Modelling sexual transmission of HIV: testing the assumptions, validating the predictions
Rebecca F. Baggaley and Christophe Fraser

Introduction
Mathematical models have played important roles facilitating understanding of HIV epidemiology and evaluating the performance of prevention initiatives [1]. From the earliest models examining the interaction between HIV and other sexually transmitted infections (STIs) [2], the effects of sexual mixing patterns between individuals by age [3] and predicting the future course of HIV epidemics [4], modelling has assisted in making projections [5], explaining past and future trends [6–8], as well as predicting the impact of existing and proposed HIV prevention initiatives [9–11]. Such analyses, in which model input parameters are believed to be estimated with sufficient accuracy, can provide quantitative predictions, often being combined with economic analyses to provide cost–efficacy or cost–benefit projections [12*,13,14]. When such precision is not attainable, modelling can explore more qualitative outcomes, able to open up new directions of enquiry, such as predicting the impact of HIV prevention technologies yet to be developed (such as vaccines and microbicides).

Both qualitative models (used for broad insights) and detailed models (developed for operational purposes) may influence HIV prevention and treatment policies, yet there may also be a lack of trust due to the opaque nature of modelling methods that are used (often quite complex and technical), or conversely, overconfidence and reliance on certain methods or research groups because of lack of understanding of mathematical models in the wider stakeholder community [1]. In this review, we include a case study that has recently received a lot of attention and in which models have been used to influence the research community, policy and beyond: mathematical models of HIV testing and antiretroviral treatment as prevention (‘test and treat’).

From efficacy to effectiveness
Mathematical models have proven especially useful for assessing interventions such as ‘test and treat’ or male circumcision, because their effect is to prevent transmission, and these interventions have individual, pairwise and population level benefits, which are very hard to estimate using empirical field studies alone. Protecting one individual from acquiring infection has an indirect protective effect on others (Fig. 1a). The efficacy of an HIV prevention intervention denotes the degree of protection against infection experienced by one individual benefiting directly from the intervention, such as the protection afforded to a man who is circumcised. Effectiveness of infectious disease interventions is more complex, as it includes the far-reaching population effects of applying the intervention to each of these individuals (as shown by the concept of herd immunity, vaccination of a
fraction of the population provides protection to unvaccinated or otherwise unprotected individuals. The relationship between individual level efficacy and population level effectiveness (Fig. 1a and b) is not straightforward because of the indirect benefits of prevention, but also because people may be exposed to HIV multiple times in a lifetime. The prevention of one new infection could in principle lead to a whole subepidemic being averted, but it could also result in a whole different epidemic due to the same person being infected at a later date. Mathematical models are ideal tools for exploring these complex relationships between different scenarios, and therefore, for relating individual efficacy to population effectiveness in different settings.

There are estimates of efficacy for a whole range of HIV interventions: voluntary counselling and testing and other behaviour change interventions, male circumcision, antiretroviral therapy (ART), preexposure and postexposure prophylaxis and treatment for bacterial STIs, herpes simplex virus (HSV) and other coinfections [14–18]. However, almost no direct estimates of effectiveness exist because sufficiently large community-based trials, which would reflect the indirect protective effect on others as illustrated in Fig. 1a, or community cluster randomized trials that could measure the effects illustrated in Fig. 1b, are increasingly viewed as prohibitively expensive. Mathematical models are, thus, often viewed as the only feasible method for assessing the effectiveness of interventions with known efficacy.

Even measurement of efficacy can be hugely expensive. Endpoints directly reflecting sexual HIV transmission, such as follow-up of discordant couples (for example, Partners in Prevention investigating the efficacy of acyclovir treatment of HSV in reducing the HIV infectiousness of dually infected sexual partners [19]), or measuring HIV incidence or prevalence, provide the most compelling evidence for HIV intervention efficacy. However, their expense means that other endpoints are often used, such as reduction in HIV viral load representing reduced infectiousness [20]. However, care must be taken in the translation from such proxy endpoints to the effect on actual infection transmission as these are extrapolations and subject to wide statistical uncertainty and potentially unknown setting-specific confounders [21].

Difficulties in measuring population-level effectiveness not only include physical scale (community-based trials need to be large) but also time scale. The indirect protective effects of HIV interventions illustrated in Fig. 1 change over time after the introduction of the intervention, as the indirect effects of prevention may accumulate for many years after each directly averted infection. For example, women indirectly benefit from the reduced HIV incidence among circumcised men.

---

**Figure 1** Demonstration of the impact of HIV prevention at the individual and population levels

(a) HIV prevention method directly prevents one transmission event but indirectly prevents an additional three transmissions, which would have also taken place. (b) The population-level effect of prevention methods can be measured through cluster randomized trials or mathematical modelling analyses, which compare scenarios in which no protection is offered (1) to those wherein all those at risk of acquiring HIV infection are protected (3) or some degree of coverage in between (2).
impacting on HIV prevalence. If these women go on to form new partnerships with uncircumcised men, then with time the protective effect of circumcision also goes on to indirectly benefit men who are not circumcised (Fig. 2, taken from Hallett et al. [22]). These benefits increase over time, so where is the cut-off at which point effectiveness should be measured? As benefits are likely to increase with the years, so their measurement becomes less achievable, with increasing difficulty in separating the effect from secular trends in natural epidemic dynamics and the impact of other interventions. In this case, modelling can be used to explore different scenarios, measuring effectiveness over many years and decades [12**,22].

Wherein large, community-based or cluster randomized trials of interventions have been feasible, there is an increasing recognition of the importance of partnership between empirical data gathering and mathematical modelling of HIV transmission. An example of this recognition is the National Institutes of Health Methods for Prevention Packages Program that has funded several multidisciplinary groups, each including mathematical modellers, to design and optimize combination prevention for HIV (see http://grants.nih.gov/grants/guide/rfa-files/RFA-AI-08-019.html for details).

**A focus on test and treat**

In 2009, research by investigators at the World Health Organisation (WHO), modelling a ‘universal test and treat’ strategy for sexual HIV transmission prevention, was published in the *Lancet* [12**]. The strategy involved annual HIV screening and immediate initiation of combination antiretroviral treatment for everyone testing positive in order to control the HIV epidemic in sub-Saharan Africa. The authors predicted that the proposed strategy could move generalized HIV epidemics such as that in South Africa to elimination (defined as HIV incidence less than one case per thousand per year) within 10 years.

The paper is interesting not only for its bold and optimistic approach (it assumes treatment coverage quickly reaches 90% with only a 1.5% per year long-term drop-out rate, no transmission of drug resistance, and substantial reductions in sexual risk behaviour), but also for the magnitude of the reaction it generated [23–32]. It was greeted by a considerable response from the research, policy making and advocacy communities, with supporters as well as opponents. Among the various objections made on the grounds of feasibility and human rights issues were those commenting on the modelling methods; that overly optimistic assumptions were made and that the simple transmission model used could not adequately capture the transmission dynamics in realistic populations.

Granich *et al.* [12**] raise as many questions as they answer. In terms of methodology, their approach returns to very simple modelling structures. The majority of mathematical models of ART and prevention of sexual HIV transmission have, like in the study by Granich *et al.*, been simple compartmental models. In these models, the variables that change over time are the number of people in a typically small number of particular states (e.g. uninfected, early untreated, treated), known as the compartments (Fig. 3a). Infection is governed by a per partnership infectiousness (or a per act infectiousness and number of acts per partnership). Contacts are typically relatively homogeneous (at least within and between the discrete classes) and partnerships are typically not modelled explicitly. These models vary in their complexity by defining fewer or more different types of individuals (and, therefore, fewer or larger numbers of compartments).

Some very generic insights, which remain broadly valid when replicated from more complex models, can be obtained from the simplest transmission models, involving the smallest number of differences between individuals. All heterogeneity is averaged out and they involve just three input parameters: the average number of sexual partners each member of the modelled population has (expressed as the rate of change of partners, often denoted ‘r’), the duration that an individual remains infected with HIV (denoted ‘D’) and the average probability of transmission of HIV for each partnership between an infected and uninfected individual (denoted ‘β’). More complex models divide these three simplest
Figure 3 Schematic illustrations of the structure of HIV transmission models incorporating antiretroviral therapy

(a) Simplest HIV transmission model: model output is dependent on the rate of treatment uptake, \( \gamma \) and the infectiousness \( \beta \), sexual activity \( c \) and duration spent in that compartment \( D \) for infected individuals in the untreated compartment and then the treated compartment. (b) HIV transmission model with further stratifications to incorporate heterogeneity between individuals in the population: stages of infection, treatment failure, treatment withdrawal and heterogeneity in sexual activity (denote by layers of compartments).

Factors into more groups to reflect heterogeneity in these features over time since infection and between individuals (e.g. Figure 3b). For example, in the context of ART, treated individuals have a lower risk of transmitting HIV through sex but live longer, thus reducing \( \beta \) but increasing \( D \). Therefore, the simplest sexual HIV transmission models incorporating ART use involve an average rate of uptake of treatment for all infected individuals, leading to an average reduction in HIV infectiousness, an average increase in life expectancy and potentially also an average change in individuals’ risk behaviours (Fig. 3a). Model output depends on values assigned to these three parameters \( (\beta,c,D) \) for treated and untreated individuals, as well as the rate of treatment uptake \( (\gamma) \). The generic insight obtained from this model is first that treatment can only reduce overall incidence, provided the reduction in \( \beta \) and \( c \) is greater than the increase in \( D \), and second that if this condition is met, then there is a threshold rate of treatment uptake above which transmission is not sustainable and the epidemic can be eliminated. Wherein more complex models are needed is in translating this generic insight into specific quantitative predictions or conditions for success. The simple model cannot be used to generate credible predictions of the magnitude of reduction in transmissibility or risk behaviour, or of treatment uptake, which are required for successful epidemic control – estimates turn out to be strongly dependent on many specifics.

Every transmission model published has involved some extension to the simple generic model to account for heterogeneity in one way or another, which is important because different groups may initiate treatment at different rates and fare differently once on treatment. Many of these heterogeneities are illustrated in Fig. 3b. There may be progression through stages of HIV infection, with
treatment uptake dependent on the stage \((\gamma_1 - \gamma_3)\). Treatment failure and withdrawal from treatment may also be explicitly incorporated, as may stratification of the populations into groups with different levels of sexual activity (layers of boxes for each compartment illustrated in Fig. 3b). Further complexity arises from stratifying by level of adherence, explicitly incorporating rates of diagnosis preceding treatment initiation and development and transmission of resistance.

Models may also be more complex by adopting fundamentally different structures, such as network models, which explicitly simulate each individual and the partnerships they form, rather than relying on the averages of groups of people implicit in the former type of model [33,34]. These models are very effective at describing extremes of heterogeneous sexual behaviour, but also and just as importantly, the effects of local saturation of epidemic spread, which arises when the transmission of infection becomes limited because sexual acts are repeated among subgroups of people who are all infected. Network models are also necessary for modelling interventions that target specific individuals, for example, through contact tracing, which may help quickly identify individuals who are contributing disproportionately to the epidemic.

There will always be a need for simple alongside complex model structures because simple models generate generic mechanistic insight, which can then be refined and matched to specific operational data with more complex and hopefully realistic models. However, there is also a need to revisit the fundamental validity of generic insights in the light of extra detail. For example, for test and treat, Dodd et al. [35*] recently explored the model predictions of Granich et al. [12**], but further adding sexual heterogeneity, with a core group of highly sexually active individuals, and concluded that the impact of universal test and treat would be highly dependent on the epidemiological context. Additionally, many may argue that veracity of conclusions by Granich et al. [12**] may be limited by not explicitly modelling transmission of drug-resistant HIV [36]. Drug resistance may not be a simple refinement of the model, but could, in some scenarios, completely govern the success or failure of any test and treat strategy. Thus, in this case, better parameters on sexual risk behaviour and drug resistance could result in qualitatively different predictions, not just refined versions of the predictions based on the simplest model [12**].

Despite the attention that Granich et al. [12**] received, ART for prevention is not a new idea. A series of high-profile papers were published when triple ART was in its infancy, which explored strategies involving treatment of individuals relatively early in their infection, with high coverage levels, but for industrialized country contexts [37,38]. Since then, many models have explored the potential impact of ART in resource-poor settings [39–42], yet their more modest assumptions regarding ART coverage did not provoke a significant response. In contrast, the high coverage assumed for the WHO authors’ strategy and questions regarding feasibility of such large-scale roll-out implicit for regions with such high HIV prevalence led to substantially more attention and criticism. Yet this is the kind of exercise to which mathematical modelling is so well suited: hypothetical scenarios involving blue skies thinking which can broaden and provoke debate, and hopefully also lead to empirical data gathering and feasibility testing of the most promising ideas.

No single modelling analysis can tell us which test and treat strategies could be successful and which could not, and Granich et al. [12**] state that better data are required for model parameterization. Since the introduction of triple ART in 1996, more than 30 articles have been published presenting results from HIV transmission models exploring the impact that ART has had or will have on HIV transmission (incidence, prevalence, infections averted) in various settings [10,37–69]. Whereas earlier models focussed on predicting ART impact in industrialized settings [37,38,46,55,56,61], more recently, many models have explored the impact in sub-Saharan Africa [42,48,50,52,57], as ART scale-up has become a reality. Although many of these restricted ART coverage to those with late-stage HIV infection [41,42], thought realistic because individuals in resource-poor settings are typically diagnosed so late [70,71], the sudden surge in ART funding from the U.S. President’s Emergency Plan for AIDS Relief (PEPFAR) and other organizations after initiatives such as the WHO’s ‘3 by 5’ caused all modelling to shift in scope and ambition, as people’s mindsets altered. Treating people at earlier stages of infection has become a realistic prospect, as demonstrated by the WHO’s revised guidelines recommending treatment of all individuals with CD4 cell counts of 350 cells/μL or less.

Granich et al. [12**] have asserted that screening the vast majority of individuals annually would be sufficient to drive HIV epidemics to an elimination phase, as infections would be identified and treated at such an early stage that the window of transmission before this would be too short to sustain transmission. The more infectious, or the more sexually active, an infected individual is, the earlier in their infection treatment must be initiated for this to hold (Fig. 4, taken from Garnett and Baggaley [24]). There has, therefore, been considerable debate on how early treatment must be initiated and on the magnitude of the role played by acute, or primary HIV infection, in ongoing HIV transmission, and on how this depends on the different epidemic contexts [28,72,73].
Conclusion
The different model structures, model assumptions and parameter ranges, as well as choice of model outcomes presented and the time scale over which the impact of ART is measured, make comparison of ART HIV transmission modelling studies difficult. However, the need to understand the implications of these variables requires a formal analysis of techniques and exploration of parameter space (evidence based). The comparison of different models from different research groups, designed to answer the same research question, can facilitate this understanding and has recently been employed to assess the impact of male circumcision [74**].

We have used test and treat as a case study in this review, but for all HIV transmission modelling, there is a need to understand as fully as possible the dependencies in the model predictions and to highlight those aspects of model structure that are imperative to incorporate. It is, thus, vital, whether aiming for broad qualitative or detailed quantitative insights, to systematically investigate the sensitivity of model output to predictions and to present predictions with a full range of uncertainty to both parameters and modelling choices. In this way, models can be objectively compared and important parameters can be identified for further empirical research. Eventually, predictions can then converge to improved levels of realism and reliability.

Acknowledgements
The present work was supported by the Wellcome Trust (IG082623MA to R.F.B.) and the Royal Society (C.F.). We thank the UK Medical Research Council (MRC) for Centre funding. We declare that we have no conflicts of interest.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:
* of special interest
** of outstanding interest
Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 350).

Sexual transmission of HIV


74. UNAIDS/WHO/SACEMA Expert Group on Modelling the Impact and Cost of Male Circumcision for HIV Prevention. Male circumcision for HIV prevention in high HIV prevalence settings: what can mathematical modelling contribute to informed decision making? PLoS Med 2009; 6:e1000109. Collaboration of experts reviewing the outcomes of six mathematical models of the population level impact of male circumcision on HIV incidence in high HIV prevalence settings. Models were from different research groups and involved different methods, assumptions and some input variables, yet results were relatively consistent, predicting that one HIV infection would be averted for every five to 15 male circumcisions performed.