Mathematical models have facilitated our understanding of infectious diseases dynamics and proved useful tools to compare control scenarios when interventional studies are not feasible or ethical. Here, we summarize evidence linking social, economic and biologic determinants to tuberculosis (TB) and review modeling approaches that have been used to understand their contribution to the epidemic dynamics of TB. Specifically, we find evidence for associations between smoking, indoor air pollution, diabetes mellitus, alcohol, nutritional status, crowding, migration, aging and economic trends, and the occurrence of TB infection and/or disease. We outline some methodological problems inherent to the study of these associations; these include study design issues, reverse causality and misclassification of both exposure and outcomes. We then go on to review two existing approaches to modeling the impact of determinants and the effect of interventions: the population attributable fraction model, which estimates the proportion of the TB burden that would be averted if exposure to a risk factor were eliminated from the population, and deterministic epidemic models that capture transmission dynamics and the indirect effects of interventions. We conclude by defining research priorities in both the study of specific determinants and the development of appropriate models to assess the impact of addressing these determinants.

**KEY WORDS:** smoking; IAP from solid fuels; diabetes mellitus; alcohol; body mass index; undernutrition

**SUMMARY**

Despite improvements in case detection and tuberculosis (TB) treatment outcomes over the past decades, TB incidence has not declined as quickly as predicted. This raises the question of whether other approaches to TB control should be considered. Mathematical models have facilitated our understanding of infectious disease dynamics and proved useful tools to compare control scenarios when interventional studies are not feasible or ethical. Over the past 20 years, models of TB transmission have been developed that have focused on interventions such as DOTS, improved diagnostics, novel vaccines and drug therapies. More recently, researchers have attempted to measure the contribution of social, environmental and biological determinants to the burden of TB and to model the potential impact of the modification of those that are amenable to public health interventions. In this article, we briefly summarize existing evidence on determinants and review modeling approaches that have been used to understand their contribution to the epidemic dynamics of TB.

**SOCIAL, ENVIRONMENTAL AND BIOLOGICAL DETERMINANTS OF TB**

Social, environmental and biological determinants of health have long been recognized as risk factors for TB; however, only recently have the contributions of specific determinants been assessed through systematic studies and summaries of existing data. Such studies have been conducted on human immunodeficiency virus (HIV), smoking, indoor air pollution (IAP), diabetes mellitus (DM), alcohol use and body mass index (BMI) and under-nutrition, factors that are widely prevalent, responsible for substantial morbidity and mortality, and can be targeted through individual or population level interventions. Other, less-well studied determinants include those related to urbanization, crowding, housing conditions, migration, aging, economic trends and host genetics and co-morbidities. As our goal is to consider modeling the impact of public health interventions on the epidemic behavior of TB, we focus here on the first set
of determinants, describing only briefly the existing data on less proximal social factors.

**HIV**

Among the determinants that have measurably increased individual level risk of active TB, HIV is the one that has had the most impact on altering TB incidence rates over the past several decades. Numerous studies have documented the increased risk of TB disease among the HIV-infected compared to the non-infected, with current estimates suggesting an incidence rate ratio as high as 20. Among the HIV-infected population, low CD4 and high viral loads have been found to be further risk factors for disease, while treatment with highly active antiretroviral therapy reduces risk. On a population level, rising HIV prevalence is strongly associated with dramatic increases in TB notification rates, especially in sub-Saharan Africa, where at the peak of the HIV epidemic TB incidence rates were three- to five-fold higher than in 1980. The Joint United Nations Programme on HIV/AIDS estimated that 33.4 million people were living with HIV in 2008, of which 22.4 resided in sub-Saharan Africa.

**Smoking**

In 2006, 5.8 trillion cigarettes were manufactured, equivalent to 2.4 cigarettes per day per person. Between 1970 and 2000, cigarette consumption declined by 0.2% annually in high-income countries, but increased by 5% annually in low- and middle-income countries (LMICs). The World Health Organization's Commission on Macroeconomics and Health concluded that the economic burden from tobacco-associated illness and premature death is a major impediment to economic development in LMICs. A joint report from the WHO and the International Union Against Tuberculosis and Lung Disease concluded that exposure to tobacco smoke was significantly associated with TB infection and disease, and several systematic reviews have found that rates of TB infection, disease and mortality were significantly higher among smokers.

**Indoor air pollution from solid fuels**

More than half of the world’s population uses solid fuels, with a particularly high prevalence (>70%) in sub-Saharan Africa and South-East Asia, where TB burden is high. Although indoor smoke from solid fuels might increase the risk of TB through a mechanism similar to that of tobacco smoke, evidence for the association between IAP and TB is far more limited. Two systematic reviews have found a possibly increased risk of TB among solid-fuel users; however, they also pointed to the need for more high-quality data.

**Diabetes mellitus**

The WHO reported that in 2005 an estimated 1.1 million people died from DM, with almost 80% of diabetes deaths occurring in LMICs, and that the number of diabetes deaths is expected to double between 2005 and 2030. A recent meta-analysis found that patients with DM were more than three times more likely to have TB than controls.

**Alcohol**

Alcohol-related diseases were estimated to account for 3.8% of all deaths, and 4.6% of the global burden of disease in 2004. Although the association of alcoholism and TB was first described in 1785, only one systematic review has examined this association. In 21 studies identified, all forms of TB disease were about three times higher, and pulmonary TB was four times more frequent, in heavy drinkers than controls.

**Body mass index and under-nutrition**

There are relatively few data on the association between nutrition or low BMI and TB, perhaps due to the complexity of mechanisms linking nutrition, the immune system and TB, and the difficulties in identifying the relevant biologic markers. One recent review on TB and low BMI has indicated a strong dose-response relationship, with TB incidence increasing exponentially as BMI decreased in the six studies summarized. On a population level, the impact of this risk factor has been predicted to be enormous, due to the widespread global prevalence of under-nutrition.

**Social and environmental factors**

Crowding and urban dwelling have long been established as risk factors for exposure to TB infection. In one large study among white male US navy recruits, lifelong residents of urban areas were 1.5 times more likely to be positive skin test reactors than were lifetime residents of farms. Housing conditions often reflect socio-economic status and can affect TB infection risk and outcomes through such mediators as poor ventilation and air quality in the home.

Immigration is also a social determinant of population TB risk. Migrants from countries with a high burden of TB are often at risk for the disease after arrival in low-burden countries; high rates of disease often occur within the first year of arrival, with a slow decline that can persist for at least a decade. While these data suggest that primary TB may be a frequent cause of disease among newly arrived immigrants, at least one study conducted among the Somali community in Toronto reported that more than half of the cases were likely due to reactivation.

The aging of populations may also increase risk of TB, as cumulative exposure to infection and the risk of progression given an infection increases with age. The institutionalization of the elderly may also lead to nosocomial disease in this vulnerable group, and the rates of TB infection and disease have been found to be high in congregate settings such as nursing homes. As LMICs go through a demographic transition, the growing population of the elderly will...
represent a huge reservoir of TB infection that will in turn present a host of unique challenges to TB treatment and control.44

Lastly, TB is clearly associated with poverty on both the individual and the country level. As poverty underlies many of the known risk factors for TB, such as under-nutrition, poor housing, and IAP, its impact may represent the individual effects of multiple proximal determinants. For example, several studies have documented an association between per capita gross domestic product and estimated national TB incidence;45,46 however, this relationship was found to be attenuated in a multivariate model that included more proximal determinants.

METHODOLOGICAL PROBLEMS IN ESTIMATING THE IMPACT OF DETERMINANTS

Data summaries are only as good as the data summarized, and the meta-analyst of observational studies runs the risk of propagating or even amplifying biased results by pooling them.47,48 Most studies on risk factors for TB use either a case-control design that compares the distribution of determinants in cases and controls, or a cohort approach in which people with and without risk factors are followed for the occurrence of disease. As they are more feasible than cohort studies, case-control studies make up the majority of the studies cited above. These studies may be prone to a number of possible biases. First, as the risk factor is ascertained at the time a case is recruited, the temporal association between the variables cannot be assessed. For example, TB disease results in weight loss that would lead to low BMI at the time of TB diagnosis49,50 and to inflammation-mediated hyperglycemia51 that could result in a diagnosis of DM. Active TB may also lead to reductions in exposures; smokers may stop smoking when afflicted with pulmonary TB disease. Ideally, cohort studies should minimize the bias introduced by this reverse causality, and some systematic reviews, such as on the association of BMI with active TB,54 have included only cohort studies in their data summaries. In some circumstances, however, the ethical obligation to treat newly diagnosed DM, for example, or to re-feed undernourished participants, might make it impossible to observe the true impact of such exposures. In addition, odds ratios from cross-sectional studies and case-control studies that included only prevalent cases of TB (instead of incident cases) would not approximate incidence rate ratios if the risk factors also change the duration of TB disease, as prevalence can be affected by both incidence and duration of disease. For example, HIV infection dramatically shortens the duration of TB disease due to the high mortality among people co-infected with HIV and TB, and some studies on the HIV-TB association have reported much smaller odds ratios from prevalence studies than incidence rate ratios from cohort studies.52,53

Case-control studies may also yield estimates biased toward the null if controls are recruited from medical facilities.54,55 As smoking, DM, alcoholism and low BMI all place individuals at risk for many illnesses other than TB, controls selected from these settings should be more likely to have these exposures than healthy community members. Although this is a well-known problem with control selection, six of 15 case-control studies on smoking and TB and three of seven on diabetes and TB selected 'medical' controls.26,30 Household controls present a different set of methodological challenges.56 Although household controls are likely to have been exposed to TB, they are also more likely to share clustered risk factors such as DM and alcoholism or, in the case of smoking, to be exposed by proxy to secondhand smoke even if they are not smokers themselves. Interestingly, children exposed to secondhand smoke were at even higher risk of TB than smokers, and their use as controls might even reverse the real direction of the effect.57,58

Another potential problem with inference from epidemiological studies of TB stems from the tools used for diagnosis of the outcomes. TB infection is diagnosed by measuring the host’s immune response, either through the tuberculin skin test (TST) or by measuring the release of interferon-gamma from host T-cells exposed to Mycobacterium tuberculosis antigen.59 As many risk factors for TB reduce the host’s immune response, they may also lead to a false-negative test for infection, systematically biasing results. For example, one study on TB infection among smokers found a dose-response effect with smaller TST induration sizes associated with higher tobacco intake.60 Misdiagnosis of TB disease may also bias results toward the null; many studies use relatively non-specific methods, such as chest radiography, to classify TB, and almost certainly overdiagnose disease. When investigators used meta-regression to estimate summary effect estimates stratified on the diagnostic methods used, they found that studies that used the more accurate bacteriologic methods reported higher odds ratios for both DM and smoking as a cause of TB.26,30

These limitations emphasize the need for carefully designed prospective studies that address the temporal sequence of exposures and outcomes, minimize potential bias introduced by the selection of controls, and select diagnostics that are unlikely to be affected by the determinants being investigated. Such studies will need to recruit participants and measure exposures before the occurrence of the outcome being studied, and will need to implement systematic procedures for diagnosis of infection and disease and relapse. As TB is a rare outcome even in very high-burden settings, such studies may need to identify a cohort at especially high risk, such as household contacts of those with active disease, and follow this group longitudinally through time.
MODELING THE IMPACT OF DETERMINANTS AND THE EFFECT OF INTERVENTIONS

One way to assess the population-level impact that determinants and their modification may have on the TB burden is to use a population attributable fraction (PAF) model to estimate the proportion of the TB burden that would be averted if exposure to a risk factor were eliminated from the population. The PAF describes the proportion of all cases in a population that can be attributed to a specific exposure, and is given by the expression:

\[
P(\text{RR} - 1) + 1
\]

where \( P \) is the prevalence of the risk factor and RR is the relative risk of the outcome conditional on the exposure.

It is important to note that the PAF model assumes that the relative risk of disease given an exposure captures a causal association between the risk factor and the outcome, rather than an association that has occurred either by chance or through the association of the risk factor and the outcome with a third factor that is causal. As TB is caused by more than one determinant, and many individuals are affected by more than one of these factors, the PAF often sums to more than 1. The PAF model has been used widely to estimate the contribution of specific risk factors to the burden of non-communicable diseases, and more recently, of TB.\(^{12,61,62}\) Lönnroth et al. used country-level data from 22 high-burden TB countries to estimate the proportion of active TB that could be attributed to malnutrition (PAF = 27%), IAP (PAF = 22%), active smoking (PAF = 21%), alcohol abuse (PAF = 13%), HIV infection (PAF = 16%), and diabetes (PAF = 10%).\(^{11}\) Similarly, Stevenson et al. used age- and sex-specific data on diabetes prevalence and TB incidence in India to estimate that 20.2% of the cases of smear-positive TB notified in 2000 could be attributed to diabetes.\(^{63}\) Such models rely heavily on the validity and homogeneity of summary effect estimates for specific determinants and on the availability of accurate assessments of the prevalence of risk factors in modeled populations.

A major but implicit premise of the PAF model is that the occurrence of disease is assumed to be independent among individuals, i.e., the disease in one individual is not the consequence of disease in another. While this generally holds for non-communicable diseases, the contagious nature of an infectious disease such as TB clearly violates the independence assumption. In the context of the attributable disease burden, removing a determinant not only prevents primary cases that are directly caused by that risk factor, but also prevents secondary cases that arise indirectly when the primary case goes on to infect others, including some in whom the risk factor may not be present.\(^{64,65}\) Failure to account for the indirect effect of risk factor reduction on preventing secondary cases will underestimate the attributable risk of a determinant and the potential impact of an intervention.

The mathematical modeling of epidemics provides an alternative approach to estimating the impact of determinants and interventions, one that addresses the problem of the potential indirect effects. The most basic of such models places the population into mutually exclusive compartments and then specifies the rates at which people transition between these states. TB models generally build on a simple ‘SIR’ approach, in which everyone in a population falls into one of three groups: ‘susceptibles’ who are at risk for infection; ‘infecteds’ who can spread infection to susceptibles; and ‘recovered’, those who have acquired immunity to infection and are no longer at risk.\(^{46}\) As tuberculosis also includes a latent state from which people can reactivate an earlier infection, the SIR model is usually modified to capture the features of the natural history of the disease that are most relevant to its epidemic behavior.\(^{4,67}\) Transitions between compartments are described by parameters such as the force of infection, the rate of progression to primary disease, the rate of reactivation, the recovery rate conditional on treatment and the disease-specific death rate. The incidence rate is modeled as a function of the model parameters and also of the prevalence of infectious TB cases at any given time; this approach consequently avoids the independence assumption. The Figure portrays one such basic model and illustrates the various transitions that might be affected by individual risk factors or determinants. Other, more complex models incorporate greater realism by explicitly including the age and social structure of populations, spatial heterogeneity and the distribution of risk factors in the population.\(^{68-70}\) These models typically either include an increasing number of compartments or represent the population at the level of individuals rather than compartment, and are consequently much more computationally intensive than their more basic counterparts.

While epidemic models of TB have often assessed the potential impact of intervention strategies, few have specifically focused on the potential impact of modifying TB determinants. In one such study, the authors assessed the impact of smoking and IAP reduction strategies on future trends of TB incidence in China.\(^{71}\) This analysis specified a TB model and then explicitly incorporated parameters obtained from previous systematic reviews that characterized how smoking and indoor air pollution altered TB transitions in the model. The model used site-specific information on the temporal trends in the prevalence of TB and the availability of other control measures such as DOTS. It then assessed the impact of specific smoking reduction scenarios (moderate control by tax and regulation vs. more aggressive tactics) by projecting their impact on smoking trends. This model showed
that although TB incidence was projected to decline under a range of smoking reduction scenarios, the impact of smoking reduction was expected to be greater in those areas where DOTS implementation was sub-optimal. As the actual strategies for tobacco control and the expected future uptake of DOTS in China is not known, modeling was used here as a way to explicitly formalize assumptions and as a tool to identify important knowledge gaps that need to be better understood in future research.

When multiple determinants are considered simultaneously, the estimation of their impact becomes increasingly complex. First, as discussed above, the determinants are often correlated with each other. For example, tobacco smoking is often associated with heavier use of alcohol. To estimate the impact of modifying one of several correlated risk factors, modelers need information on the joint distribution of these determinants and how a change in one is expected to affect the others. Second, the effect of one determinant might be mediated through another. For example, while increases in BMI were found to predicted against TB, they also increase the risk for diabetes, itself an independent risk factor for TB. Third, the effect of one determinant might be modified by another. For example, the effect of smoking on TB might differ depending on exposure to IAP, if these risk factors act through similar mechanisms. Although methods exist to address some of these issues for PAF modeling on non-communicable diseases, further work is needed on these important problems for both the PAF and epidemic modeling approaches to infectious diseases.

CONCLUSION

In this article, we reviewed the methodological issues in modeling the impact of social determinants of TB and their modification. As some of these issues reflect the limitations of currently available epidemiologic studies on the associations between determinants and TB, we expect that in the future better-designed cohort studies may provide better evidence on which to base projections. Other problems result from the complex natural history of TB and the intricate interplay among the determinants and their causal relationship with TB. Modeling provides an important tool for exploring the impact of such interaction, and for gaining insights into the potential impact of social determinants on TB dynamics and the development and evaluation of strategies designed to address these risk factors and improve TB control globally. Future research priorities include well-designed cohort studies to assess the impact of determinants, the integration of single and multiple determinants into dynamic epidemic models, the linkage between epidemiologic studies of TB determinants and dynamic TB models, and the estimation of population-attributable fractions within the context of dynamic models.

References

Les modèles mathématiques ont facilité notre compréhension de la dynamique des maladies infectieuses et se sont avéré des outils utiles pour comparer les scénarios de lutte lorsque des études internationales ne sont pas réalisables ou pas éthiques. Dans ce travail, nous avons résumé les preuves reliant les déterminants socio-économiques et biologiques à la tuberculose (TB) et fait la revue des approches modélisées qui ont été utilisées pour comprendre leur contribution aux dynamiques épidémiques de la TB. Spécifiquement, nous avons trouvé des preuves d’association entre le fait de fumer, la pollution de l’air intérieur, le diabète sucré, l’alcool, l’état nutritionnel, la surpopulation, la migration, les progrès de l’âge et les tendances économiques et, d’autre part, l’apparition d’une infection et/ou d’une maladie TB. Nous délimitons certains problèmes méthodologiques inhérents à l’étude de ces associations ; ceux-ci comportent des problèmes de schéma d’études, d’inversion de causalité et d’erreurs de classification portant à la fois sur l’exposition et sur les résultats. Nous poursuivons par la revue de deux approches existantes pour la modélisation de l’impact des déterminants et de l’effet des interventions : le modèle de réaction attribuable à la population (PAF) qui estime la proportion du fardeau de TB qui serait éliminé dans une population et, d’autre part, les modèles épidémiques déterministes qui saisissent la dynamique de transmission et les effets indirects des interventions. Nous concluons en définissant les priorités de recherche à la fois pour l’étude des déterminants spécifiques et pour l’élaboration de modèles appropriés à l’évaluation de l’impact qu’aurait le fait de prendre ces déterminants en compte.

Los modelos matemáticos han facilitado la comprensión de la dinámica de las enfermedades infecciosas y se ha demostrado que son instrumentos útiles en la comparación de las estrategias de control, cuando los estudios de intervención no son factibles o son inaceptables desde una perspectiva ética. En el presente artículo se resumen los datos existentes que establecen un vínculo entre los determinantes sociales, económicos y biológicos y la tuberculosis (TB) y se examinan los estudios de simulación con modelos que se han utilizado con el objeto de comprender la contribución de estos factores a la dinámica de las epidemias de TB. En concreto, se demostró una asociación entre la aparición de infección o enfermedad tuberculosa y el tabaquismo, la contaminación del aire de interiores, la diabetes, el alcoholismo, el estado nutricional, el hacinamiento, las migraciones, el envejecimiento y las tendencias económicas. Se describen además algunos problemas metodológicos inherentes al estudio de estas asociaciones, que incluyen aspectos metodológicos, la causalidad inversa y la clasificación errónea de la exposición y de los resultados. Se examinaron luego dos estrategias existentes de modelización del impacto de los determinantes y del efecto de las intervenciones: el modelo del factor de riesgo atribuible a la población, en el cual se calcula la proporción de morbilidad por TB que se evitaría en caso de suprimir la exposición a un factor de riesgo en la población y los modelos deterministas de las epidemias que captan la dinámica de transmisión y los efectos indirectos de las intervenciones. Por último, se concluye con la definición de las prioridades de investigación en el estudio de los factores determinantes específicos y en la creación de modelos apropiados de evaluación de las medidas que responden a estos factores.