID</td>
<td>BCG (Bulgaria)</td>
<td>90</td>
</tr>
</tbody>
</table>

Figure 2. Risk of bias assessments for included studies. Cochrane collaboration risk of bias tool used for randomised trials. C1: random sequence generation (selection bias); C2: allocation concealment (selection bias); C3: blinding of participants and researchers (performance bias); C4: blinding of outcome detection (detection bias); C5: incomplete outcome data (attrition bias); C6: selective reporting (reporting bias); and C7: other bias. Newcastle-Ottawa Score used for non-randomised studies. D1: bias due to selection; D2: bias due to comparability; D3: bias due to outcome measurement; and overall score (0-3 = very high risk of bias; 4-6 = high risk of bias [some concerns]; and 7-9 = low risk of bias). Risk of bias assessment includes 26/27 studies. The excluded study is a trial protocol that has not yet reported results (https://clinicaltrials.gov/ct2/show/NCT02709278). Matsumiya 2015 included non-randomised participants reported in Harris 2013.
intradermal BCG on subsequent development of all cause lower respiratory tract infections. Prior BCG vaccination was used very variably in different studies, either as a specific exclusion criterion, specific inclusion criterion, or with an a priori aim to explore differences in response to mycobacterial challenge between BCG naïve and BCG exposed groups. Several studies did not report prior BCG vaccine status. In addition, studies also variably used dermal response to purified protein derivative and performed interferon-gamma release assays (IGRA) as surrogates of immune response when selecting participants. As detailed in Table 3 and Table 4, most studies were conducted in geographical regions with low TB prevalence. The selection of participants for TB CHIM studies in TB endemic areas (greater need for context-relevant therapeutics and vaccines) was more complex. In South Africa, Davids et al. categorized participants using clinical (history of treatment; presence of active disease; and household contact of index case), radiological (chest X-ray) and immunodiagnostics features (IGRA or QuantiFERON TB Gold-in-tube testing) to define “protective” and “susceptible” participant phenotypes.

**Challenge agent administration route and subsequent sampling.** Most included studies employed intradermal administration of BCG, modified BCG or modified *M. tb* and used peripheral blood derived endpoints as the most convenient route of both administration and sampling. There are a smaller number of important pulmonary studies to consider. For example, Davids et al. directly instilled BCG into a lung segment of participants with varying susceptibility phenotypes (e.g., exposed but uninfected versus history of recurrent TB), and compared BCG with PPD and saline instillation in alternative lung segments. That study demonstrated differential responses in bronchoalveolar lavage (BAL) immune cellular profile between the lung segments challenged and subsequently sampled. This study established the minimal BCG allowing the detection of responses and the safety of this model, the key advantage of which is interrogation of responses at the site of disease and a precise quantity of mycobacteria can be delivered to the alveolar compartment. A follow-on study to interrogate 6-month post-installation lung responses and to directly compare lung versus nasal versus skin installation is underway.

BCG administration by nebulised delivery is under investigation in a separate study but results are not yet reported. Nebulised BCG has been tested in a historical study in both adults and children but this study was assessed at high risk of bias and the described study methodology are unlikely to meet contemporary ethical committee criteria for approval. Oral BCG administration has been explored but in this review these studies were assessed as at risk of bias and did not isolate mycobacteria after challenge. Oral BCG challenge, however, does show evidence of being immunoregulatory of other responses, although evidence is lacking for enhanced pulmonary defence induced by mucosally-directed vaccination.

In skin studies, the theoretical environmental risk of non-participant population exposure with intradermal administration was systematically explored in studies of both wild-type BCG and genetically modified mycobacterial strains. Persistent shedding of viable mycobacteria from ulcerated intradermal injection sites was described but this was at low concentrations, with no recorded transmission such that this potential risk was concluded to be extremely low. Clearly skin biopsy is a valuable site for sampling in those studies, being safe and easily controlled. Reliable mycobacterial detection required large (≥4mm) punch biopsies which may limit the ability to conduct longitudinal sampling. Skin microbiopsy (<1mm) techniques could potentially be useful in this respect if diagnostic accuracy can be demonstrated.

The natural route of *M. tb* infection is inhalation of aerosolized bacteria via the respiratory tract therefore respiratory mucosal immune responses to intradermal mycobacterial challenge may not represent natural exposure, or elicit protective pulmonary responses. Alveolar macrophage responses to intradermal BCG administration but found that induced sputum (a less invasive technique than direct bronchoalveolar lavage sampling) was associated with cellular activation but the sampling technique likely introduced artefact, impairing their ability to infer immune response to challenge.

**Challenge agent strains and doses, for intradermal and mucosal routes.** Optimal conditions for intradermal controlled human infection with wild-type BCG have been explored, testing different strains and doses. The choice of BCG strain did not influence subsequent recovery of bacteria from skin punch biopsy after 14 days nor significantly alter immune responses within these studies. Higher BCG doses (6 x 10⁸ – 2.4 x 10⁹ CFU) were associated with improved bacterial detection from a day 14 skin punch biopsy with no measurable impact on adverse side effects. Early phase trials of intradermal recombinant BCG and attenuated *M. tb* (MTBVAC) vaccine candidates incorporated dose escalation within their study designs. Several recombinant BCG vaccine candidates have been compared with standard BCG. These include high dose (1 x 10⁶ ≥ CFU < 1 x 10⁷) AERAS-422 and high dose (5 x 10⁶ CFU) VPM1002. AERAS-422 included inserted genes to increase antigen expression and the perfringolysin gene to promote phagolysosome perforation and antigen presentation—see adverse reactions section. VPM1002 included insertion of the listeriolysin gene to promote phagolysosome perforation and antigen presentation and was found to be safe and immunogenic.

The attenuated *M. tb* strain MTBVAC (double deletion of independent virulence factors *phoP* and *fadB26*) was trialled at several doses. The highest dose used was 5 x 10⁶ CFU and was found to be safe and immunogenic (n=9) when compared to standard BCG in a healthy Swiss population. Subsequently, this dose was used in nine South African adults (TB endemic area) with one serious adverse event reported (graded as...
unlikely related to the study vaccine) and similar reactivity and immunogenicity compared to BCG. The infant data from this study are not reported as these fall outside of SR inclusion criteria.

For bronchoscopically instilled BCG, a lower dose of 1 x 10^4 CFU was used and found to induce significant changes to BAL cellular profiles, and antibody responses, which were not detectable below this dose. There was clear evidence of differentially expressed genes in BAL, and dysregulated proteins. Doses higher than 1 x 10^5 CFU were not explored within this study. Aerosolized BCG doses up to 1 x 10^7 CFU will be explored (McShane, personal communication) in a current trial, however the results are not yet reported. Orally administered BCG has been given in much higher doses within studies reporting dosage, with administration of 2 x 10^8 CFU in two studies and 1 x 10^8 CFU in a third. The potential for differential impact of stomach acid neutralization was not systematically explored within the included studies.

**Microbiological confirmation after challenge administration.** Microbiological confirmation of infection and clearance are critical in a CHIM. Microbiological recovery methods including sampling and processing techniques were evaluated in multiple studies within this review. Culture methods were the most widely used, supported by molecular techniques. Robust quantification of microbiological endpoints is essential for a controlled human infection model in order to detect differences between vaccinated and/or treated participants compared to placebo control participants within interventional trials.

Hoft et al., used serial skin punch biopsies at days 2, 7 and 14 after intradermal BCG administration with paired histopathological evaluation to examine granuloma formation. Mycobacteria were recovered from inoculation site culture in some samples but were more reliably detected by PCR (optimal time point at D14 with detection in 6/7 samples). Subsequently, a single punch biopsy at D14 has been used by the McShane group in recent studies. Optimisation of the punch biopsy technique is described by Minnich et al., who demonstrated that BCG could be recovered in all specimens (40/40) by both culture and qPCR with strong correlation between the techniques (r=0.664) and between BCG challenge dose and subsequent biopsy CFU count (r=0.749). Harris et al., demonstrated that qPCR estimated copy numbers were 1-2 logs higher (detected 48/48 samples) than cultures (detected 45/48 samples) from punch biopsy. In addition to D14 punch biopsy, Minassian et al., induced suction blisters of the BCG administration site but these were examined for immunological cellular infiltrates and microbiological results were not reported. This study demonstrated that BCG was detected in 28/28 punch biopsy specimens by qPCR and 19/28 by culture and that quantification by qPCR was a mean 1 log higher than by culture. Longitudinal swabbing of the intradermal injection site was systematically evaluated by Blazevic et al., who demonstrated that paired classical and qPCR microbiological quantification techniques demonstrated significant kinetic association. Another study, however, was not able to culture BCG by longitudinal swabbing of the injection site.

Opportunistic swabbing from ulcerated/discharging BCG intradermal injection sites was applied in an early phase trial with culture recovery (5/10 samples) for both interventional (MTBVAC, n=1) and control (BCG, n=4) mycobacteria.

In a bronchoscopic BCG installation study, Davids et al., were able to recover BCG in 6/54 (11%) of BAL microbiological samples.

**Measurement of immune response to mycobacterial challenge.** Early BCG studies were used to describe the now well recognized induction of interferon gamma responses by mycobacteria. There are still no validated immunological correlates of protection against *M. tb* infection and indeed most correlates fail in some subjects, particularly if dependent on a cellular response.

Exciting recent progress, however, has shown the potential of transcriptomics in defining a signature for active infection, and particularly single cell sequencing in defining immunological phenotype and linking it to clinical function. In the meantime, therefore, microbiological recovery and quantification is critically important as the primary endpoint of a TB-CHIM model.

In the studies reviewed, many different research approaches were taken to evaluate cellular and humoral responses to BCG. These include and are not limited to innate, humoral, and cellular responses but the focus has been on interferon gamma signatures at cytokine or cellular levels. The literature from non-human primate infection with *M. tb* and human studies of active TB are extensive and provide a robust background literature against which to check any human challenge findings. This will assist in validating any future TB-CHIM but are beyond the scope of this review.

**Adverse event reporting.** There were no standardised reporting criteria for adverse events after mycobacterial administration in these studies, but rather an adherence to the internationally accepted adverse event, severe adverse event, and adverse events of special interest (AE/SAE/AESI) definitions in use in clinical trials. Intradermal injection of BCG was associated with local, and expected side effects following routine vaccination. The maximum follow up period for AEs was 90 days. These AEs were reported in all included studies incorporating intradermal injection and no SAEs were reported in any of the identified studies. Hoft et al., reported that all BCG intradermal injection sites were healed by 90 days, which may therefore represent a pragmatic cut off point for reporting serious adverse drug reactions (SADR) in TB-CHIM studies. Davids et al., identified the minimal (safest) immunogenic dose (1 x 10^4 CFU) of bronchoscopically instilled BCG. The investigators found that AEs developed in 9.4% of participants (10/106), not usually associated with bronchoscopy, and with no significant difference between BCG, purified protein derivative (PPD) and normal saline. All adverse events were reported as mild and managed in an outpatient setting.

Serious adverse events were reported, however, in studies investigating candidate TB vaccines (MTBVAC and AERAS-422).
During a phase two trial in South Africa one participant was diagnosed with newly acquired HIV infection and aseptic meningitis three months after randomised allocation to MTBVAC vaccination\textsuperscript{54}. The participant was treated with empirical broad-spectrum antibiotics and anti-tuberculous therapy and was discharged from hospital five days after admission with no neurological sequelae. Six months of anti-tuberculous therapy were completed before commencement of anti-retroviral therapy with the participant reported as well with no detectable viral load at study completion by investigators. High dose AERAS-422 (n=8) was associated with varicella zoster virus reactivation in two participants\textsuperscript{41} resulting in discontinuation of the vaccine development programme. Investigators within this study identified immunological and transcriptomic correlations between TB immunity and varicella zoster virus infection but no definitive causative mechanism was identified.

**Summary of systematic review findings**

There are no data on controlled human infection studies using wild-type *M.tuberculosis*. There are considerable data on BCG human challenge, modified BCG and modified live *M.tuberculosis* vaccination both in carefully observed experimental studies and in vaccine trials. These data suggest that:

- BCG is a safe and acceptable experimental model of infection, using skin, pulmonary and oral routes.

  - The intradermal BCG model shows some immune responses that resemble those found in TB, but intradermal BCG can only be a model of infection and not pulmonary TB disease.

- Intrapulmonary BCG may be a more promising model in that the immune responses in lungs are more representative of TB, but recovery of inoculated mycobacteria in a quantifiable manner is problematic, making this model unlikely to be useful in vaccine testing if microbiological endpoints are required. However, an alternative readout could be a host biomarker-based one including post vaccinated responses associated with protection (e.g., alveolar polyfunctional T cells, lung resident T cells etc.). Another approach using this model could be to use alveolar lavage cells and blood post pulmonary vaccination in an *in vitro* killing model with live mycobacteria (serving as a proxy for vaccine efficacy). Thus, different vaccines given *via* the intrapulmonary route could be compared. An additional potential here is the ability to study the route of infection (lung *versus* skin [intradermal] *versus* gut).

- Orally administered BCG is not likely to be a useful model of TB due to the high doses used, and the inability to recover mycobacteria.

- Vaccination with BCG gives variable results depending on age, global location and immune status suggesting that any CHIM data obtained from BCG human challenge must be interpreted with close attention to these parameters.

- Given the safety record of BCG, and recognising the BCG limitations, modified BCG including additional antigens or detection systems could be developed to further refine the BCG model.

- In the event that this BCG work was successful, particularly in detecting and controlling the inoculum strain, then further consideration of modified wild-type *M.tuberculosis* in CHIM models would be warranted.

**Workshop discussion of practical considerations for development of TB-CHIM**

We conducted two workshops (all attendees are listed as authors or as part of the “TB Controlled Human Infection Model Development Group” in the acknowledgements section) to review the current data from human CHIM for TB, and the practical steps needed to take the work forwards. One workshop (residential, 2 days) was held in the UK (24–25 September 2019, Inglewood Manor, Cheshire, UK) and the other in Malawi (21\textsuperscript{st} June 2022, Malawi Liverpool Wellcome Programme, Blantyre, Malawi; meeting on-line and face to face). Both workshops were supportive of TB-CHIM as a worthy research aim, for the urgent reasons introduced earlier in this paper, provided that international standards of safety and volunteer consent could be achieved. Several specific topics were discussed in detail as below.

**Acceptability, specifically in UK and Malawi**

In terms of practical steps, both community and stakeholder acceptability of a TB CHIM to accelerate vaccine development in any specific location (we particularly discussed the UK and Malawi) must first be assessed. There are published experiences of this CHIM acceptability enquiry process in both Kenya for malaria, and Malawi for pneumococcal carriage\textsuperscript{55}. The Malawi stakeholder community has previously included and would still include current CHIM researchers; the District Health Office; members of the Research Ethics Committee; potential volunteers; health experts; physicians; and public health opinion leaders\textsuperscript{56}.

Neither the malaria nor pneumococcal CHIM were first-in-human experiments in Africa. These models were both successfully and safely transferred from the UK after considerable experience in a non-endemic population\textsuperscript{57,58}. In our Malawi workshop, there were two strongly articulated views. One was that the moral imperative to develop vaccines urgently for and in Africa made it important to pioneer these techniques in Africa, in relevant populations with endemic disease. The differences in immune experience of endemic populations might offer either increased risk or protection from the model but in either case an experiment conducted in a non-endemic region would not be informative about that risk. The alternative view was that regulators in Africa would continue to expect research protocols to be tested in research-rich environments supported by sophisticated health care systems before allowing studies of this type in settings with more limited health care systems, particularly Malawi. For example, in Kenya, guidance states that “CHIM models should be developed in maximally
resourced settings before introduction to Kenya”. In any context, therefore, careful stakeholder engagement and community enquiry will be needed to determine where the community opinion lies on this spectrum. Guidance by WHO states that “there has been increased recognition of the potential value of supporting the development of infrastructure and research capacity to enable CHIS to be conducted in disease-relevant local populations in LMIC, where this meets local disease/research priorities and where such resources may not already exist.”

Feasibility evaluation and volunteer recruitment
TB-CHIM using BCG would require appropriate clinical facilities; approved protocols; and recruitment strategy. Given that BCG is used to vaccinate the public in Malawi, offering no transmission risk, there is no requirement for containment facilities for preparation of the inoculum or to accommodate potential volunteers. Modification of BCG, or use of modified \( M.\text{tb} \) in future would require a detailed assessment of transmission risk and re-evaluation, as well as government regulator’s approval for release of a genetically modified organism.

Selection of volunteers in a TB endemic area allows clinical and immunological stratification to high, middle, and low risk of past or current TB. In an African study, the risk of adverse events was highest in volunteers with positive IGRA tests (Quantiferon) and/or abnormal chest radiographs. The workshops therefore concluded that BCG CHIM in Africa should begin with IGRA negative healthy volunteers.

Informed consent quality
Informed consent to human challenge studies in any setting is a little different to other clinical research studies because of the inoculum step proceeding towards disease, rather than a study focusing entirely on prevention, observation, or treatment. As such, information for volunteers must be clear and comprehensible, with risks explained in an appropriate manner. The quality of consent can best be obtained by evaluation of volunteer comprehension both before and after participation, with lessons learned being iteratively applied to improve the quality of the consent process in future. Further, research veterans can offer new volunteers a clear explanation of their experience in a more interesting and engaging manner than researchers, albeit with the risk of bias if only “research champions” are selected. All these methods are applicable and in use in Africa.

Challenge agent and route of delivery
There was unanimous assent in our workshop discussions that wild-type \( M.\text{tb} \) could not currently be used in a pulmonary CHIM experiment because complete cure could not be guaranteed, and tissue damage remains a possibility in pulmonary models. Further, evaluation of the bacterial load of infection in the lung is not yet possible. The development of a TB-CHIM must therefore consider other agents, primarily BCG, modified BCG and modified \( M.\text{tb} \).

A BCG model is the currently preferred design from which to start a mycobacterial CHIM. BCG has been administered to humans for many years by intradermal, oral, and pulmonary delivery. Intradermal and oral delivery are well trusted as safe, having been used to vaccinate billions of people world-wide. Intradermal BCG has been carefully studied in clinical trials as the preferred control in vaccine studies of novel anti-tuberculous vaccines that are also injected. Further, the optimal dose for immunological discovery experiments has been defined and optimised for intradermal delivery. Reservations remain about this model, however, as protection from cutaneous infection may not represent pulmonary protection, and the punch biopsies of the skin are too invasive for repeated serial sampling. There are good data, however, that show the molecular signature of skin biopsies from tuberculin skin test (TST) sites and lungs closely reflect each other both in healthy adults and subjects living with HIV. Exaggerated IL-17 and Th17 responses found in patients with TB compared to subjects with subclinical disease drive the pulmonary damage and are also found in TST but not normal skin. Further, dose of BCG delivered, inflammatory response and inhibition of BCG growth in vivo have all been shown to be closely related, suggesting a useful model. Although successful testing of a vaccine in development has not yet occurred with this model, a vaccine evaluation of MVA85A showed no vaccine effect in the BCG CHIM, consistent with the clinical trial of the same vaccine.

The oral BCG route requires very high dosing, is non-physiological and minimally immunogenic of the systemic compartment, albeit strongly inducing of mucosal immunoglobulin. Also, it is not possible to reliably recover BCG from stool using this route. The pulmonary bronchoscopy instillation route is technically difficult and particularly challenging to determine post-inoculation infection. Currently an Oxford group are working to develop an inhaled model in volunteers with previous BCG experience using post-exposure bronchoscopy to detect persistent pulmonary BCG. The instilled and nebulised pulmonary models are highly sophisticated and simpler models would be preferred in vaccine testing. A key aspect to address though, is measurement of mycobacterial killing versus translocation into lymph nodes.

There are modified BCG models that have shown promise in pre-clinical studies and may soon be applicable, subject to regulatory approval. A very useful potential immediate modification is to use fluorophore insertion to replace microbiological BCG measurement with fluorescent signal. This would have the enormous advantage of continuously measurable readout over time, potentially tracking mycobacterium translocation within the lymphatic system. This method, however, will require regulatory approval of the genetically modified organism; the signal detection technology; and pilot studies to show that the bacteriology and optical readouts correlate tightly within a measurable range.

Conditionally replicating or suicide strains of \( M.\text{tb} \) are also in development, needing some further testing in animal models to be sure of safety and stability of the strains, along with assured suicide in human subjects on removal of the conditioning agent. It is highly probable that such strains, together with appropriate short-course treatment would be very safe. Although there
is a theoretical risk of long-term mycobacterial persistence in pulmonary and extrapulmonary sites\textsuperscript{86}, this risk is low, and steadily decreases each year after exposure. If mycobacterial detection challenges could be overcome using the BCG model, this would mitigate some of the safety challenges with a future conditionally replicating \textit{M}.\textit{tuberculosis} model.

**Different CHIM models that could be applied to TB**

The pathogenesis of pulmonary TB disease leads from primary infection to pulmonary granuloma formation, which may lead to extensive caseation, cavitation and transmission at one extreme, whilst at the other extreme, successful walling off of the infection and eventual sterilisation. In between, pathogenesis may lead to walling off, but failure of sterilisation leading to the risk of subsequent recurrence of disease\textsuperscript{87,88}. Re-infection with a new strain remains possible, but disease is often immunising and re-infection in otherwise healthy adults is considered less common. Many cured patients following TB, however, have longstanding residual lung damage demonstrable by both CT scanning and pulmonary function testing. TB-CHIM models of pulmonary disease might similarly result in lung damage and would therefore be unethical. TB-CHIM of primary infection, or re-infection followed by full early natural sterilisation would incur less risk and should be considered. In TB-endemic regions where most people develop measurable anti-mycobacterial immune responses at a very young age such that a TB-CHIM of primary infection would have limited applicability. A TB-CHIM of value in Africa would therefore be a model of re-infection, either resulting in natural sterilising immunity in otherwise healthy control subjects or the CHIM being controlled at an early stage to prevent any tissue damage. At this point, models utilising administration are more accessible than the pulmonary route. The main purpose of the CHIM in this scenario would be to evaluate vaccines as a means of preventing re-infection.

**Mycobacterial confirmation**

Mycobacterial confirmation of BCG challenge has been consistently possible in some studies using homogenised biopsies of tissue cultured on Middlebrook agar for 5 weeks\textsuperscript{17}. Accurate determination of copy number has been determined using PCR techniques developed in Oxford and applied on the biopsies tissue. BCG detection on swabs of purulent indurated BCG inoculation sites has been possible but not in a quantitative manner suitable for vaccine evaluation\textsuperscript{18}. Culture and PCR from nasal fluid and BAL, correlating with immune responses in these compartments has been possible, but BCG and immune response detection in stools are problematic\textsuperscript{17}. Overall, given the preference for intradermal models, optimisation of less invasive or micro-biopsy sampling and culture/PCR is the immediate priority, with transfer to an optical readout being a most appealing future prospect.

**Immunogenicity endpoints**

It is very difficult to make an early diagnosis of pulmonary \textit{M}.\textit{tuberculosis} infection because the mycobacteria are not only contained in macrophages, but also surrounded by a ball of inflammatory cells and fibroblasts\textsuperscript{87,88}, long before the formation of sterilising granulomas, which themselves show variable structure and function\textsuperscript{89}. Very recent breakthrough studies of exhaled aerosol have demonstrated whole mycobacteria in patients with subclinical disease\textsuperscript{89} but this is a recent discovery; identification is by morphology rather than culture; the application is not yet widespread. In the absence of microbiological endpoints, vaccine efficacy studies have utilised the conversion of previously interferon-gamma-release-assay (IGRA) negative recruits to become IGRA positive as confirmation of infection\textsuperscript{90}. This endpoint has limitations as the immunological response to \textit{M}.\textit{tuberculosis} shows variation between subjects, and can be altered by concurrent infection (e.g., HIV and COVID-19). Nevertheless, just as in prevention of infection vaccine trials, an interferon-based signal is currently the most reliable endpoint in a TB-CHIM, whether it be a release assay or a transcriptome-based assay. Alternative endpoints may potentially include measurement of polyfunctional T cells in the blood and resident in the lung. Transcriptomic studies have shown potentially improved precision in predicting the progression of subclinical (“latent”) disease to active disease\textsuperscript{91}, and these studies have subsequently been confirmed with PET/CT scanning to confirm current active infection in subjects lacking any symptoms\textsuperscript{92}. In the development of a TB-CHIM, consideration must be given to compartmental differences in immune response. Circulating blood lymphocytes, respiratory tract mucosal tissue, bronchoalveolar lavage\textsuperscript{93} and skin will exhibit different populations and percentages of macrophage, T cell and innate cell phenotypes, albeit with substantial overlap. Any endpoint must be validated for reproducibility and reliability in reporting infection, ideally by comparison with a microbiological endpoint. Further refinement of the intradermal CHIM to derive and validate immunological endpoints would be valuable to inform decisions to down-select vaccine candidates; this development work is a priority to make the model fit for purpose.

**Vaccine testing and power calculation**

Many novel vaccine development strategies are in advanced discovery stages, even reaching pre-clinical and phase 1, first in human evaluation. These include repeat BCG, modified BCG including \textit{M}.\textit{tuberculosis} antigens\textsuperscript{92}, modified \textit{M}.\textit{tuberculosis}, vector delivered antigens and RNA vaccines, delivered by a variety of routes, with and without BCG boosting\textsuperscript{93}. Pre-clinical evaluation can include transcriptional profiling \textit{ex vivo}, or in animal models, computer modelling and human studies. If an intradermal BCG was used as a vaccine testing TB-CHIM, what would be the power of this study? In a proof-of-concept design, Harris \textit{et al}., used 12 subjects in each of four groups to compare BCG naïve/vaccinated and MVA85A vaccine \textit{vs}. placebo\textsuperscript{86}. No power calculation was included in this study, and although a study using area under curve of microbiological endpoints would be ideal, multiple biopsy and multiple bronchoscopy endpoints are not likely to be feasible in a TB-CHIM. To be most useful, TB-CHIM studies must report accurate, reproducible data with a much smaller number of subjects than clinical trials—for example, a maximum of 200 subjects tested. Standardisation of the TB-CHIM model in terms of strain; dose; mycobacterial detection methods; and associated immunological
secondary endpoints through collaboration between investigators in the field would represent a major advance and should be prioritised.

Antimicrobials and drug discovery
After many years of frustration, there are now more than 25 new drugs or drug combinations to treat TB in a registered clinical pipeline\textsuperscript{93}. It is widely accepted that combination therapy has been successful in limiting the emergence of drug-resistant TB (MDR-TB) by the targeting of distinct essential biological functions. Furthermore, with the use of combination therapies it has been possible to shorten treatment duration and reduce relapse. However, identifying optimum combinations from a pharmacokinetic and pharmacodynamic perspective is not without significant challenges, and whilst \textit{in vitro} and \textit{in vivo} (animal) pre-clinical models that assess anti-TB drug activity are available, the combination of models that is most predictive of clinical treatment outcome remains unclear. This forms a significant barrier to the prioritization of promising drug regimens and there remains therefore significant room to improve the evidence base prior to commitment to Phase II/III clinical studies. A human challenge model that can be used to uncover drug combination synergies that translate clinically has the potential to accelerate the TB drug development pipeline. The use of a TB-CHIM to test drugs in development has not yet been described and a BCG model has not been used in this way. Anti-tuberculous treatment is offered to patients developing the rare clinical presentation of disseminated BCG disease\textsuperscript{94}. It would therefore be reasonable to expect a TB-CHIM using intradermal BCG to report reduced mycobacterial growth or accelerated clearance when treated with effective regimens or potentially useful novel agents. It is also reasonable to expect that this model would be safe given the published literature reviewed above.

Scientific opportunity
In addition to testing vaccine efficacy and drug potency against mycobacteria, there could be wider scientific interest in a cutaneous TB-CHIM. There are increasing data to show that BCG modifies the human response to infection via interferon dependent mechanisms\textsuperscript{95}, by epigenetic modifications and particularly by modification of macrophage function—“trained immunity”\textsuperscript{96}. Early publications showed increased survival in young children attributable to protection from respiratory infections, and more recent work has confirmed both resilience to respiratory viral infections in older adult subjects\textsuperscript{97,98}, and enhanced response to influenza vaccination\textsuperscript{99}. TB-CHIM using a skin BCG model would potentially allow sophisticated new techniques to be applied indirectly to determine non-specific effects of BCG in defence against respiratory infection\textsuperscript{100}.

A controlled human infection model of TB in Malawi - could it last to utility?
Currently, multiple controlled human infection models are being established in populations with endemic infectious disease throughout the world including \textit{Streptococcus pneumoniae} in Malawi\textsuperscript{101}, falciparum malaria in Kenya and Gabon, vivax malaria in Thailand, hookworm in Brazil; schistosomiasis in Uganda, rotavirus in Zambia (using live vaccine not wild-type virus) and shigella in Kenya. Our group has successfully established the controlled human infection model for \textit{Streptococcus pneumoniae} infection in Malawi. The process of introducing this model included robust stakeholder engagement activities\textsuperscript{72,86}; assessment by both Malawi and UK national ethical and regulatory bodies\textsuperscript{90}; and evaluation of participant acceptability\textsuperscript{86} during the conduct of our feasibility study\textsuperscript{95}. Informed by this work, we have now completed a trial to test the efficacy of licensed pneumococcal conjugate vaccines and measure immunological responses in the Malawian context\textsuperscript{88}. The team conducting this study included Malawian doctors, nurses, data scientists, microbiologists, immunologists, and social scientists. Capacity development of this team was identified as a critical determinant of the longevity of this newly established tradition of CHIM research in Malawi.

There is a longer established tradition of TB research in Malawi, including the detailed study of mucosal immunity\textsuperscript{86,93}. TB remains a major cause of morbidity and mortality with public health interventions designed to control endemic infection impaired by the COVID-19 pandemic\textsuperscript{102}. Introduction of a highly efficacious TB vaccine could have transformative public health benefits. There is both the clinical and laboratory research capacity and the driving need for the TB-CHIM in Malawi, if acceptability, feasibility, and practical challenges can be overcome\textsuperscript{71}. From our experience of introducing pneumococcal CHIM, we would recommend a stepwise approach to introduction of a TB CHIM, de-risking safety aspects in UK based studies before transfer to Malawi. It is likely that introduction of an intradermal (ID) model using currently licensed BCG strains to demonstrate participant safety and develop local researcher skills and experience would be advantageous before potential exploration of different routes of administration and GMO strains.

Conclusions and next steps
There is strong support for a TB-CHIM in the global research community. There is sufficient research capacity and local support to plan a path to develop a TB-CHIM for vaccine and drug testing in both the UK and Malawi. TB controlled human infection models in skin and lung have been established in both the United Kingdom, USA and South Africa using the BCG strain with robust participant safety data reported. There are, however, considerations of risk, advantage, and disadvantage in human infection models of TB infection in Africa that require local and expert consideration. We have outlined potential advantages and disadvantages of alternate approaches in Table 5. Using a pathway for introduction of a relevant and de-risked controlled human infection model (cutaneous BCG) to Malawi for the first time, we now seek to explore if TB-CHIM studies would potentially be acceptable in principle. As part of our stakeholder engagement, we will co-create the optimal design for the Malawian setting, discussing the various trade-offs for each potential approach and exploring local acceptability for each model.
Table 5. Comparison of cutaneous versus pulmonary routes of administration for a mycobacterial controlled human infection model. Table describes administration route considerations for a controlled human infection model of infection. Mtb: Mycobacterium tuberculosis; TICE: US brand name for intravesical BCG. SSI: Statens Serum Institut (Denmark) strain.

<table>
<thead>
<tr>
<th>Administration route</th>
<th>Cutaneous (intradermal)</th>
<th>Pulmonary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk to participants</td>
<td>• Localised mycobacterial administration, facilitates visual monitoring of compartmentalized area</td>
<td>• Lack of compartmentalization and inability to visualize increases theoretical risk</td>
</tr>
<tr>
<td></td>
<td>• More established safety record with multiple cutaneous studies conducted</td>
<td>• Enhanced participant monitoring required compared to skin model.</td>
</tr>
<tr>
<td>Risk to community</td>
<td>• Skin compartmentalized mycobacteria pose minimal onward community transmission risk</td>
<td>• Potential transmission risk to community members from coughing/droplet formation</td>
</tr>
<tr>
<td>Invasive procedures</td>
<td>• Skin sampling may be minimally (swab) to moderately (biopsy) invasive.</td>
<td>• Research bronchoscopy is an invasive procedure requiring trained staff and equipment. May be unpleasant for participants who require local anaesthetic and potentially sedation.</td>
</tr>
<tr>
<td>Recruitment</td>
<td>• Model logistically easier to facilitate potentially rapid recruitment for interventional therapeutic and vaccination trials</td>
<td>• Complex and more costly model to deliver limits the number of participants for interventional therapeutic and vaccination trials</td>
</tr>
<tr>
<td>Mycobacterial recovery</td>
<td>• Reliable culture and molecular quantification from skin administration site.</td>
<td>• Mycobacteria recovered in 6/54 (11%) of participants using bronchoalveolar lavage fluid(^2)</td>
</tr>
<tr>
<td></td>
<td>• Optimal detection method impacted by BCG strain (TICE by skin swab and SSI by punch biopsy)</td>
<td></td>
</tr>
<tr>
<td>Immunological response</td>
<td>• Not the primary site of Mtb infection such that responses may not be representative of disease.</td>
<td>• Mirrors Mtb infection route such that observed responses are more likely reflective of disease.</td>
</tr>
<tr>
<td></td>
<td>• Potential to measure intercompartmental immune responses (skin → blood → lung) after vaccination.</td>
<td>• Need to understand impact of aerosolized vs direct lung installation administration of challenge agent</td>
</tr>
</tbody>
</table>

Data availability
Underlying data
All data underlying the results are available as part of the article and no additional source data are required.

Extended data
Harvard Dataverse: CHIM SR extended and underlying data. https://doi.org/10.7910/DVN/U8IIWZ\(^2\).

This project includes the following extended data:
- PRISMA_2020_checklist.docx
- Reflexivity statement - TB CHIM.docx
- Table S1.xlsx

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

Author information
TB Controlled Human Infection Model Development Group
Thomas Baer\(^1\), Peter Banda\(^2\), David Barr\(^3\), Emma Carter\(^4\), Tarsizio Chikaonda\(^5\), Andrea Collins\(^6\), Graham Cooke\(^7\), Dingase Dula\(^8\), Fred Fyles\(^9\), Kristine Guinn\(^10\), Willem Hanekom\(^11\), Marc Y.R. Henrion\(^12\), Maideyi Kanjewe\(^13\), Nitya Krishnan\(^14\), David G. Lallo\(^15\), Fiona Marston\(^16\), Andrew McCallum\(^17\), Joseph Mfutso-Bengo\(^18\), Alfred Muyaya\(^19\), John Ndaferakhande\(^20\), Brian Ngwira\(^21\), Edna Nsomba\(^22\), Shaun Pennington\(^23\), Anil Pooran\(^24\), Ryan Robinson\(^25\), Eric Rubin\(^26\), David Russell\(^27\), Jamie Ryland\(^28\), Hannah Scott\(^29\), Neema Toto\(^30\), Britta Urban\(^31\), Claire Waddington\(^32\), Steve Ward\(^33\), Robert J. Wilkinson\(^34\), Tom Wingfield\(^35\), Angie Wright\(^36\).

1. Stanford Photonics Research Center, California, USA
2. Malawi-Liverpool Wellcome Trust Clinical Research Programme, Blantyre, Malawi
3. Kamuzu University of Health Sciences, Blantyre, Malawi
4. University of Glasgow, Glasgow, Scotland
5. Liverpool School of Tropical Medicine, Liverpool, UK
6. Imperial College, London, UK
7. Harvard T.H. Chan School of Public Health, Harvard University, Boston, USA
8. Africa Health Research Institute, KwaZulu-Natal, South Africa
9. University of Cape Town, Cape Town, South Africa
10. Cornell University College of Veterinary Medicine, New York, USA
11. The Jenner Institute, University of Oxford, Oxford, UK

Rosenthal SR, McEnery JT, Raisys N; Monteiro-Maia R, Ortigão-de-Sampaio MB, Pinho RT; Minassian AM, Satti I, Poulton ID; Koeken VA, van der Pasch ES, Leijte GP; Hoft DF, Kemp EB, Marinaro M; Hoft DF, Blazevic A, Abate G; Wardhana, Datau EA, Sultana A; Grode L, Ganoza CA, Brohm C.


Meta Roestenberg

1 Malaria Research Group, Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands
2 Malaria Research Group, Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands

The article presents a thorough literature review and report of several discussions between the authors on the next steps regarding a TB CHIM model. Given the impact of TB and the difficulty of progressing novel drugs and vaccines along the clinical development pipeline, the question which is being addressed is extremely timely and important. Although I feel the review was done robustly, and the results are carefully presented, as well as the discussion, the article could benefit from some additional structuring to make the arguments clear for an unfamiliar reader. Already in the introduction the authors state the challenges which come with establishing TB-CHIM (which are substantial):

- Safety aspects related to eradication, treatment, serious adverse events, recurrent infection.
- Validity of the model: e.g. are modified strains such as BCG representative of TB infection?
- Endpoint measures: how to reliably detect infection (and preferably quantify it over time).

Although these aspects are somewhat discussed, the main focus of the review is on route of administration, participant selection, doses and immune responses. It would have been helpful to first address the proposed challenges and provide clear insights into the ultimate compromise of a human model for TB and describe how routes of administration are relevant in the context of these questions.

In the "conclusion and next steps" section, the author seem to settle on cutaneous BCG as the preferred model to be taken forward for testing into endemic areas. However, why this was chosen exactly is less clear from the current text. P.16 states: "a BCG model is the currently preferred design...", but p.15 also has some limitations "intradermal BCG can only be a model of infection, not pulmonary TB disease. Although I fully appreciate (and support) the choices the authors are making, I think it would benefit readers with unfamiliar background if the pros and cons were clearly stated and summarized in the concluding paragraph.
Some additional (minor) comments:
- The summary of immunological findings are only relevant as endpoint to a CHIM if these findings are somehow a correlate of protection. If this is not the case, then this is interesting from a scientific perspective but does not necessarily accelerate the vaccine/drug development. This should then be stated.
- The need for a TB-CHIM in Africa is scientific (as the authors describe) but I think also has a very important strategic aspect in providing endemic sites with tools to take a leading position in the clinical development of novel products for TB. This can be added to page 9.
- Page 14 section on microbiological confirmation lacks discussion on the quantification of microbiological endpoints. This would greatly increase the power of the model to detect differences between vaccinated/treated individuals and placebo controls.
- Page 17 first paragraph addressed "TB-CHIM" in a very general way – I assume challenge with BCG (intradermal?) is the methodology which is referred to here?
- Page 17 second paragraph: the utility of NHP models is an important consideration in the relevance or necessity of a CHIM model. I thus suggest to move this section up (to the intro) then describe what the added value of a human model in addition to the existing NHP models would be.

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**
Yes

**Are sufficient details of the methods and analysis provided to allow replication by others?**
Yes

**Is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are the conclusions drawn adequately supported by the results presented in the review?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Controlled human infections

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Jun 2023

**Ben Morton**

In the "conclusion and next steps" section, the author seem to settle on cutaneous BCG as the preferred model to be taken forward for testing into endemic areas. However, why this
was chosen exactly is less clear from the current text. P.16 states: "a BCG model is the currently preferred design...", but p.15 also has some limitations “intradermal BCG can only be a model of infection, not pulmonary TB disease. Although I fully appreciate (and support) the choices the authors are making, I think it would benefit readers with unfamiliar background if the pros and cons were clearly stated and summarized in the concluding paragraph.

Thank you for this carefully considered review. We agree that further explanation would be beneficial and have now included an additional table to outline the pros and cons around the choice of design.

Some additional (minor) comments:

- The summary of immunological findings are only relevant as endpoint to a CHIM if these findings are somehow a correlate of protection. If this is not the case, then this is interesting from a scientific perspective but does not necessarily accelerate the vaccine/drug development. This should then be stated.

  We agree. We have now made this point clearer in the manuscript in the immunogenicity endpoints section.

- The need for a TB-CHIM in Africa is scientific (as the authors describe) but I think also has a very important strategic aspect in providing endemic sites with tools to take a leading position in the clinical development of novel products for TB. This can be added to page 9.

  We agree completely with this sentiment and have now added a statement to this effect within the “Potential for a TB-CHIM” section.

- Page 14 section on microbiological confirmation lacks discussion on the quantification of microbiological endpoints. This would greatly increase the power of the model to detect differences between vaccinated/treated individuals and placebo controls.

  We agree that robust quantification is required to detect differences between vaccinated/treated individuals and placebo controls. We have now highlighted the importance of this within the section.

- Page 17 first paragraph addressed “TB-CHIM” in a very general way – I assume challenge with BCG (intradermal?) is the methodology which is referred to here?

  We have reworked this paragraph to provide more clarity

- Page 17 second paragraph: the utility of NHP models is an important consideration in the relevance or necessity of a CHIM model. I thus suggest to move this section up (to the intro) then describe what the added value of a human model in addition to the existing NHP models would be.

  We have now moved this paragraph to the introduction section and described what the added value of the human model would be.

Competing Interests: No competing interests were disclosed.
This article explains the need for a controlled human infection model (CHIM) of tuberculosis (TB). It is part systematic review of clinical trials where *Mycobacterium tuberculosis* or the BCG vaccine were administered to adults (27 studies are described, no meta-analysis conducted) and part opinion of the TB-CHIM development group on the best way forward.

No doubt safe TB-CHIM would accelerate vaccine development to alleviate morbidity and mortality caused by TB. It is great that this Malawi-UK collaborative team shares their considerations at the beginning this process. The paper contains a wealth of information but readers would find it more quickly if the structure was streamlined and repetition avoided.

In my mind it would be:

1. Rationale for TB-CHIM;

2. Open questions and then for each question:
   i) review of literature - can the answer to the question be found?
   ii) interpretation - which aspects are unknown?
   iii) how does the committee propose to answer these open questions to move forward? This last part is a little weak at present - more detail and timelines are required.

From this structure it would then be great to include a summary Figure - key points in visually appealing way - so people can put it up in their presentations to broadcast your message!

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**
Yes

**Are sufficient details of the methods and analysis provided to allow replication by others?**
Yes

**Is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are the conclusions drawn adequately supported by the results presented in the review?**
Yes

**Competing Interests:** No competing interests were disclosed.
Reviewer Expertise: Controlled human malaria infection trials, immunology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Jun 2023

Ben Morton

We thank the reviewer for their positive feedback and careful review of our manuscript. We agree that the structure of the piece could potentially be presented in a more accessible way. We have considered this suggestion carefully but feel that restructuring in this way would weaken our reporting from the workshops in Liverpool and Blantyre. We feel it is important to emphasise these stakeholder discussions within the piece. These demonstrate our engagement on this important but potentially controversial area with co-creation of context-sensitive approaches to progression.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 March 2023

https://doi.org/10.21956/wellcomeopenres.20808.r54857

© 2023 McCarthy J. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

James McCarthy

1 Peter Doherty Institute, The University of Melbourne, Melbourne, Victoria, Australia
2 Peter Doherty Institute, The University of Melbourne, Melbourne, Victoria, Australia

The authors present a well-documented systematic review of human infection challenge studies with Mycobacteria, focussing on vaccines for tuberculosis. Their review has been undertaken according to the PRISMA guidelines, and is there is appropriate coverage of the potential role of challenge studies in vaccine selection, identification of immune correlates, possible drug efficacy studies and more fundamental questions around pathogenesis and immunity.

The review summarises the different mycobacterial CHIM models and presents the results in a clear logical manner. The diversity of models, study endpoints and possible correlates of protection are well covered, as are volunteer safety issues. Specific technical challenges of the models, namely choice of challenge agent BCG/GMO etc, route of infection, endpoint selection (quantification of mycobacteria, immune correlates), other factors such overall feasibility of the model and transmission risk are all covered.

The powerful argument that there is an imperative that such translational research be undertaken in disease-endemic low and middle income countries, such as Malawi, is well articulated. It was
pleasing to see that the second workshop was conducted there and local stakeholders are included in this process and authorship.

This review would be strengthened if the issue of how a TB CHIM would provide a convincing case for continued product development (or termination) beyond the challenge study. In malaria and some bacterial and viral challenge models, such systems have established themselves a go-no-go stage gates for vaccine or drug development. The authors conclude that at present the ID BCG infection route is currently the most practical methodology. One is left with the impression that this model as currently used is not entirely fit-for-purpose to down select vaccines. Given the issues the authors highlight, including variability in study design (e.g. dose, recovered mycobacteria detection methods [culture, PCR etc]) and diverse endpoints), the paper would be strengthened if it also made recommendations in regards to standardisation around these parameters, and what development work is a priority to make a model fit-for-purpose.

Implementing the ID BCG model in a high burden setting may provide very useful collateral information, but will it provide the evidence base to advance or stop the development a vaccine along the critical path to registration? A critical summary of research priorities, perhaps in table form of ways that such models can assume a place would add value to the review. For example, do we need to wait for GMO challenge strains, new assays, new approaches to define study endpoints, or new routes of administration such as aerosol?

Minor comments:
- The literature search was carried out more than a year ago. Have any key studies been reported since then?

- I am aware of two prospective studies in the 70s where BCG was administering via aerosol to lung cancer patients; these were not included in the results. They both appear to meet the inclusion criteria in the paper. Is this because the primary endpoint for these studies was safety/lung cancer treatment? Immune responses are reported in the studies e.g. skin test conversion.

- Table 3: there are two errors regarding the TB041 trial. As per the published protocol, the control group was not saline but rather intradermally administered BCG. Also the maximum aerosolised dose was $1 \times 10^7$ CFU not $1 \times 10^6$ CFU (also stated incorrectly on page 14, first paragraph.)

- It would be good to standardise manuscript references or clinicaltrials.gov number (when unpublished) when referring to CHIM studies in Africa (p.9 second paragraph). Some assertions are made without any references i.e. page 9 third paragraph – references would be helpful here. Especially for assertions such as “... (TB) exposure increases susceptibility to mycobacterial infection”.

- The authors mention that large biopsies may be required to recover BCG in ID BCG CHIMS
but clarification of what “large” means would be helpful. Do the authors believe this to be a problem. i.e. increased AEs, requirement for suturing etc? Does this affect consideration of the utility of an ID CHIM models in future.

- The authors propose that the PO administration is not useful model. A more explicit statement why would be helpful. Do the high doses used pose safety concerns? Is it because of inability to recover mycobacteria (as noted by the authors in the conclusion), or is it not a common route of infection (ileal TB being rare).

- Duplicate use of remain page 16: “Reservations remain about this model remain,”

- The ordering of the clinical trials in tables 3 and 4 is not consistent (table 3 appears to be alphabetical, while table 4 is reverse chronological). I prefer a chronologic order.

References

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results presented in the review?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Human challenge models

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Jun 2023
Ben Morton

This review would be strengthened if the issue of how a TB CHIM would provide a
convincing case for continued product development (or termination) beyond the challenge study. In malaria and some bacterial and viral challenge models, such systems have established themselves a go-no-go stage gates for vaccine or drug development. The authors conclude that at present the ID BCG infection route is currently the most practical methodology. One is left with the impression that this model as currently used is not entirely fit-for-purpose to down select vaccines. Given the issues the authors highlight, including variability in study design (e.g. dose, recovered mycobacteria detection methods [culture, PCR etc]) and diverse endpoints), the paper would be strengthened if it also made recommendations in regards to standardisation around these parameters, and what development work is a priority to make a model fit-for-purpose.

**Thank you for these important points. We have now highlighted that further refinement of the ID CHIM model is required to develop for vaccine down selection in the “immunogenicity endpoint” section. We have also highlighted that standardisation of study design is a priority in the “vaccine testing and power calculation” section.**

Implementing the ID BCG model in a high burden setting may provide very useful collateral information, but will it provide the evidence base to advance or stop the development a vaccine along the critical path to registration? A critical summary of research priorities, perhaps in table form of ways that such models can assume a place would add value to the review. For example, do we need to wait for GMO challenge strains, new assays, new approaches to define study endpoints, or new routes of administration such as aerosol?

**Thank you for raising this. We have now expanded the section “A controlled human infection model of TB in Malawi - could it last to utility?” section to explore how we would envisage introduction of the TB-CHIM model in Malawi.**

**Minor comments:**
- The literature search was carried out more than a year ago. Have any key studies been reported since then?

**We have re-run our systematic search (24/04/2023) and no further key studies have been published since our original search.**
- I am aware of two prospective studies in the 70s where BCG was administering via aerosol to lung cancer patients; these were not included in the results. They both appear to meet the inclusion criteria in the paper. Is this because the primary endpoint for these studies was safety/lung cancer treatment? Immune responses are reported in the studies e.g. skin test conversion.


**Thank you for drawing attention to these two interesting studies in patients with lung cancer. Similarly, BCG is also given via the intravesical route for TCC bladder with multiple studies reporting this. We decided to exclude such studies in patients who are likely to have altered immune response to infection but neglected to include this**
criterion in the submitted manuscript. We have now corrected this in the revised manuscript.

- Table 3: there are two errors regarding the TB041 trial. As per the published protocol, the control group was not saline but rather intradermally administered BCG. Also the maximum aerosolised dose was 1x10^7 CFU not 1x10^6 CFU (also stated incorrectly on page 14, first paragraph.)

We have now clarified use of intradermal BCG for the control group in the Table legend (the control group also receive aerosolised saline). We have now checked with Prof McShane (CI for TB041), personal communication, who confirmed the maximum dose was 1x10^7 CFUs. We have now corrected this issue in the manuscript.

- It would be good to standardise manuscript references or clinicaltrials.gov number (when unpublished) when referring to CHIM studies in Africa (p.9 second paragraph). Some assertions are made without any references i.e. page 9 third paragraph – references would be helpful here. Especially for assertions such as "...(TB) exposure increases susceptibility to mycobacterial infection".

We have now added references as suggested and removed examples where references were not accessible.

- The authors mention that large biopsies may be required to recover BCG in ID BCG CHIMS but clarification of what “large” means would be helpful. Do the authors believe this to be a problem. i.e. increased AEs, requirement for suturing etc? Does this affect consideration of the utility of an ID CHIM models in future.

We have now clarified what large means (≥4mm as the standard diameter of a skin punch biopsy) and highlighted the inherent limitation of this. We have also highlighted that exploration of smaller (microbiopsies) in further development/refinement work may be beneficial.

- The authors propose that the PO administration is not useful model. A more explicit statement why would be helpful. Do the high doses used pose safety concerns? Is it because of inability to recover mycobacteria (as noted by the authors in the conclusion), or is it not a common route of infection (ileal TB being rare).

We have now more explicitly highlighted in the review that the main reason we propose PO administration is not useful for this indication is due to the inability to recover mycobacteria.

- Duplicate use of remain page 16: “Reservations remain about this model remain,”

Corrected, thank you.

- The ordering of the clinical trials in tables 3 and 4 is not consistent (table 3 appears to be alphabetical, while table 4 is reverse chronological). I prefer a chronologic order.

We have now updated both tables to be in chronological order.

Competing Interests: No competing interests were disclosed.