AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy

COMMENTARY

Toward an Evidence-Based Nonclinical Road Map for Evaluating the Efficacy of New Tuberculosis (TB) Drug Regimens: Proceedings of a Critical Path to TB Drug Regimens-National Institute of Allergy and Infectious Diseases *In Vivo* Pharmacology Workshop for TB Drug Development

Eric Nuermberger,[a] Christine Sizemore,[b] Klaus Romero,[c] Debra Hanna[c]

Center for Tuberculosis Research, Johns Hopkins University, Baltimore, Maryland, USA[a]; National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, Maryland, USA[b]; Critical Path Institute, Tucson, Arizona, USA[c]

Novel tuberculosis (TB) drug regimens are urgently needed, and their development will be enabled by improved preclinical approaches that more effectively inform and ensure safe selection of clinical candidates and drug combination/regimens. An evidence-based approach for the assessment of nonclinical models supporting TB drug development has been proposed by a joint partnership between the National Institute of Allergy and Infectious Diseases (NIAID) and the Critical Path to TB Drug Regimens (CPTR) Consortium.

General attrition of drug candidates during the development process represents a significant cost and risk to developers. This attrition is particularly damaging to the current tuberculosis (TB) drug pipeline since few commercial organizations are engaged in drug discovery for this neglected disease, which takes the lives of close to 1.5 million people each year. Recognizing the need to assess and improve the predictive accuracy of nonclinical models for TB drug development in order to help improve safe selection of clinical candidates and drug combination/regimens to provide more interpretable tools to drug developers, the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) and the Critical Path to TB Drug Regimens (CPTR), which is a consortium founded by the Bill & Melinda Gates Foundation, the Critical Path Institute (C-Path), and the Global Alliance for TB Drug Development (TB Alliance), convened a workshop in August 2014 to discuss whether currently available and emerging pharmacology-based efficacy models are able to provide drug developers with the appropriate data to confidently select the most promising drugs and drug combinations for clinical development. These discussions were structured to identify gaps in the nonclinical drug development process that are not addressed through the current approaches. As an outcome of the workshop, a road map will be developed to outline what additional studies are required to create evidence demonstrating the utility and validation of *in vitro* and *in vivo* efficacy models to improve the efficiency of development of new drugs and regimens for treatment of TB. While the workshop was focused on approaches to model efficacy of drugs and combinations, the importance of modeling regimen safety was also noted.

The summary below outlines the key elements of the discussions that occurred at the workshop and will be used as the basis to develop an evidence-based nonclinical road map to establish validated models which will provide data critical to the selection of drugs and regimens to improve TB treatment.

BACKGROUND

According to the World Health Organization (WHO), an estimated 9 million people developed TB and 1.5 million died from the disease in 2013 (1). While these numbers reflect a decline in incidence over the past 2 decades, owing in large part to successful drug regimens developed and implemented by WHO and its partners, limitations in implementing these regimens also contributed to the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains and, combined with the worldwide HIV/AIDS epidemic and economic and political instability, have stifled progress, particularly in high-burden countries in regions in Africa, India, China, and the Western Pacific (1).

With currently available drug regimens, patients are typically treated for at least 6 months for drug-sensitive TB cases or for at least 20 months for MDR cases (1). Treatment of MDR and XDR cases is not only more time-consuming but also more toxic and less effective than treatment of drug-sensitive TB. New drug combinations are being developed that are intended to shorten treatment duration, especially combinations comprised exclusively of drugs with new mechanisms of action that would be effective against both drug-sensitive and drug-resistant strains of bacilli, thereby providing new treatment paradigms for all patients. For TB as well as other infectious diseases, pharmacology-based animal models have proven to be useful tools to inform the process of developing drugs and regimens and are a critical data element described in documents providing U.S. Food and Drug Administration (FDA) guidance for TB drug development. However, an-

Accepted manuscript posted online 11 January 2016

Citation Nuermberger E, Sizemore C, Romero K, Hanna D. 2016. Toward an evidence-based nonclinical road map for evaluating the efficacy of new tuberculosis (TB) drug regimens: proceedings of a Critical Path to TB Drug Regimens-National Institute of Allergy and Infectious Diseases *in vivo* pharmacology workshop for TB drug development. Antimicrob Agents Chemother 60:1177–1182. doi:10.1128/AAC.02041-15.

Address correspondence to Debra Hanna, dhanna@c-path.org.

imal models intended to inform drug development and clinical dose selection decisions may capture relevant aspects of drug exposure-response relationships (endpoints that are transferrable from animals to humans) with only varying degrees of accuracy. Therefore, validating the utility of animal models for estimating treatment efficacy in humans will require assessment of the quality of the evidence used to define pharmacological relationships measured in animal models in comparison to those that drive efficacy in humans.

CURRENT AND EMERGING MODELS

Various nonclinical strategies are employed by pharmaceutical companies and product development partnerships to select and evaluate new TB drugs and regimens. Most of these approaches are empirical in nature. There is heavy emphasis on the use of single-agent dose-finding studies in infected mice that are based on human-equivalent drug exposures, if known. This is followed by screening combination efficacy studies in mouse infection models to evaluate bactericidal activity over relatively short durations of treatment. Although these models provide some information on efficacy, data about critical pharmacokinetic/pharmacodynamic (PK/PD) relationships are not always gathered and translation to cure is unknown. Quantifying the contribution of caseous lesions on the PK/PD of TB drugs is an important step toward the development of quantitative mechanistic models that account for this information in the broader context of a more physiologically diverse population of bacteria and the interaction between the pathogen, the host immune system, and the drug treatments. These short-term models are often used to select promising regimens for further study in longer-term efficacy studies using relapse after treatment completion as an endpoint (2–4). The latter models provide information about the ability of drugs to impact actively replicating and also persistent populations of *Mycobacterium tuberculosis* and thereby inform about their sterilizing potential. While technically tractable, these mouse models are nevertheless time-, labor-, and resource-intensive and provide limited opportunity to determine a thorough dose-response relationship or effectiveness against a range of *M. tuberculosis* strains that may show various levels of sensitivity to treatment.

The mouse strains most widely used in TB models (e.g., BALB/c, C57BL/6, Swiss) develop lesions in target organs (lung/spleen) in which *M. tuberculosis* resides predominantly intracellularly and do not develop caseous lesions or cavitation in response to *M. tuberculosis* infection. The latter lesions are pathological hallmarks of cases of human TB in which large numbers of bacilli reside outside cellular compartments in necrotic zones with reduced vascular supply and under altered microenvironmental conditions (e.g., hypoxia) that may affect drug delivery, phenotypic drug susceptibility of *M. tuberculosis*, and/or the mechanism of drug action or activation differently than residence inside macrophages. Therefore, it is logical to conclude that PK/PD relationships established in commonly used mouse strains may not translate directly to such lesions. An emerging mouse model, the C3HeB/FeJ (Kramnik) model, does, however, produce caseous necrotic lesions and even cavities (5) and likely presents a more physiologically diverse population of bacteria and thus may provide greater translational accuracy. Likewise, marmosets develop pathological changes similar to those seen with humans, including cavitation, in response to *M. tuberculosis* infection, but such models are less expensive than larger-nonhuman-primate models such as macaque models and have the additional advantage of frequent twinning, which provides excellent experimental controls. Comparative drug studies employing commonly used mouse models and C3HeB/FeJ mice are needed to show what additional information can be gained using these larger-animal models.

While infection studies in mouse models have been shown to reasonably replicate the bactericidal and sterilizing activity of most existing TB drug regimens, commonly used mouse models do not recapitulate all aspects of human disease and therefore have to be interpreted carefully in selecting new drugs, doses, and regimens for human trials. Commonly used mouse models are designed to deliver reliable, reproducible rates of *M. tuberculosis* infection and stable initial levels of CFU in target organs and are not meant to model the wide heterogeneity seen in human disease and the variability of drug concentrations seen in patients. If the animal systems are modeled and an exposure-response relationship can be generated, then the human PK profiles and their variances can be inserted and outcomes modeled through use of Monte Carlo simulation. For example, because mice of commonly used strains do not develop cavitary disease, their use may not forecast the differences in treatment duration required to effectively treat cavitary versus noncavitary disease. Similarly, unless experiments are designed to model the range of drug exposures observed for key sterilizing agents in humans, results in mice may not provide sufficient insight into the range of human treatment responses. For these and other reasons, the observed number of days that treatment can be shortened in mice with a new regimen may not translate directly to treatment of humans; however, studies employing mice are nevertheless useful in modeling the dynamics of metabolically diverse or drug-resistant subpopulations and are able to suggest regimen-to-regimen differences in sterilizing potential and resistance prevention and hence contribute to regimen selection.

To address various aspects of human disease and PK variability, nonclinical drug development efforts for TB treatment would greatly benefit from the availability of qualified models that are able to query various endpoints. As stated above, current models, i.e., using one strain of mice, one strain of *M. tuberculosis*, one inoculation size, and one drug dose, were designed to be reproducible and to display low variability. It is understood that the mouse model will not be able to reproduce all aspects of human disease, but different mouse models can be utilized and possibly supplemented with other animal models to query various aspects of human relevant pathology, microbial subpopulations, and/or PK/PD parameters.

NONCLINICAL EVALUATION OF DRUG REGIMENS

The discussions in this workshop were targeted at establishing a nonclinical road map for the development and selection of efficacious new drug combinations and are likely applicable to the development of individual drugs that can later be added to or substituted for drugs used in existing TB treatment regimens. In this context, the workshop initially focused on the regulatory guidance that is available to inform studies that are required for later investigational new drug (IND) or new drug application (NDA) submission.

Data needed to inform phase II trials include nonclinical toxicology, nonclinical efficacy, biological rationale for a specific combination of drugs, and evidence of contribution of each drug

to the regimen, as defined by regulatory agencies. In the United States, two FDA guidance documents are relevant in this context: *Guidance for Industry: Codevelopment of Two or More New Investigational Drugs for Use in Combination* (6) and *Guidance for Industry—Pulmonary Tuberculosis: Developing Drugs for Treatment* (7). The focus of this workshop was on discussing key assays and approaches for a nonclinical road map to inform decision-making related specifically to selection of efficacious new drug combinations and their most effective doses. While such concerns are equally important, the workshop did not focus on toxicology and safety considerations for regimen selection.

The individual components listed in the guidance documents served to anchor participant discussions and to address whether nonclinical endpoints of value for selection of clinical drug regimens are indeed addressed by the existing nonclinical *in vivo* models and where more model development may be required. It should be understood, however, that none of the current TB models represent accepted safety models for IND or NDA submissions and as such are not referred to in this text as representing "regulatory safety studies" but rather to identify unexpected untoward effects early that can help guide regulatory-quality animal studies. For codevelopment of combinations, the guidance document states that each component must make a contribution such that the combination is safe and effective. Evidence must be provided, preferably from *in vivo* models, to support the biological rationale, safety, and efficacy of a combination. *In vitro* PK/PD models may also be used to support the *in vivo* data. One of the approaches discussed in the codevelopment guidance is the use of factorial designs, which test components alone as well as in combination. However, because of the risk of development of resistance associated with monotherapy in TB, individual drugs can be studied in patients for a maximum of only 2 weeks before effective combination therapy must be initiated.

(i) Demonstrating efficacy. Demonstrating efficacy with nonclinical models requires identification of relevant endpoints and of the most appropriate model or set of models with which to investigate those endpoints, as well as the context of the use for which regulatory endorsement by FDA and European Medicines Agency (EMA) would be sought. The context-of-use statement represents the main construct behind the regulatory endorsement of drug development tools, as it constitutes the description of what the tool (animal model) can provide in terms of information and of how such information can be used for specific purposes during the drug development process. Nonclinical models need to capture data that correlate with clinical efficacy, i.e., prevention of death, cure without relapse, or prevention of resistance. Yet while the gold standard clinical definition of "cure" of TB has long been prevention of relapse after discontinuation of treatment, there remains uncertainty about how well differences in the rates of clearance of bacteria from human sputum or animal organs during therapy predict differences in the durations of treatment needed to produce nonrelapsing cure. Relapse in murine TB models is caused by organisms that are viable but noncultivable on solid media and that are induced and/or selected by treatment. Whether a phenotypically similar population would be responsible for relapse in humans, thus linking the nonclinical and clinical conditions, is unknown, but the possibility is suggested by the superior sensitivity of liquid culture methods and the identification of viable but noncultivable bacilli in human sputum (8). More-sensitive culture methods or the use of molecular markers that have been validated against current standard methodologies could clarify the translational value of the relapse endpoint in mice and other nonclinical species and establish more-efficient surrogate markers.

Most clinical relapses occur in the first 6 months after completion of treatment, with similar observations reported in mouse models, which suggests that relapse in mouse models should be evaluated at 3 to 6 months posttherapy. While it is recognized that mouse models often do not precisely represent human drug pharmacokinetics or the full range of pathology and disease dynamics of human TB but rather provide relevant data on the relative ability of a drug to affect bacterial loads in target organs, these animal models are nevertheless important in helping relate quantitative changes in bacterial loads to relapse and the governing drug exposure-response relationships. PK/PD parameters initially modeled in *in vitro* systems can be tested in animal studies to arrive at drug combinations and doses for clinical evaluation; conversely, the PK/PD data obtained in *in vivo* systems can contribute to the establishment of new hypotheses that can be modeled *in vitro*.

(ii) Biological rationale for the selection of regimens. Closely related to the issue of efficacy is the biological rationale for selection of the components of a combination regimen. To shorten therapy, for example, one assumption is that larger overall drug exposures are needed to achieve faster and more-effective bacterial killing. However, considering that the correlation of initial antibacterial activity with duration of treatment and lack of relapse has not been established, it is more likely that factors beyond the overall magnitude of drug exposure, such as efficacy against various metabolically different bacterial populations, need to be considered. While targeting initial maximal bacterial killing has proven of value in the treatment of other infectious diseases, the slow growth of *M. tuberculosis* and the hypothesized presence of persister populations of bacteria will likely require additional strategies for combining drugs beyond assessing initial bacterial killing.

Ideally, *in vitro* models, such as hollow-fiber infection models, as well as *in vivo* studies would demonstrate these key factors for an effective drug combination: *(a)* drug synergy or antagonism; *(b)* cross-resistance of regimen components; *(c)* differential and complementary PK; *(d)* appropriate tissue and lesion penetration; and *(e)* activity against various metabolically different bacterial populations that are thought to determine the response to therapy (4). While many of these parameters have not been validated as predicting human relevant outcomes, they nevertheless contribute to the list of characteristics for a given drug or regimen so that the most promising or diversely active candidate can be selected for clinical trials. These models would also optimally contribute to the establishment of optimized doses of individual components and give insight into the toxicodynamics of a specific multidrug combination (9). Drugs that kill by different mechanisms may produce different optimal effects against replicating and persistent bacteria and may contribute differentially to bacterial sterilization and relapse (4, 10). The collection of nonclinical data addressing the parameters noted above may also improve predictions of the efficacy and safety of regimens against extrapulmonary forms of disease (e.g., TB meningitis) or in special populations (e.g., patients with HIV/AIDS, children).

Investigations of the ability of individual drugs to penetrate target tissues and lesions have recently gained increased interest since they may help to minimize the possibility of delivering sub-

optimal drug exposures or functional monotherapy to key regions or organs where bacteria reside. Tissue penetration can be modeled in animals using radiolabeled drug derivatives to assess penetration into various compartments (e.g., lung, granulomas, and peripheral circulation) and to determine the ratio between drug levels in serum and levels in target organs. A promising method for determining the dynamics of drug penetration in human lungs is the assessment of drug levels in epithelium lining fluid (ELF). *In vivo* imaging methods using bioluminescence, positron emission tomography, and fluorescence (11) and *ex vivo* approaches based on mass spectroscopy (12) are currently being studied as tools to assess tissue penetration. Once the methods are standardized, the resulting imaging and PK data could be fed into physiology-based PK (PB/PK) modeling platforms to establish the link between nonclinical measures of tissue penetration and treatment outcome.

In studies of other infectious diseases, dynamic *in vitro* models such as hollow-fiber models have been used to estimate the dose-response range for anti-infectives that are then tested in animal models. Estimation of clinically effective exposure profiles based on animal efficacy model experiments using single doses (with a sole focus on matching an average human PK parameter value rather than a range of clinically relevant exposures), while providing important information, does not provide a reliable assessment of when biological efficacy may be lost and when a maximal biological effect is obtained. However, these data are critical for further *in vitro* or *in vivo* modeling that takes drug exposure variability in more-complex systems, i.e., humans, into consideration (13–15). This issue is illustrated in the failure of clinical trials in which drug doses were not adequate to provide efficacious drug exposures in many patients. On the other side, relying on maximally tolerated doses to achieve the desired biological effect may result in massive overtreatment of patients. Teasing out these conflicting issues in a nonclinical setting for combination regimens with *in silico*, *in vitro*, and in-animal models is complex but essential. For example, one reason for avoiding high doses is the assumption that treatment will be needed for a long period of time whereas using a shorter-duration regimen administered at a higher dose or switching regimens and using shorter treatment durations may be achievable and provide the desired biological effect while limiting toxicity (16).

Several scientific hypotheses currently exist that describe diverse metabolic populations of bacilli present throughout the various stages of TB disease that are thought to be responsible for the need for long durations of treatment with currently available drugs. The use of animal models that reflect populations of replicating bacilli, nonreplicating "persisters," and mutants with reduced drug susceptibility is important to assess the potential of drugs to eradicate infection and prevent relapse. One potential advantage of the hollow-fiber model is that it can efficiently determine isolated dose-response effects on a variety of bacterial strains and metabolic subpopulations and thus give insight into microbiologic variability and the development of resistance (4, 10). These data, combined with mouse dose-response efficacy data and patient PK variability data, can be integrated into complex models and simulations to determine the range of drug exposure profiles that may be needed to provide efficacious doses for the largest number of patients. The high predictive accuracy of the hollow-fiber model led to the announcement of a recent positive qualification opinion from the EMA regarding the use of this platform as a drug development tool for treatment of TB. However, the relevance of specific metabolic subpopulations created under the stress conditions imposed in the hollow-fiber model to the persister populations responsible for human relapses must be confirmed with further study.

It is clearly important to assess dose response in nonclinical models so that, by the time a regimen is administered to patients, the range of combination doses that have the greatest likelihood of being efficacious has been identified. However, one would optimally also evaluate multiple doses in humans to confirm exposure response parameters. The expectation is that nonclinical models could help enable more-rational selection of doses for initial clinical trials.

(iii) Modeling drugs in combination. Treatment with drug combinations is the standard of care for active TB. Since the benefit, or lack thereof, of a regimen or a regimen component may not become apparent until late-stage clinical development, nonclinical models play an important role in dissecting the contributions of different components of the regimen and their potential for toxicity, drug-drug interactions, and resistance development.

To dissect the contributions of individual drugs to a regimen, one could consider developing models where animals are infected with a strain of *M. tuberculosis* that is genetically altered to no longer respond to one of the components of the regimen while comparing efficacy to that of a model infected with the fully drug-sensitive, isogenic parent strain. Sequentially testing the effectiveness of the regimen in this way rather than testing regimens minus one drug at a time offers the advantage that mutually protective or antagonistic effects of the drug combinations are maintained under all conditions while the responsiveness of the infecting pathogen is varied. However, the results could be confounded if the drug-resistant mutant has a level of fitness different from that of the parent. An alternative or complementary methodology could be that of identifying promising drug combinations using similar approaches in the hollow-fiber model and then validating results in animal models, thus reducing the numbers of animals required for study.

Last, to address the effect of a regimen on different levels of bacterial burden and bacterial populations, including genotypically resistant subpopulations, that arise or are selected during the course of treatment, *in vitro* studies can be conducted with different inoculum sizes or metabolically distinct populations. These initial observations can then be validated in animal models that favor these parameters, such as an immunocompromised mouse that displays high bacterial burden in lungs or C3HeB/FeJ mice that, due to the presence of distinct necrotic lung granulomas, present a more diverse population of *M. tuberculosis* in tissues (5–7, 11, 12, 17). These approaches, combined with quantitative modeling, may provide additional information about variations in drug efficacy against discrete subpopulations of bacteria over the course of treatment.

(iv) Role of nonclinical data in phase II combination trials. While phase II clinical trials in TB have shown notable effects of new drug combinations on reducing bacterial counts in human sputum over the course of 2 weeks or 2 months, the value of such data for predicting differences in long-term treatment outcomes and reduction of relapse remains unclear. In contrast, weak results in phase II trials do not necessarily exclude the possibility that a regimen would have desirable long-term effects, since changes in sputum cultures over the initial 2 weeks or even 2 months may not

reflect the effects of delayed tissue penetration, time-dependent changes in drug metabolism or PK interactions, host factors such as immune status or resolution of cavitation, and clearance of important bacterial subpopulations, including drug-resistant mutants and phenotypically tolerant persisters. Due to the challenges of studying such effects in early clinical development, appropriate use of nonclinical models to explore such effects can better inform clinical trial design, manage expectations, and eventually improve the linkage between nonclinical and clinical outcomes. Greater confidence in this linkage would promote development of new treatment paradigms. For example, a more frequent and targeted change of drug combinations (drug sequencing) to optimize efficacy during various stages of treatment could be considered rather than adhering to the traditional paradigm of a 2-month intensive phase followed by a shortened continuation phase (16).

With regard to the contribution of the immune system to the reduction in bacterial counts, studies evaluating different regimens and dosing schedules in athymic nude mice have shown that, despite their showing PK profiles similar to those seen with immunocompetent mice, the magnitude of the treatment effect in these immunodeficient mice is smaller and there is a greater propensity to select for drug-resistant mutants (4). This is likely due to a combination of a higher intrinsic bacterial growth rate and reduced fitness cost for drug-resistant mutants due to the lack of an adaptive immune response. Testing drug regimens in an immunocompromised model may be useful to estimate a regimen's effectiveness in immunocompromised patients and may also contribute data regarding its potential to select for resistance in such patients or in lesions where the host immune response is ineffective, such as cavities.

Studies in C3HeB/FeJ mice (17) have also demonstrated an increased likelihood of selecting drug-resistant mutants which is likely determined by higher bacterial burdens as well as by less-effective immune responses, particularly in large caseous lesions. However, drug-specific issues related to reduced penetration or activity in such lesions may also play a role.

TOWARD AN EVIDENCE-BASED NONCLINICAL ROAD MAP

An evidence-based road map is needed to define which nonclinical models are best suited to address specific biological issues and which nonclinical endpoints best represent accepted clinical efficacy endpoints in order to better translate nonclinical data to the clinic. On the basis of the discussions outlined in this paper, mouse models could help elucidate what role the size of the bacterial burden at the beginning of treatment may play for treatment shortening or for initial reduction of bacterial counts in target organs. Studies using mouse models, since they are relatively easy to conduct, are affordable and provide important information about the direct effect of drugs on viable bacterial counts and can be used to bridge *in vitro* modeling data on kill rates and PK/PD estimates to more-complex *in vivo* situations and to confirm the relationship between *in vitro* drug MICs and actual drug exposure. Mouse models can also be useful in recreation of the heterogeneity of bacterial populations, especially through the use of mouse strains that develop caseous lung lesions that may harbor a greater range of metabolically different populations of bacteria or of strains with compromised immune systems that carry extremely high bacterial loads and therefore allow selection of preexisting resistant bacilli such as may be present in patients. Other animal models which feature caseous lesions, such as marmosets, may be useful in a similar capacity (18). Better definition of the clinically relevant biological endpoints in animal models will foster better understanding of the composite of nonclinical data needed to increase confidence in the selection of the best drug and regimen doses and combinations for human clinical trials.

In order to improve the utility of animal models for bridging nonclinical and clinical settings, more dedicated PK/PD-based approaches are needed to define dose-response relationships of drugs and drug combinations in animals to investigate the impact of human population-level PK variability on outcomes utilizing commonly used endpoints. These refined models could be used to further analyze the relationship between results from mouse efficacy models and those from recent phase II clinical trials and longer phase III clinical trials, such as those in which 4-month fluoroquinolone-containing regimens did not meet noninferiority criteria relative to the standard of care. Such PK/PD-based analyses should put the sequence of nonclinical and human phase II and III trials into better perspective and help improve the interpretation of current nonclinical and clinical approaches for dose and regimen selection.

To define a road map for the nonclinical evaluation of candidate combinations, a series of critical biological questions relevant to human efficacy need more careful analysis to identify data gaps that currently prevent more confident bridging of nonclinical and clinical settings, including the following:

- Is there a quantifiable, operational definition of "cure" that can be modeled/measured in *in vivo* systems?
- What is the relationship between the kinetics of the early reduction of bacterial load in lungs and the duration of treatment needed for cure?
- What critical PK/PD relationship(s) for individual drugs or drug combinations can be translated from *in vitro* and *in vivo* models into human clinical trials?
- What aspects of human physiology, immunology, and pathology modify the treatment response and should be modeled or measured in nonclinical models?
- What aspects of bacterial physiology, pathogenesis, or drug susceptibility modify the treatment response and should be modeled or measured in nonclinical models?
- Is there a rationale for changing the composition or dosing of a regimen during treatment to maximize bacterial killing and resistance prevention while minimizing toxicity?

To address these questions in an evidence-based manner, the current *in vivo* and *in vitro* models applied to TB drug development need additional scrutiny to identify the most suitable, relevant, and interpretable approach with answers to questions such as the following:

- What *in vitro* or *in vivo* model(s) is best suited to addressing the various parameters outlined above, i.e., effect against diverse bacterial populations, effect against bacteria that reside within or outside cells *in vivo*, effect against large bacterial loads, role of immune system in clearance, establishment of "portable" PK/PD parameters to estimate clinical doses, etc.?
- What is the role of emerging models featuring caseation, e.g., models using C3HeB/FeJ mice or marmosets? Do they

contribute unique data sets for critical development decisions? What are the PK/PD relationships for TB drugs in these models, and do they track with clinical experiences/other established models?

- What is the value of a murine model of persisters and its translatability to humans?
- What combinations of models and repeat experiments, e.g., using different bacterial strains, are needed to address performance of drugs/regimens in the complex clinical setting?
- Can quantitative PK/PD-based models that better account for variability in human PK, disease severity, and bacterial drug susceptibility improve the predictive accuracy of nonclinical models and simulations?
- Can biomarkers (e.g., indication of bacterial load, presence of resistant populations, altered drug metabolism and negative impact on exposure profiles, etc.) be established that mark stages of drug efficacy along the course of treatment and that correlate animal and human studies?

CONCLUSIONS

Drug development is a process of building confidence in a candidate molecule or regimen in a step-by-step fashion. Selection of new TB drugs and regimens to provide a go/no-go decision based on a single, uniform sequence of studies in a single efficacy model will likely not be possible but will more likely be based on a composite of data derived from a series of models addressing key aspects of efficacy. Nonclinical models will never replace clinical trials but may be useful in predicting clinically observed phenomena while keeping safety and efficacy as the primary mandates. Therefore, development of quantitative and nonquantitative nonclinical models must go hand in hand with evolving strategies for incorporating the variability in disease state, host immune function, drug PK, and bacterial susceptibility that occurs clinically.

While drug development is more complicated for TB than for other bacterial indications due to the requirement for multidrug therapy and the current long duration of treatment, the field can nevertheless benefit from using animal and *in vitro* modeling approaches that have been shown to contribute to the development of therapies against other infectious diseases. Available tools in TB drug development research can be aligned in a more rational manner to create diverse data sets that cover clinically relevant considerations and contribute to confidence in carrying a new regimen into advanced clinical trials. The establishment of a road map for nonclinical development of drug regimens will be an important tool to help guide this process.

REFERENCES

1. **World Health Organization.** 2013. Global tuberculosis report 2013. World Health Organization, Geneva, Switzerland.
2. **Ahmad Z, Nuermberger EL, Tasneen R, Pinn ML, Williams KN, Peloquin CA, Grosset JH, Karakousis PC.** 2010. Comparison of the 'Denver regimen' against acute tuberculosis in the mouse and guinea pig. J Antimicrob Chemother **65:**729–734. http://dx.doi.org/10.1093/jac/dkq007.
3. **Gumbo T, Angulo-Barturen I, Ferrer-Bazaga S.** 2015. Pharmacokinetic-pharmacodynamic and dose-response relationships of antituberculosis drugs: recommendations and standards for industry and academia. J Infect Dis **211**(Suppl 3)**:**S96–S106. http://dx.doi.org/10.1093/infdis/jiu610.
4. **Gumbo T, Lenaerts AJ, Hanna D, Romero K, Nuermberger E.** 2015. Nonclinical models for antituberculosis drug development: a landscape analysis. J Infect Dis **211**(Suppl 3)**:**S83–S95. http://dx.doi.org/10.1093/infdis/jiv183.
5. **Driver ER, Ryan GJ, Hoff DR, Irwin SM, Basaraba RJ, Kramnik I, Lenaerts AJ.** 2012. Evaluation of a mouse model of necrotic granuloma formation using C3HeB/FeJ mice for testing of drugs against Mycobacterium tuberculosis. Antimicrob Agents Chemother **56:**3181–3195. http://dx.doi.org/10.1128/AAC.00217-12.
6. **Food and Drug Administration.** 2013. Guidance for industry: codevelopment of two or more new investigational drugs for use in combination. Food and Drug Administration, Silver Spring, MD. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm236669.pdf. Accessed 22 August 2015.
7. **Food and Drug Administration.** 2013. Guidance for industry—pulmonary tuberculosis: developing drugs for treatment. Food and Drug Administration, Silver Spring, MD. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm373580.pdf. Accessed 22 August 2015.
8. **Mukamolova GV, Turapov O, Malkin J, Woltmann G, Barer MR.** 2010. Resuscitation-promoting factors reveal an occult population of tubercle bacilli in sputum. Am J Respir Crit Care Med **181:**174–180. http://dx.doi.org/10.1164/rccm.200905-0661OC.
9. **Brown AN, Drusano GL, Adams JR, Rodriguez JL, Jambunathan K, Baluya DL, Brown DL, Kwara A, Mirsalis JC, Hafner R, Louie A.** 2015. Preclinical evaluations to identify optimal linezolid regimens for tuberculosis therapy. mBio **6:**e01741-15. http://dx.doi.org/10.1128/mBio.01741-15.
10. **Drusano GL, Sgambati N, Eichas A, Brown DL, Kulawy R, Louie A.** 2010. The combination of rifampin plus moxifloxacin is synergistic for suppression of resistance but antagonistic for cell kill of Mycobacterium tuberculosis as determined in a hollow-fiber infection model. mBio **1:**e00139-10. http://dx.doi.org/10.1128/mBio.00139-10.
11. **Zelmer A, Carroll P, Andreu N, Hagens K, Mahlo J, Redinger N, Robertson BD, Wiles S, Ward TH, Parish T, Ripoll J, Bancroft GJ, Schaible UE.** 2012. A new in vivo model to test anti-tuberculosis drugs using fluorescence imaging. J Antimicrob Chemother **67:**1948–1960. http://dx.doi.org/10.1093/jac/dks161.
12. **Dartois V.** 2014. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. Nat Rev Microbiol **12:**159–167. http://dx.doi.org/10.1038/nrmicro3200.
13. **Pasipanodya JG, Nuermberger E, Romero K, Hanna D, Gumbo T.** 2015. Systematic analysis of hollow fiber model of tuberculosis experiments. Clin Infect Dis **61**(Suppl 1)**:**S10–S17. http://dx.doi.org/10.1093/cid/civ425.
14. **Gumbo T, Pasipanodya JG, Nuermberger E, Romero K, Hanna D.** 2015. Correlations between the hollow fiber model of tuberculosis and therapeutic events in tuberculosis patients: learn and confirm. Clin Infect Dis **61**(Suppl 1)**:**S18–S24. http://dx.doi.org/10.1093/cid/civ426.
15. **Gumbo T, Pasipanodya JG, Romero K, Hanna D, Nuermberger E.** 2015. Forecasting accuracy of the hollow fiber model of tuberculosis for clinical therapeutic outcomes. Clin Infect Dis **61**(Suppl 1)**:**S25–S31. http://dx.doi.org/10.1093/cid/civ427.
16. **Drusano GL, Neely M, Van Guilder M, Schumitzky A, Brown D, Fikes S, Peloquin C, Louie A.** 2014. Analysis of combination drug therapy to develop regimens with shortened duration of treatment for tuberculosis. PLoS One **9:**e101311. http://dx.doi.org/10.1371/journal.pone.0101311.
17. **Lanoix JP, Lenaerts AJ, Nuermberger EL.** 2015. Heterogeneous disease progression and treatment response in a C3HeB/FeJ mouse model of tuberculosis. Dis Model Mech **8:**603–610. http://dx.doi.org/10.1242/dmm.019513.
18. **Via LE, Weiner DM, Schimel D, Lin PL, Dayao E, Tankersley SL, Cai Y, Coleman MT, Tomko J, Paripati P, Orandle M, Kastenmayer RJ, Tartakovsky M, Rosenthal A, Portevin D, Eum SY, Lahouar S, Gagneux S, Young DB, Flynn JL, Barry CE, III.** 2013. Differential virulence and disease progression following Mycobacterium tuberculosis complex infection of the common marmoset (Callithrix jacchus). Infect Immun **81:**2909–2919. http://dx.doi.org/10.1128/IAI.00632-13.