Can we know in advance whether models will get it right?

Validation of model predictions against real-world data is worthwhile,[1] yet few have been able, or are brave enough, to actually do it. Thankfully, in *The Lancet Global Health,* Jeffrey Eaton and colleagues[2] have stepped up to the challenge, using ten models calibrated to data from 2002–12 to predict HIV prevalence in South Africa in 2012 before the release of estimates from the national household survey done in 2012.

Overall, the models got many of the details correct, such as a shift in HIV burden from younger to older age groups, but got the big picture wrong—predicting stable or declining overall prevalence, whereas prevalence actually increased from 16·9% to 18·8% (difference 1·9, 95% CI –0·1 to 3·9). Only one model predicted a noticeable increase in HIV prevalence towards the level measured in the survey, and the best estimates of only two of the ten models were within the 95% CIs of the 2012 household survey data. This finding raises the sobering question: if we can get model predictions so wrong in the data-rich setting of South Africa, where there are ten leading HIV epidemiological modelling groups focusing their attention, where can we get it right with confidence?

One possible answer is to redefine what is meant by getting it right. Three of the ten models included in this study incorporated uncertainties—credible intervals based on Bayesian analysis—in their results. For most indicators for these three models, the empirical data did fall within the uncertainty bounds of the models. If all models provide wide limits of possible epidemic projections based on all plausible trajectories, which ultimately include what does occur in future findings, then they could be regarded as right, but they would not be very helpful.

It is also possible that models can correctly predict what would have been expected to occur, had unforeseen changes in underlying conditions not affected the epidemic. These conditions could, for example, be political, financial, programmatic, or behavioural in nature. In such circumstances, the models project a counterfactual that can be compared with the actual outcome to assess the effect of the changes in conditions, but in themselves might not be able to be validated. Despite this limitation, this type of epidemiological modelling is the best approach we have for informing decisions about changing conditions at the time when decisions need to be made. Although some conditions inevitably differ over time compared with the assumptions included in models, the dominant relevant condition to have changed in South Africa over the period of projections was the uptake of, and adherence to, effective antiretroviral therapy (ART), which most model predictions were able to closely resemble.

The fact that the models in this study were consistently wrong in their overall predictions of change in prevalence is itself telling. For example, all models assumed that ART was 90% effective at preventing HIV transmission. This assumption was based on empirical estimates in ideal settings that might not accurately represent real-world retention in care and treatment adherence. One possible explanation for the inaccuracies of the models is that their assumed ART effectiveness levels were too high and the real-world effectiveness of ART for reducing transmission is actually substantially lower. This explanation is an example of how models, when calibrated against empirical data, can provide scientific or programmatic insight. It also highlights again that one of the largest questions remaining for the HIV specialty is the amount of real-world prevention effectiveness afforded by ART at various levels of viral suppression.

Adjusting the models to take into account the new data in South Africa, potentially by decreasing the assumed preventive effectiveness of ART until further evidence becomes available, and also calibrating the models against additional sources of data (eg, sex-specific incidence, prevalence, and mortality data) will further refine the precision of the projections for future use and increase the confidence in model outputs for South Africa. The largest implication of the study by Eaton and colleagues[2] is that future resource needs will probably be greater than previously expected. With heavily constrained budgets, this finding highlights the substantial need to achieve efficiency gains—both targeting programmes geographically by population group with the right combination of cost-effective interventions for the epidemic context[3] and particularly through delivery of these services in the most efficient manner.

In other settings with fewer available data, the confidence that can be placed in model outputs is related to the quality of the inputs. However, models are not designed to be authoritatively right in predictions. Rather, despite their limitations, they are instruments for assessment of the available data, often attempting to reconcile several sources of data together, to provide implications, inferences, and further insights with more rigorous predictions from the knowledge base than could be achieved otherwise through simple extrapolation of past trends or speculation.

Important global and public health decisions are made on the basis of knowledge of the magnitude and trends in incidence and prevalence of infection or disease, which can generally only be inferred at a population level through the use of epidemiological models. Therefore, appraisal of the validity of models, particularly through comparisons of model predictions with data that are later reported, is crucial but rarely done. We commend Eaton and colleagues for comparing outputs from different models for the same context and undertaking an honest appraisal of model projections. Their study is a very useful exercise and we hope to see more of these comparison studies in the future.

*David P Wilson, Cliff C Kerr
Kirby Institute, University of New South Wales, Sydney, NSW 2052, Australia (DPW, CCK); and Complex Systems Group, School of Physics, University of Sydney, NSW 2006, Australia (CCK)
dwilson@unsw.edu.au

1   Delva W, Wilson DP, Abu-Raddad L, et al. HIV treatment as prevention: principles of good HIV epidemiology modelling for public health decision-making in all modes of prevention and evaluation. *PLoS Med* 2012; **9:** e1001239.
2   Eaton JW, Bacaër N, Bershteyn A, et al. Assessment of epidemic projections using recent HIV survey data in South Africa: a validation analysis of ten mathematical models of HIV epidemiology in the antiretroviral therapy era. *Lancet Glob Health* 2015; **3:** e598–608.
3   Kerr CC, Stuart RM, Gray RT, et al. Optima: a model for HIV epidemic analysis, program prioritization, and resource optimization. *J Acquir Immune Defic Syndr* 2015; **69:** 365–76.