Epidemics 18 (2017) 16–28

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Predicting lymphatic filariasis transmission and elimination dynamics using a multi-model ensemble framework

Morgan E. Smith[a], Brajendra K. Singh[a], Michael A. Irvine[b], Wilma A. Stolk[c], Swaminathan Subramanian[d], T. Déirdre Hollingsworth[b,e], Edwin Michael[a,*]

[a] Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA
[b] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[c] Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
[d] Vector Control Research Centre (Indian Council of Medical Research), Indira Nagar, Pondicherry 650 006, India
[e] Mathematics Institute, University of Warwick, Gibbet Hill Road, CV4 7AL Coventry, UK

ARTICLE INFO

Article history:
Received 9 December 2016
Received in revised form 1 February 2017
Accepted 1 February 2017

Keywords:
Neglected tropical disease
Lymphatic filariasis
Macroparasite dynamics
Multi-model ensemble
Model calibration and validation
Control dynamics

ABSTRACT

Mathematical models of parasite transmission provide powerful tools for assessing the impacts of interventions. Owing to complexity and uncertainty, no single model may capture all features of transmission and elimination dynamics. Multi-model ensemble modelling offers a framework to help overcome biases of single models. We report on the development of a first multi-model ensemble of three lymphatic filariasis (LF) models (EPIFIL, LYMFASIM, and TRANSFIL), and evaluate its predictive performance in comparison with that of the constituents using calibration and validation data from three case study sites, one each from the three major LF endemic regions: Africa, Southeast Asia and Papua New Guinea (PNG). We assessed the performance of the respective models for predicting the outcomes of annual MDA strategies for various baseline scenarios thought to exemplify the current endemic conditions in the three regions. The results show that the constructed multi-model ensemble outperformed the single models when evaluated across all sites. Single models that best fitted calibration data tended to do less well in simulating the out-of-sample, or validation, intervention data. Scenario modelling results demonstrate that the multi-model ensemble is able to compensate for variance between single models in order to produce more plausible predictions of intervention impacts. Our results highlight the value of an ensemble approach to modelling parasite control dynamics. However, its optimal use will require further methodological improvements as well as consideration of the organizational mechanisms required to ensure that modelling results and data are shared effectively between all stakeholders.

1. Introduction

There is increasing appreciation that large-scale parasite control or elimination problems belong to a decision and policy domain marked by significant uncertainty, complexity, and spatial heterogeneity (Vespignani, 2012; Klepac et al., 2013; Marathe and Vullikanti, 2013; Bhatt et al., 2015; Heesterbeek et al., 2015). Solving these problems is particularly germane for the current global strategies aiming to eliminate complex vector-borne macroparasitic diseases, such as lymphatic filariasis (LF), which exhibit a high degree of geographic heterogeneity in transmission dynamics and infection patterns, and consequently in extinction dynamics (Gambhir et al., 2010; Irvine et al., 2015; Michael and Singh, 2016; Duerr et al., 2005; Singh and Michael, 2015; Jambulingam et al., 2016; Stolk et al., 2006; Swaminathan et al., 2008). Although mathematical models of transmission can capture many features of these complexities, it is recognized that any single model may be inadequate to fully explore and predict the whole spectrum of system behavior (Oreskes et al., 1994; Neuman, 2003). This is partly a consequence of the inherent complexity of natural systems that give rise to multiple conceptualizations and mathematical descriptions (Oreskes et al., 1994). It is also a reflection of the fact that many different model structures and parameter sets can acceptably reproduce the observed behavior of a complex dynamical system, such that model acceptance in one or more settings may not constitute evidence for general model applicability (Beven and Freer, 2001; Ramin et al., 2012; Hoeting et al., 1999; Christakos,

* Corresponding author at: 349 Galvin Life Science Center, University of Notre Dame, Notre Dame, IN 46556, USA.
E-mail address: emichael@nd.edu (E. Michael).

http://dx.doi.org/10.1016/j.epidem.2017.02.006

2003). Indeed, it is increasingly realized in this context that even if approaches based on single models are able to explain the observed behavior of a dynamical system for a given set of data, such models may not generalize well enough to predict future system behavior, particularly under changed conditions – constituting the so-called "out of sample" problem (Simidjievski et al., 2015a, 2015b). Taken together, these uncertainties mean that relying upon forecasts or future predictions generated by a single model for parasite management can lead to significant bias and uncertainty in policy decisions (Lindström et al., 2015).

Recognizing that there may not be a true model of a natural dynamical system, but rather several adequate descriptions reflecting different conceptual bases and structures (Reichert and Omlin, 1997), recent work has focused on using ensemble-based approaches to explicitly account for the uncertainty inherent in the model selection process (Hoeting et al., 1999; Raftery et al., 2005; Gal et al., 2014). Thus, a single-model ensemble involves the use of a number of realizations of an individual model, with distinct predictions obtained for each realization by either introducing stochastic elements, perturbing the input data or initial conditions, or selecting different sets of model parameters (Gal et al., 2014; Viney et al., 2009). By contrast, in a multi-model ensemble, several different models are used, wherein rather than picking the single "best-fitting" model to predict responses, the aim typically is to provide some averaged prediction from different models using various combinatory methods (Hoeting et al., 1999; Raftery et al., 2005). Multi-model ensemble studies, in applications ranging from weather forecasting to cell and population dynamics modelling (Hoeting et al., 1999; Simidjievski et al., 2015a, 2015b; Raftery et al., 2005; Kuepfer et al., 2007), have highlighted the utility of this approach to significantly overcome the problems of over-fitting and model uncertainties, resulting in significant predictive performance gain by these models as compared to that of a single model. Further, studies have shown that even if a multi-model ensemble may not always be the most skillful, its performance is better than the worst single model case, and, as it is often also not possible to predict which of the constituent single-model ensembles will be worst at a given time and location, the use of multi-model ensembles is highly advantageous (Matsueda et al., 2007).

Despite the increasing success of the use of multi-model ensemble methods in other research fields, their application to epidemiological modelling has thus far been limited. However, recent developments in comparing outputs of different influenza models by the MIDAS network (Halloran et al., 2008), assessment of different vaccination strategies (Smith et al., 2012) and impacts of long-term changes in climatic conditions (Ruiz et al., 2014) for malaria, and ensemble-based predictions of Foot and Mouth Disease (FMD) epidemics (Lindström et al., 2015), point to the growing application and value of the method to infectious disease modelling. This body of work demonstrates how combining multiple models can be used to answer critical questions in epidemiology, ranging from the provision of greater confidence in health outcome predictions to improving the ways disease models inform disease control policy, suggesting that the epidemiological use of ensemble-based models are only going to increase in the future.

In this paper, we describe the construction and evaluation of an ensemble of three well-known simulation models of LF epidemiology that incorporate different modelling approaches (deterministic *versus* stochastic), structures (population *versus* individual-based) and parameterization methods (Gambhir et al., 2010; Irvine et al., 2015; Jambulingam et al., 2016; Chan et al., 1998; Norman et al., 2000; Plaisier et al., 1998; Subramanian et al., 2004a; Stolk et al., 2008), in order to better describe the population dynamics of LF and generate more accurate predictions of the impacts of drug and vector-based interventions in communities. The following sections describe the ensemble modelling procedure, analyze prediction accuracy of the single models as well as the multi-model ensemble, and assess the constructed ensemble model's performance in predicting the population dynamics of LF and the outcomes of various intervention strategies on infection. We end by discussing future work to enhance the ensemble model system for supporting policy-relevant predictions, including potential technical improvements in ensemble construction, and the international coordination mechanisms which will be required to link the system effectively to LF data and to policy making.

2. Methods

2.1. The models

The three single LF models that make up this study are: EPIFIL, LYMFASIM and TRANSFIL, which are a Monte-Carlo population-based deterministic (EPIFIL), and stochastic individual-based (LYMFASIM, TRANSFIL) models. These models thus differ in complexity from being individual to population-level based, but also in the overall number of parameters used, and in parameterization methods followed. There are also other more subtle differences among the models, including how effects of infection aggregation are handled, and how drug and vector control are incorporated (Gambhir et al., 2010; Irvine et al., 2015; Michael and Singh, 2016; Singh and Michael, 2015; Jambulingam et al., 2016; Stolk et al., 2006; Swaminathan et al., 2008; Norman et al., 2000; Plaisier et al., 1998; Subramanian et al., 2004a; Michael et al., 2004; Singh et al., 2013; Plaisier et al., 2000). These primary inter-model differences are summarized in Table 1, while Table S4 in part B of the Supplementary Information (SI) captures the key similarities and differences in terms of the model parameters used and optimized during model induction and data fitting, and in running simulations of interventions using annual mass drug administrations (MDAs) and vector control. The full details of the three models and their implementation and fitting procedures for LF infection data have been described extensively previously (Gambhir et al., 2010; Irvine et al., 2015; Jambulingam et al., 2016; Swaminathan et al., 2008; Chan et al., 1998; Norman et al., 2000; Plaisier et al., 1998; Subramanian et al., 2004b), and are summarized in part A of the SI.

2.2. Experimental setup

We employed an experimental design in which each LF model was prepared, calibrated and operated by the respective modelling group, following which the relevant simulation outputs from each single model were provided for use in constructing the multi-model LF ensemble. This experimental setup comprised the following steps. First, the three groups were provided with LF baseline infection and post-intervention data from three community sites chosen to represent the vector-mediated transmission dynamics specific to each of the three major LF endemic regions of Africa (primarily *Anopheles*-mediated transmission), Papua New Guinea – PNG – (*Anopheles*) and India (*Culex*) (Singh et al., 2013; Njenga et al., 2008; Rajagopalan et al., 1989; Subramanian et al., 1989; Das et al., 1992; Rajagopalan et al., 1988) (Table 2A and B). The groups were asked to calibrate their models to the baseline microfilariae (mf) age-prevalence data (="training" data) from these sites, and to provide an ensemble of simulations for the construction and analysis of the multi-model ensemble. Each model aimed to generate 500 fits, or model members, but the number of initial simulations drawn by each group varied from 10,000 (LYMFASIM) to 200,000 (EPIFIL) as a result of differences in the fitting procedures followed and computational intricacies of the three models (see part A of the SI). We deem these as single-model ensembles, which are calibrated and