

**Original citation:**

Hollingsworth, T. Déirdre and Medley, Graham. (2017) Learning from multi-model comparisons : collaboration leads to insights, but limitations remain. *Epidemics*, 18. pp. 1-3.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/87030>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

**A note on versions:**

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)



## Learning from multi-model comparisons: Collaboration leads to insights, but limitations remain



T.D. Hollingsworth<sup>a</sup>, G.F. Medley<sup>b</sup>

<sup>a</sup> Zeeman Institute, University of Warwick, Coventry CV4 7AL, United Kingdom

<sup>b</sup> Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

Neglected tropical diseases (known as NTDs) are a group of diseases predominantly affecting the poorest populations of the globe (sometimes called the ‘bottom billion’). The risk of disease is related to poor housing, poor sanitation and poor health systems as well as the environmental suitability of tropical areas. The NTDs are not a well-defined group, and comprise a variety of pathogens with different transmission routes, life cycles and behavioural risk factors, although they are similar in that they are currently relatively hard to diagnose. The host population factors and relatively poor surveillance data present particular challenges for providing transmission dynamics models. This collection reflects the current state-of-the-art for modelling NTDs, as well as judging on the suitability of models to provide quantitative policy advice.

The World Health Organization (WHO) lists 17 diseases as NTDs, and in 2012 produced a report posing ambitious targets for the eradication, control and elimination of the burden of these diseases as public health problems – often defined as prevalence or incidence below particular thresholds (WHO, 2012). Ten of these targets were explicitly supported by a number of stakeholders, including un-national and multi-national donors, pharmaceutical companies, national and international health bodies and researchers in the London Declaration on NTDs (London, 2012). Here we focus on nine of these diseases, excluding Guinea worm, for which modelling was not considered necessary to ensure eradication.

At the time of the London Declaration it was realised that there was a need for epidemiological modelling to inform many of the key questions regarding interventions and targets. Most of the NTDs have been modelled, but the efforts have been of varying quality, partly due to limited data and information on key biological factors, such as the incubation period or relationship between infection and symptoms (reviewed in two issues of *Advances in Parasitology* (Basanez and Anderson, 2016; Basanez and Anderson, 2015)). However, to support the development of models to address policy needs, and to provide a mechanism to support and strengthen NTD modelling efforts, the NTD Modelling Consortium was formed as an international network of epidemiological modellers selected by independent scientific review (Hollingsworth et al., 2015). This consortium contained at least two modelling groups for each disease, in order to provide more robust scientific insights, and to ensure that previously conflicting guidance could be discussed and

some consensus opinion be given. The individual models have been published in a special issue of *Parasites and Vectors* (Hollingsworth et al., 2015). The purpose of the current collection is to present the comparisons between the models, resulting in 9 disease-specific papers. Additionally, there is one paper that provides modelling insight into the problem of non-adherence to mass treatment. Each paper presents the combination of models and data, in keeping with the ethos of *Epidemics*, and we are very grateful to the journal, and particularly the reviewers, for the rapid turn-around and processing.

### 1. Multiple model comparisons: the basis of robust policy support

Mathematical models of infectious disease have a long history, and are now coming to fruition in terms of their ability and position in policy making. Certainly in the UK, and in other administrations, it would be highly unlikely for decisions about major infectious disease interventions to be made without evidence of effectiveness and cost-effectiveness derived from quantitative analysis and mathematical models, and global strategy to control NTDs should be no different. But a model is only a model, and there is always a concern that relying on one model is insufficient – a different (equally valid) model, might give different evidence. There is a need to provide robust evidence to decision-makers that is based on understanding of transmission dynamics, and is not (individual) model dependent. Consequently, multi-model comparisons are increasingly regarded as a standard, and the basis of robust policy support (see for example: Okell et al. (2015); Eaton et al. (2012); Smith et al. (2012); Johansson et al. (2016)).

Key to the relationship between modelling and policy is the question of how to handle uncertainty. Uncertainty arises from multiple sources, but the two principal sources are parameter uncertainty (i.e. uncertainty in results due to variability in parameters) and structural uncertainty (i.e. uncertainty in evidence derived from different model structures). Stochastic uncertainty arising from stochastic models provides further complication. Whilst the methods for estimating and presenting parameter uncertainty are well advanced, the same is not true for structural uncertainty. For example, it is not clear how many models are

required. Given a set of competing models, there are various methods for deciding which is “better” (e.g. [Touloupou et al. \(2015\)](#)), but it is not clear whether or how models should be weighted to give a combined result. A further complication is the relationship between models and data. There is agreement that validation of a model derives from its ability to forecast data that are not included in its fitting, and in this collection all the authors used (with some exceptions) the same data to fit models, and then forecasted the same data to compare and validate the models. But beyond that there are a wide variety of techniques described. For NTDs in particular, due to the limited number of high quality datasets, the ability of a modelling suite to replicate or be validated against one or two, potentially unrepresentative datasets, does not necessarily validate the extrapolation of these models to multiple settings.

This collection is comprised of research papers, in which comparisons between models are made to expose their differences, and to estimate structural uncertainty. But for policy purposes, decision-makers prefer less uncertainty, so there is a political pressure to reduce structural uncertainty, and models tend to converge to the same answers. Whilst this is not a feature of the current collection, it is something that should be guarded against.

## 2. Data: sources of information for models

This collection highlights that the NTDs, and consequently the modelling efforts, vary in terms of data availability. Some authors have used routine surveillance data ([Blok et al., 2017](#); [Pinsent et al., 2017](#); [Rock et al., 2017](#); [Truscott et al., 2017](#)), whereas others have used more detailed data from epidemiological studies ([Coffeng et al., 2017](#)). Either way, the available data is not usually sufficient to distinguish between different life history assumptions and therefore these papers do not seek to define those life history details, but rather investigate the impact of their assumptions on policy-related forward projections. A general conclusion is that current surveillance data alone are often not sufficient to provide useful forward projections for policy planning.

The London Declaration NTDs are generally divided in two groups, those that are controlled by mass drug administration (MDA), and those that are controlled through intensified disease management (IDM), i.e. through increased case detection and treatment ([Hollingsworth et al., 2015](#)). This distinction also characterises the data which are collected as part of routine surveillance, and this is the way that we characterise the papers to serve as an introduction to the collection.

### 2.1. Cross-sectional prevalence survey data & mass treatment (MDA)

Infections which are controlled by MDA are generally monitored by cross-sectional surveys of prevalence in selected sites between treatment rounds. For many of these diseases surveys are conducted every year, but for blinding trachoma in many sites the data collection is much sparser ([Pinsent et al., 2017](#)).

For the soil-transmitted helminth (STH) study, the authors fitted to detailed age-specific data from an epidemiological study of hookworms and roundworms (*Ascaris lumbricoides*) ([Coffeng et al., 2017](#)), whereas for schistosomiasis, surveillance data in children (the age-group being treated) were used ([Truscott et al., 2017](#)). For both comparisons, the short-term dynamics were similar, but longer-term dynamics were highly dependent on model assumptions. For schistosomiasis in particular, there were very different assumptions on worm distributions and dynamics in the snail host. Similarly for STH, the effect of age-dependent transmission rates in different demographics and egg survival dynamics can currently only be investigated through modelling studies as they are noto-

riously difficult to observe directly. However, novel methods of model-led data collection could help identify key measurements which could at least rule out some of the potential scenarios.

For onchocerciasis and lymphatic filariasis, there are similar uncertainties regarding some aspects of the dynamics of worms within and between hosts as for STH and schistosomiasis, but the epidemic dynamics are much slower, meaning that drops in prevalence between treatment rounds are more closely related to coverage and efficacy of drugs. The longer-term dynamics are, of course, highly sensitive to underlying model assumptions. In contrast to schistosomiasis and STH, onchocerciasis and lymphatic filariasis models have been fitted to multiple-timepoint epidemiological data for many years, and therefore have similar conclusions on the impact of repeat treatment with a known coverage. The lymphatic filariasis multi-model comparison (including two individual-based models and a deterministic model) highlighted the short-term similarities in predictions, but longer-term uncertainties remain ([Smith et al., 2017](#)). The onchocerciasis modellers analysed long-term datasets, and so were able to investigate elimination dynamics ([Walker et al., 2017](#)). In particular they identified a key aggregate output of their model assumptions, the relationship between the annual biting rate and equilibrium prevalence, as a key driver of discrepant results on elimination.

The remaining MDA infection studied here was blinding trachoma, caused by an ocular bacterial infection. Some existing models have been validated against epidemiological data, and so the challenge for this group was to use routine surveillance data, which are extremely infrequent prevalence surveys. The authors analysed 7 deterministic, mechanistic models and statistical models forecasting between 7 and 14 years in the future ([Pinsent et al., 2017](#)). By combining these different modelling approaches they highlighted the strengths and weaknesses of these different approaches, and the need for combined efforts such as these.

In this part of the collection we also include a cross-cutting article which investigates assumptions regarding how MDA campaigns cover the population ([Dyson et al., 2017](#)). The authors review different modelling frameworks and propose a new mechanism which captures many of the existing frameworks, showing that only a small amount of additional surveillance data could help identify the extent to which the programme is at risk of not reaching its goals. There is a need for similar studies to consider the elimination dynamics under different assumptions regarding heterogeneities, migration and dynamics at different spatial scales.

### 2.2. Rates of case detection

Mass drug administration is not available for four of the NTDs considered: leprosy, human African trypanosomiasis (HAT), visceral leishmaniasis (VL) and Chagas disease. Three of these diseases are controlled by IDM. For HAT, active surveillance is used to seek out potential infections, whereas leprosy and VL rely on individuals presenting themselves for diagnosis: passive surveillance. Because of active surveillance, the HAT modellers had the most data available, but this meant that they were more selective in which data they incorporated in the comparison ([Rock et al., 2017](#)). The VL modellers were confronted with seasonality, and demonstrated that more data gives a better fit, but perhaps more importantly that case data alone (especially over a short time period) were insufficiently informative to enable an accurate assessment of structural uncertainty ([Le Rutte et al., 2017](#)). A key challenge is to decide which minimal additional data are required. The leprosy modellers used the widest variety of model structures, from stochastic individual-based simulations to purely statistical approaches ([Blok et al., 2017](#)). Perhaps surprisingly the models agreed in terms of broad outcomes, and where they differed in detail, these differences were explainable. For leprosy at least, the work presented

here shows that structural uncertainty is not a barrier for applying models to policy.

### 2.3. Serological surveys

Chagas disease is perhaps the most awkward of the NTDs considered here. Human infection is a spill over from animal transmission cycles and patterns of disease are largely determined by exposure to vectors (triatomine bugs) which occurs within households. As far as we are aware, there are no large-scale data which take into account the small spatial scales required for accurate modelling of transmission dynamics, such as within- and between-household dynamics. The Chagas disease modellers instead used a time series of national, cross-sectional, sero-surveillance data to estimate changes in infection rates over time, but acknowledged that the models and data are not at the same scale (Bartsch et al., 2017). Such data are becoming increasingly common in NTD surveillance, but the sampling strategies need to be designed to provide the right information for models.

## 3. Characterising uncertainty

This collection is a unique set of comparisons of deterministic, stochastic, individual-based mechanistic and statistical models. Traditionally, different modelling groups have presented very different kinds of uncertainty, for example due to uncertainties in the data, parameters, model structure or stochastic variability between simulations, without necessarily explaining which is which, or justifying underlying assumptions. In this collection, we have sought to make clear the drivers of uncertainty in the different models, but clearly this is an area for future improvement.

## 4. Summary

As with any research, the publications are markers of progress rather than the principal aim. If we were advising others beginning a similar process we would highlight (i) that there is not currently a uniformly accepted statistical framework for comparing and combining such diverse models with such diverse datasets, and that this is an area ripe for development, and (ii) the quality of insight gained in these publications are the result of the trust built up between the groups through previous work, allowing them to ‘expose’ their models to each other’s scrutiny. An important outcome has been the increased collaboration between different groups, and the inevitable competition has spurred the improvement of the models generally. We hope that this collection will, in turn, demonstrate the need for high quality, high volume data. Policy-makers can only get the level of advice and support that the data allow, and no amount of modelling can substitute for accurate observation. The example of onchocerciasis demonstrates that a cycle of model development and comparison fuelled by good data can produce a consensus framework on which policy can be sensibly founded, in an area where historically the models have given highly discrepant results on key policy outputs such as the number of years a treatment program would need to be run to achieve the programme goals (Walker et al., 2017; Basanez et al., 2016; Stolk et al., 2015).

It should be noted that the models presented here are missing the dimension of health economics. The current policy impetus is

to achieve defined goals, so that it is the effectiveness of interventions that are being tested. However, as these goals are achieved, so the emphasis will move increasingly towards cost-effectiveness, i.e. how can control/elimination be achieved most efficiently. This will require a novel set of models, building on those presented here, but including the complications of health systems, diagnostics, surveillance systems and delivery systems.

In summary, this collection represents the state-of-the-art for modelling NTD transmission dynamics. Across all nine diseases we provide projections of the outcomes of current interventions that go beyond qualitative predictions. The future focus of the NTD Modelling Consortium will be on continued development and use of these quantitative models to support policy decisions.

## References

- Bartsch, S.M., et al., 2017. Comparison and validation of two computational models of Chagas disease: a thirty year perspective from Venezuela. *Epidemics* 18, 81–91.
- Basanez, M.G., Anderson, R.M., 2015. Mathematical models for neglected tropical diseases: essential tools for control and elimination, part A. *Adv. Parasitol.* 87, xiii–xvii.
- Basanez, M.G., Anderson, R.M., 2016. Preface. *Adv. Parasitol.* 94, xi–xvi.
- Basanez, M.G., et al., 2016. River blindness: mathematical models for control and elimination. *Adv. Parasitol.* 94, 247–341.
- Blok, D.J., et al., 2017. Forecasting the new case detection rate of leprosy in four states of Brazil: a comparison of modelling approaches. *Epidemics* 18, 92–100.
- Coffeng, L., et al., 2017. Comparison and validation of two mathematical models for the impact of mass drug administration on *Ascaris lumbricoides* and hookworm infection. *Epidemics* 18, 38–47.
- Dyson, L., et al., 2017. Measuring and modelling the effects of systematic non-adherence to mass drug administration. *Epidemics* 18, 56–66.
- Eaton, J.W., et al., 2012. HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. *PLoS Med.* 9 (7), e1001245.
- Hollingsworth, T.D., et al., 2015. Quantitative analyses and modelling to support achievement of the 2020 goals for nine neglected tropical diseases. *Parasites Vectors* 8, 630.
- Johansson, M.A., et al., 2016. Evaluating the performance of infectious disease forecasts: a comparison of climate-driven and seasonal dengue forecasts for Mexico. *Sci. Rep.* 6.
- Le Rutte, E.A., et al., 2017. Elimination of visceral leishmaniasis in the Indian subcontinent: a comparison of three transmission models. *Epidemics* 18, 67–80.
- London, 2012. London Declaration on Neglected Tropical Diseases.
- Okell, L., et al. Consensus modelling evidence to support the design of mass drug administration programmes. In: Malaria Policy Advisory Committee meeting, 16–18 September 2015, Geneva Switzerland, Background document for Session 1. 2015. World Health Organization.
- Pinsent, A., et al., 2017. Probabilistic forecasts of trachoma transmission at the district level: a statistical model comparison. *Epidemics* 18, 48–55.
- Rock, K.S., et al., 2017. Data-driven models to predict the elimination of sleeping sickness in former Equateur province of DRC. *Epidemics* 18, 101–112.
- Smith, T., et al., 2012. Ensemble modelling of the likely public health impact of a pre-erythrocytic malaria vaccine. *PLoS Med.* 9 (1), e1001157.
- Smith, M.E., et al., 2017. Predicting lymphatic filariasis transmission and elimination dynamics using a multi-model ensemble framework. *Epidemics* 18, 16–28.
- Stolk, W.A., et al., 2015. Required duration of mass ivermectin treatment for onchocerciasis elimination in Africa: a comparative modelling analysis. *Parasites Vectors* 8 (552).
- Touloupou, P., et al., 2015. Model Comparison with Missing Data Using MCMC and Importance Sampling. arXiv preprint arXiv:1512.04743.
- Truscott, J.E., et al., 2017. A comparison of two mathematical models of the impact of mass drug administration on the transmission and control of schistosomiasis. *Epidemics* 18, 29–37.
- WHO, 2012. Accelerating Work to Overcome the Global Impact of Neglected Tropical Diseases – A Roadmap for Implementation. World Health Organization, Geneva.
- Walker, M., Stolk, W., Basanez, M.G., 2017. Modelling the elimination of river blindness using long-term epidemiological and programmatic data from Mali and Senegal. *Epidemics* 18, 4–15.