Algorithms for efficiently and effectively matching agents in microsimulations of sexually transmitted infections

Author: Nathan Geffen

Supervisor: Michelle Kuttel

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Computer Science

UNIVERSITY OF CAPE TOWN

March 5, 2018

In memory of Hazel Geffen

Contents

Abstract

Mathematical models of the HIV epidemic have been used to estimate incidence, prevalence and life-expectancy, as well the benefits and costs of public health interventions, such as the provision of antiretroviral treatment. Models of sexually transmitted infection epidemics attempt to account for varying levels of risk across a population based on diverse — or heterogeneous — sexual behaviour. Microsimulations are a type of model that can account for fine-grained heterogeneous sexual behaviour. This requires pairing individuals, or agents, into sexual partnerships whose distribution matches that of the population being studied, to the extent this is known. But pair-matching is computationally expensive. There is a need for computer algorithms that pair-match quickly.

In this work we describe the role of modelling in responses to the South African HIV epidemic. We also chronicle a three-decade debate, greatly influenced since 2008 by a mathematical model, on the optimal time for people with HIV to start antiretroviral treatment. We then present and analyse several pair-matching algorithms, and compare them in a microsimulation of a fictitious STI. We find that there are algorithms, such as Cluster Shuffle Pair-Matching, that offer a good compromise between speed and approximating the distribution of sexual relationships of the study-population. An interesting further finding is that infection incidence decreases as population increases, all other things being equal. Whether this is an artefact of our methodology or a natural world phenomenon is unclear and a topic for further research.

Abbreviations

ART ..... Antiretroviral Treatment

ARV ..... Antiretroviral

ASSA ..... Actuarial Society of South Africa

BFPM ..... Brute-Force Pair Matching

CSPM ..... Cluster-Shuffle Pair Matching

HSRC ..... Human Sciences Research Council

PMTCT ..... Prevention of mother-to-child transmission of HIV

RPM ..... Random Pair Matching

RKPM ..... Random-k Pair Matching

STI ..... Sexually Transmitted Infection

Acknowledgements

Many thanks to my supervisor Michelle Kuttel for reviewing this work and helping me to improve it. Leigh Johnson, Nicoli Nattrass, Edwin Cameron, James Gain, Andrew Boulle and Eduard Grebe all offered important advice. I am especially grateful to my co-authors of some of the articles included in this dissertation: Alex Welte, Marcus Low and Stefan Scholz. The three examiners offered many useful suggestions that have been incorporated in the final text. Thanks also to my husband Faizel Slamang and my colleagues at Community Media Trust and GroundUp.

List of publications

This is a thesis by publication. Table 1 lists the publications included in this dissertation. Several further articles were authored or co-authored by this author in the course of this PhD, but are not included in this dissertation (although most are cited in the publications). These are listed in Table 2.

Title	Status	Journal	Authors	Contributions	Ref
Modelling the HIV epidemic: A review of the substance and role of models in South Africa.	Published	Southern African Journal of HIV Medicine	Geffen, N and Welte, A	Geffen conceived and drafted the article. Welte proposed including the terminology on model world, model and scenario, and drafted this section of the article. The authors reviewed and edited the article together.	Geffen and Welte (2018)
A history of controversies involving HIV models in South Africa (supplementary online material for above article)	Online addendum to above article	Southern African Journal of HIV Medicine	Geffen, N and Welte, A	Geffen conceived and drafted the article. The authors reviewed and edited the article together.	Geffen and Welte (2018)
When to start antiretroviral treatment? A history and analysis of a scientific controversy	Published	Southern African Journal of HIV Medicine	Geffen, N and Low, M	Geffen conceived, drafted and edited the paper. Low made additions, reviewed and edited the paper.	Geffen and Low (2017)
Efficient and Effective Pair-Matching Algorithms for Agent-Based Models.	Published	Journal of Artificial Societies and Social Simulation	Geffen,N and Scholz, S	Geffen conceived and implemented the algorithms, and programmed the algorithms in C++. Both authors wrote R scripts to analyse the results. Geffen drafted the manuscript. Scholz made additions. Both authors reviewed and edited it.	Geffen and Scholz (2017)
The influence of design decisions on incidence in microsimulations of sexually transmitted infections.	Undergoing peer review	Draft paper included here	Geffen,N and Scholz, S	Geffen conceived and programmed the microsimulation, and the pair-matching algorithms. We both wrote R scripts to analyse the results. Scholz developed the technique for initialising agents into partnerships at the beginning of the simulation. He also analysed and provided the data sets on demographics, and partnership formation and dissolution in the German population. The article was drafted by Geffen. Scholz made additions. Both authors reviewed and edited the manuscript.	N/A

Table 1: List of publications included in this dissertation

Title	Reason not included	Journal	Authors	Contributions	Ref
A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology.	Geffen is 2nd author. However, the challenges encountered with the pair-matching technique used in this article was the inspiration for the work presented in this dissertation.	Sexually Transmitted Diseases	Johnson, LF and Geffen, N	Johnson conceived the article, implemented the microsimulation and drafted the article. Geffen assisted with the pair-matching algorithm.	Johnson and Geffen (2016)
Community perspective on the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial.	Although this article discusses when to start antiretroviral treatment (a theme of this dissertation), it doesn't deal with modelling.	HIV Medicine	Geffen N1, Aagaard P, Corbelli GM, Meulbroek M, Peavy D, Rappoport C, Schwarze S, Collins S; International Network for Strategic Initiatives in Global HIV Trials (INSIGHT) Community Advisory Board.	Geffen drafted the article. Some of the authors made editing suggestions that were incorporated.	Geffen et al. (2015)
Anything to Stay Alive: The Challenges of a Campaign for an Experimental Drug	Although an earlier conception of this dissertation intended to include tuberculosis (the subject of this article), the final dissertation did not.	Developing World Bioethics	Geffen, N	Geffen was the sole author.	Geffen (2015)
The role of activists in access to HIV and tuberculosis treatment and prevention.	Although it deals with themes covered in this dissertation it does not discuss modelling.	Current Opinion in HIV/AIDS	Geffen, N	Geffen was the sole author.	Geffen (2017)
Community views: balancing the public health benefits of earlier antiretroviral treatment with the implications for individual patients - perspectives from the community	Geffen is the 2nd author. Although the article deals with themes covered in this dissertation, it does not deal with modelling.	Current Opinion in HIV/AIDS	Collins, S and Geffen, N	Collins drafted most of the article. Geffen contributed the sections specific to South Africa. Both authors reviewed and edited the draft.	Collins and Geffen (2014)

Table 2: List of publications not included in this dissertation

Software

The software developed in the course of this research is available in two Github respositories: `https://github.com/nathangeffen/pairmatchingalgorithms` and `https://github.com/nathangeffen/faststi`. All software, except for a third-party implementation of the Blossom V algorithm, is licensed under the GNU General Public License version 3.

Chapter 1

Introduction

1.1 Mathematical models and the South African HIV epidemic

Nearly 7 million people were living with HIV in South Africa in 2015. About 150,000 people died of AIDS that year, bringing the total number of people in the country who have died since the onset of the epidemic in the 1980s to 3.5 million (Johnson et al., 2016). Globally in 2015, just under 37 million people lived with the disease, 2.1 million became infected, and 1.1 million died (UNAIDS, 2016).

These statistics are necessarily imprecise; they are estimated by mathematical models that make use of multiple sources of data such as prevalence surveys that count how many people have HIV, incidence studies that try to determine the rate at which people become infected with HIV, studies that estimate life-expectancy with HIV, estimates of background life-expectancy,

and death certificates. These sources of data are imperfectly collected, as is our knowledge of how these factors interact. Nevertheless since the onset of the HIV epidemic both data collection and modelling techniques have improved, enhancing our understanding of this and other sexually transmitted infection (STI) epidemics.

Since AIDS was discovered (Gottlieb et al., 1981), information has been needed to inform life-and-death policy decisions. But because of imperfect or absent data, modelling is used to fill in the gaps. For example, since the late 1980s and early to mid-1990s models have been used to estimate the past, current and future size of the South African epidemic. Until the 2000s there were no countrywide surveys of the number of people infected. So models were constructed using the bits of data available, such as prevalence rates of pregnant women attending antenatal clinics and death registrations. These models, such as Doyle (1993), ASSA500 and ASSA600 (Dorrington, 1998) then estimated past, current and future prevalence, the effect of HIV on life-expectancy, and how the epidemic affected people by gender and age.[1]

These models were important for policy-making. The early models warned of the impending size of the epidemic, indicating that preventative measures needed to be taken. (That such measures were either not taken or failed is a separate complex discussion beyond the scope of this work.) From the 2000s, models such as ASSA2000, ASSA2002 and ASSA2008 were used by researchers to estimate the cost and benefits of treating people with HIV

[1]Most, but not all, of these early models were deterministic compartmental models: The population was divided into compartments based on, for example, risk, sex and age. Then differential equations were used to calculate outputs, for example prevalence, for each compartment.

using antiretroviral (ARV) medicines, e.g. Geffen et al. (2003), Boulle et al. (2003) and Nattrass and Geffen (2005).

Then in the late 2000s, especially following the publication of a highly cited and discussed article by Granich et al. (2009), models were part of the arsenal in a great debate about the optimal time for people with HIV to initiate ARV treatment (ART). Proponents of a test-and-treat strategy (i.e. testing as many people as possible for HIV and immediately recommending treatment to people who tested positive) argued this could reduce the number of new infections because people with HIV on ARVs are less infectious, even uninfectious when the virus in their bloodstream, or plasma, is reduced to undetectable levels. But others, including this author, wanted a more cautious approach, because it was not yet clear when the best time to start treatment for optimal patient health was. Granich et al. (2009) advocated for test-and-treat, because their models estimated that the epidemic could be virtually eliminated using this approach. (By late 2015, clinical trials had shown that both from the perspective of patient health and reducing new infections, immediate treatment is optimal.)

1.2 Modelling methods

Mathematical models of infectious diseases estimate information such as the prevalence, incidence and effect on life-expectancy of the disease at time-points for which there have been no direct measurements. These models usually tell us about the state of an epidemic now and in the future under

different scenarios. [2]

The most common approach to modelling the HIV epidemic has been to use deterministic compartmental methods. These "stratify the population into groups according to each individual's characteristics and HIV infection status and use differential or difference equations to track the rate of movement of individuals between these groups."(Eaton et al., 2012).

For example, using an explanation provided by Johnson (2004), here is a simple deterministic compartmental model of HIV infection:

We have a sexually active population divided into HIV-positive, Y, and HIV-negative, X, adults. People enter the population at a rate π and die in the absence of AIDS at a rate μ. We assume β is the probability of HIV infection in a sexual relationship where the partners are serodiscordant, i.e. one is infected with HIV and the other not, and c is the rate of new partnerships. Once infected, the rate of mortality increases by α.

This is illustrated in Figure 1.1 which is taken from Johnson (2004).

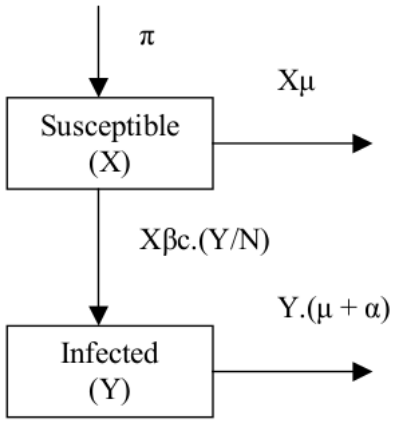

Figure 1.1: Simple model of HIV transmission by Johnson (2004).

[2]We first expressed some of the ideas in this section in Geffen (2013).

Figure 1.2 shows the deterministic compartmental model used in Granich et al. (2009).

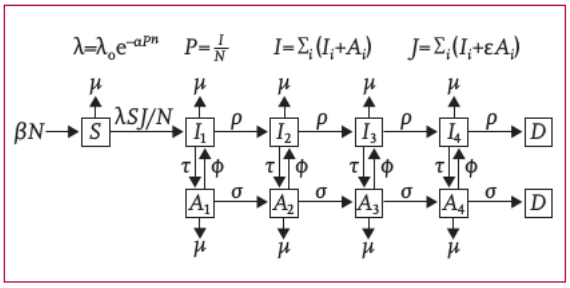

Figure 1.2: Model of HIV transmission by Granich et al. (2009). The authors explain: "N represents population aged 15 years and above. People enter into the susceptible class (S) at a rate βN, become infected at a rate $\lambda SJ/N$, progress through four stages of HIV (I_i, i = 1–4) at a rate between each stage, and then die (D). The background mortality rate is μ and people are tested at a rate τ. If they are tested and put onto ART, they move to the corresponding ART box $A_i (i = 1–4)$, where they progress through four stages at a rate σ and then die. The term governing transmission contains the factor $J\alpha(I_i + \epsilon A_i)$ where ϵ allows for the fact that people receiving ART are less infectious than are those who are not. They might also stop treatment or the treatment might become ineffective, in which case they return to the corresponding non-ART state at a rate ϕ. To allow for heterogeneity in sexual behaviour and for the observed steady state prevalence of HIV, we let the transmission decrease with the prevalence, P. If $n = 1$, the decrease is exponential; if $n = \infty$, the decrease is a step function."

1.2.1 Microsimulation

Orcutt (1957), within the context of economics, described the basis for what are now called microsimulations. "The severe difficulties of testing hypotheses and of estimating relations by use of highly aggregative economic time series are by now fairly widely understood by economic statisticians and are beginning to be more adequately recognized and faced by the economic profession in general," he wrote. "These difficulties and the resulting failure

to achieve satisfactory testing or estimation at a highly aggregative level have been among the elements leading to the large interest now exhibited in formulating and testing hypotheses about the behavior of such elemental decision-making units as *individuals*, households, and firms." (our emphasis)

While deterministic compartmental models estimate aggregate information about the natural world population being studied, microsimulations consist of agents that typically represent individual people and whose behaviour is determined by a set of events and parameters. The parameters can be set per agent, so that each agent has its own unique behaviour. Aggregate information is a function of all the agents and their interactions. Moreover, whether an event is executed on a particular agent, is determined using random (almost always pseudo-random, actually) numbers, for example, whether an agent becomes "infected" with HIV. So microsimulations are stochastic: given the same inputs (but with a different seed for the pseudo-random number generator) these models typically produce different results on different executions.

Many new techniques for modelling infectious diseases have been developed since the early 1990s. The need for more sophisticated models has been driven by the global HIV pandemic. But it is the increase in computing power that has made microsimulations more feasible. These often require substantial computing power to be able to deliver useful results in a reasonable period of time.

In 2012 PLoS Medicine published a *Featured Collection* on mathematical models of the HIV epidemic. The collection represents a useful landmark for how far the technical development of models has progressed since the

earliest models in the late 1980s and early 1990s. The articles in the collection examined the effect of providing antiretroviral treatment to people with HIV before clinically indicated, in order to prevent new HIV infections, one of the most important and debated questions in HIV research at the time (Abbasi et al., 2012).

As an example of the work done consider one of the articles in the collection that compared twelve models, arguably the leading ones in the field. Four were microsimulations and eight were deterministic compartmental models (Eaton et al., 2012). All the models considered the future trajectory of the HIV epidemic under competing scenarios, ranging from no ART to availability of ART to all people with HIV. The authors concluded: "Mathematical models evaluating the impact of ART vary substantially in structure, complexity, and parameter choices, but all suggest that ART, at high levels of access and with high adherence, has the potential to substantially reduce new HIV infections."

It is unlikely that most of these models, especially the microsimulations, would have been feasible with the computing power available in the 1990s, especially not the affordable consumer hardware of that time.

Aside on terminology: microsimulation vs agent-based model

The terms *microsimulation* and *agent-based model* are used frequently in the literature often describing similar or the same techniques. Finding a useful definition that describes the difference between microsimulation

and agent-based models is illusive. For example Heard et al. (2015) is titled *Agent-Based Models and Microsimulation*. Yet except for the title of the article nowhere further in the article is there any mention of the word microsimulation.

The Java Based Agent Simulation Library provides the most comprehensive explanation of the difference between the two terms we have found:

> The main differences between the two approaches can be traced down to the following: (i), microsimulations are more policy-oriented, while agent-based models are more theory oriented; (ii) microsimulations generally rely on a partial equilibrium approach, while agent-based models are most often closed models. The initial population in an agent-based model is typically not meant to reproduce a real population of agents: for example, in a labor market model with firm creation all individuals might be initiated as unemployed, or randomly employed. The focus is on the emergence of aggregate patterns from the interaction of the individual agents, with the aim to replicate some observed stylized fact (business cycle fluctuations, for instance). Accordingly, the value of the parameters that drive the processes are chosen ad-hoc, or only roughly calibrated with real data. However, in their struggle to replace dynamic stochastic general equilibrium (DSGE) models, agent-based models are becoming more empirically

oriented. At the same time, microsimulations are becoming more complex, by including more behavioral responses and general equilibrium feedbacks.

If the two approaches retain different goals and perspectives, from a mathematical and computational perspective they are identical. Both agent-based models and microsimulations are recursive models, where the number and individual states of the agents in the system are evolved by applying a sequence of algorithms to an initial population. (JAS, 2017)

We have not found the differences in terminology useful. We use the terms interchangeably in this work.

1.2.2 Discrete microsimulation

Microsimulations can either model the time until events take place, in which case they are continuous, or they can iterate for discrete time periods, with zero or more events occurring in each time period. Each event changes the state of zero or more agents, and each agent typically represents a person. Our work deals solely with discrete microsimulation. The algorithmic template of the discrete microsimulations we implement is given in Algorithm 1.

The practicality of a simulation is dependent on the efficiency class of the events. A particularly simple event is to age each agent. This merely requires incrementing an age property for each agent by the size of the time-step (typically a day in our work). The efficiency is linear with the number

Algorithm 1 Structure of a discrete microsimulation

```
for each time step do
    for each event e do
        for each agent a do
            if e has to be applied to a then
                Apply e to a
            end if
        end for
    end for
end for
```

of agents.

We usually want the efficiency class of events to be linear or at worst linearithmic [3] with the number of agents. Events with quadratic efficiency slow simulations with large numbers of agents to the point that it often becomes unfeasible to run the simulation many thousands of times in order to generate confidence intervals, conduct sensitivity testing, or estimate parameters. Also slow events can limit the number of agents the microsimulation can accommodate to fewer than we would like to use.

1.2.3 The pair matching problem

Johnson and Geffen (2016) [4] compared a deterministic compartmental model against a microsimulation for six STIs. As far as possible the implementation of the microsimulation matched the deterministic model, except for calculating how sexual partners are matched.

The authors explained that the deterministic compartmental model makes "the simplifying assumption that STI incidence is proportional to STI preva-

[3] Linearithmic means the execution is proportional to $n \log n$ where n is the number of agents.

[4] The second author of this article is the author of this dissertation.

lence in the population," while the microsimulation calculates "STI incidence more realistically by classifying individuals according to their partners' STI status."

The authors found that for all six STIs the deterministic compartmental model estimated higher prevalence than the microsimulation. Fitting the parameters of the two models so as to produce the same prevalence, the deterministic model suggested "more immunity and lower transmission probabilities". It also estimated that eliminating concurrent partnerships and a reduction in commercial sex work had a smaller effect than doing the same in the microsimulation, while the latter model "estimated a smaller impact of a reduction in unprotected sex in spousal relationships."

The two models calculated different prevalences for STIs because of the way the microsimulation matched sexual partners. The authors wrote that the deterministic models "make the simplifying assumption that STI incidence is proportional to STI prevalence in the population". By contrast microsimulations calculate incidence "more realistically by classifying individuals according to their partners' STI status."

The initial implementations of the microsimulation were impractically slow because of the algorithm that matched sexual partners into pairs. We introduced various optimisations to the pair-matching algorithm that reduced the time for a single complete simulation of 20,000 agents, coded in C++, from about 10 minutes to less than half-a-minute on fast consumer hardware. By speeding up the algorithm it became practical to run thousands of simulations. This made it possible to estimate some of the microsimulation's parameters and compare them to those of the deterministic model.

This spurred Geffen's interest in how to improve the vital pair-matching algorithms of microsimulations of STIs. Our aim is to find pair-matching algorithms that are both fast and effective, in that they approximate the distribution of sexual partners (based on various characteristics that may be defined by the modeller, e.g. age, location, sexual orientation, risk behaviour, socioeconomic class) of the natural-world population being studied.

The problem of pair-matching is as follows: Given a set of n agents, each representing a person seeking a sexual partnership [5] find a set of partnerships such that all (or as many as possible) agents are paired with one and only one other agent.

A typical pair-matching event for a microsimulation will, on a particular time step, identify all agents who are looking for sexual partners, place them in a mating pool, and then apply a pair-matching algorithm to match these agents.

This problem can be cast as a graph problem. We can define an undirected complete graph, G, as follows:

1. Define a distance function such that given any two agents, a and b, it calculates a non-negative real number that indicates the compatibility of the two agents for a sexual relationship. Note that $distance(a, b) < distance(a, c)$ implies agent b is a more compatible partner for a than c. Also the function is commutative, so $distance(a, b) = distance(b, a)$.

2. Every agent in the mating pool is a vertex in the graph.

3. The edge between any two vertices is the distance between the agents

[5] We call the set of agents looking for partners a *mating pool*.

represented by the two vertices.

There exists for G numerous sets of distinct pair-matchings. One or more of these is a minimum-distance, or perfect, set in that the sum of all the distances between the pairs is less than or equal to every other pair-matching set. There are algorithms that find the minimum-distance set of pair-matchings, for example Blossom V, which is an advance on an implementation of Edmonds' algorithm by Cook and Rohe (Edmonds, 1965; Cook and Rohe, 1999; Kolmogorov, 2009). But Blossom V is very slow (the worst-case execution time appears to increase with the cube of the number of agents, and the graph has to be constructed first, an operation whose time increases quadratically with the number of agents). Also, it may be a problem for some models that it is not stochastic — it always produces the same set of pairs, unless there are multiple perfect pair-matching sets, in which case the algorithm could be adapted to use a random tie-breaking mechanism. But some models may require more stochasticism than this. Also, reproducing the expected value of a probabilistic distribution with certainty may not be a desirable statistical attribute of some microsimulations.

To address these shortcomings we are interested in pair-matching algorithms that approximate the underlying distribution of partnerships, and do so quickly. This dissertation presents several such pair-matching algorithms, some of them novel. They are tested, compared and analysed. Optimised open-source C++ implementations of the algorithms have been made available, though it is likely that most modellers will wish to slightly adapt the algorithms for their specific needs. Some of the algorithms are very fast and

sufficiently maintain the distribution characteristics of the population being studied. We hope that these algorithms will lead to faster more accurate, easier-to-implement microsimulations of STI epidemics. These algorithms can also be used by any pair-matching modelling event. They could therefore be useful, for example, in modelling animal population dynamics.

1.3 Limitations of modelling

In Isaac Asimov's Foundation, a series of science fiction novels set throughout the galaxy thousands of years into the future, mathematical modelling of human behaviour has advanced to the point where the rise and fall of civilisation can be predicted with astonishing accuracy (Asimov, 1991). Alas, as of 2016 Asimov's vision, while fascinating and entertaining, appears unattainable. We cannot even agree on our predictions of a well-studied epidemic such as HIV. Mathematical modelling is far from an exact science. It is often a controversial endeavour, as this dissertation makes clear. [6]

Assume we wish to estimate the number of people who will be infected with HIV in South Africa five years from now. We need to know the number of people currently infected, the rate at which they die, which on aggregate is determined by age and number of years infected. We need to know the rate at which people become infected, which on aggregate is determined by how frequently people have sex, how many partners they have, and the risk of transmission from a person with HIV to a person without. Further factors are that men and women have different risks of infection. Moreover rates of

[6] We have made the comparison with Asimov's Foundation series previously in Geffen (2013).

transmission are also different among men who have sex with men.

Besides that all of these factors are imprecisely known based on flawed data, there are further complexities. There is heterogeneity across the population. Some people have sex with many different partners within short periods of time, such as sex workers, while on the other end of the spectrum there are people who do not have sex for their entire lives. Some people are more infectious than others. Possibly some people are more predisposed to infection per sexual act. There are also different networks of transmission based on who people mix with which is in turn determined by where they live.

There are no models in existence that take all of these factors into account, and even the most sophisticated models, such as Johnson et al. (2016), have to simplify their assumptions.

Of the various forms of evidence available in medical and epidemiological science — such as explanations of phenomena based on biology, observational studies of groups of people, surveys and randomised controlled trials — population models are often considered one of the weakest. But there are times when models are crucial for informing policy decisions because they provide insights that these other forms of evidence cannot.

For example, consider how many people are currently infected with HIV in South Africa. In a near-perfect world we would test the entire population in a short period of time — not dissimilar to the way a census is carried out — and calculate directly the number of infections. This has never been done and cannot be done, because it is too costly and complex. The closest to it has been a series of household surveys by the HSRC (Simbayi et al., 2014).

But even assuming such a massive study has been done, how do we know what the prevalence rate will be a year or five from now? Questions such as this about the future can only be answered by mathematical models.

Furthermore, consider that this hypothetical survey will test many people in the window period of HIV infection (i.e. people who have recently been infected but even the most sensitive tests do not yet detect this), as well as the tiny but at a population level not insignificant number of false positives, and the fact that a substantial number of people will refuse to be tested. Here modelling has to be done to account for these flaws in our massive survey and improve our estimate. In the real world where such a survey isn't possible, models use the much less comprehensive surveys that have been done to provide estimates broken down by country, province, sex, age and even sexual orientation.

Also consider determining the number of deaths caused by HIV. South Africa's death registration data was greatly improved between the end of apartheid and the onset of the millennium. But, it remains incomplete and imperfect, especially for children (Dorrington et al., 2001; Statistics South Africa, 2002). Furthermore, the underlying cause of death is frequently improperly recorded (Birnbaum et al., 2011). Here models have been vital for filling in the gaps in the data, thereby providing more plausible estimates of the number of deaths across the population by age and sex, and the number of deaths in which HIV is the underlying cause.

Another important question is whether the additional techniques used in newer models offer a better understanding or representation of the problems we are modelling, or if they merely add complexity and, worse, a false sense

of sophistication and accuracy.

For example, the model of Granich et al. (2009), mentioned above, is the most cited one of the HIV epidemic, referenced over 1,600 times according to Google Scholar at the time of writing. The model is distinguished by its simplicity. It does not differentiate between males and females, and includes a very simple, arguably simplistic, mechanism to model heterogeneity of risk of infection. Sexual relationships are not modelled explicitly. By contrast the microsimulations discussed in this thesis model do model sexual relationships explicitly. The models differentiate between sexes, can and sometimes do assign specific sexual behaviour and risk of infection to each individual in the population, and differentiate agents based on sexual orientation. One of our models, very primitively admittedly, makes agents more likely to match if they "live" closer to each other.

It is an open question whether this added complexity adds anything to our understanding to the dynamics of STI epidemics. Our assumptions about differences in risk based on sex, sexual orientation and heterogeneity are based on incomplete and disputed data. We might be exacerbating ignorance by modelling based on these data. On the other hand as data collection of these factors improves, the techniques discussed here might become more useful and accurate.

This debate is touched upon by Brian Williams, one of the authors of Granich et al. (2009), in response to a set of microsimulations that, with stepwise increasing complexity, gave different results to the Granich et al. model (Hontelez et al., 2013; Williams, 2014).

1.4 Aims

This dissertation primarily consists of five articles. [7] The first three articles analyse models of the South African HIV epidemic and their consequent role in policy debates, such as the optimal time for people with HIV to initiate ART. The next two articles offer solutions to a computer science problem: how to efficiently and effectively match agents in sexual partnerships in microsimulations.

The dissertation therefore proceeds from a high-level overview of modelling and its relationship to the response to the HIV epidemic, to a low-level technical contribution to one aspect of modelling, albeit an important one.

This work, therefore, aims to provide a small part of the answer to this broad question: How can we improve the modelling of the HIV epidemic, as well as other STIs, in order to better inform policy debates and decisions?

1.4.1 Context

This collection of articles reflects 17 years of this author's involvement in efforts to reduce HIV-related death, ranging form high-level policy review to technical implementations of models of the HIV epidemic.

Questions about the costs and benefits of providing ARVs to pregnant or breast-feeding women for the prevention of antenatal and postnatal transmission, providing ARVs to people with HIV to arrest the course of their disease, and providing ARVs to sexually active people, with or without HIV,

[7] At the time of writing, one of the articles has been published in a peer-reviewed journal. Two have been accepted for publication in a peer-reviewed journal. One will be published as supplementary online material for one of the peer-reviewed articles. The fifth is undergoing peer-review.

for the purpose of preventing new infections, depend on models for reasonable answers. And the usefulness of those models depends upon technical issues, such as how people are matched in sexual relationships.

Our initial involvement in modelling the HIV epidemic, in the early 2000s, was in the exploration of the costs and benefits of introducing PMTCT and ART into the public health system. Discussions with the developers of the ASSA models, and the consequent ASSA interventions model, which estimated the demographic outcomes of antiretroviral treatment, was followed by our work on costing the provision of antiretroviral treatment (Dorrington, 1998; Dorrington et al., 2001; Dorrington, 2002; Geffen et al., 2003; Nattrass and Geffen, 2005). We also costed the introduction of PMTCT, which albeit a technically relatively simple exercise, depended on the estimates of the ASSA models (Nattrass, 2001; Geffen, 2001).

Later Leigh Johnson, one of the main developers of the ASSA models, began exploring the additional insights into the HIV epidemic that he believed could be gained from microsimulations of the HIV and other STI epidemics. Here we encountered a problem. The most crucial difference between the outputs of a deterministic model and a microsimulation comes from the way sexual partnerships are modelled. Johnson had developed an algorithm for capturing the age distribution of heterosexual relationships based on the — admittedly limited and flawed — data available. But the algorithm to match pairs of individuals was so slow that it rendered the microsimulation largely impractical for his purposes. Together we worked on optimising the pair-matching algorithm, ultimately implementing a very complex domain-specific one which was fast enough for Johnson's purposes (Johnson and

Geffen, 2016).

This led to this author's interest in exploring pair-matching algorithms which were easily usable by other modellers.

Three features are desired of a pair-matching algorithm. It needs to be

1. simple or generic enough that other modellers could either adapt it or use generic implementations;

2. fast, so that simulations can be executed hundreds or even thousands of times on consumer hardware in a reasonable time period; and

3. effective in that it should closely approximate the distribution being used to model sexual partnerships in the population.

This research presents several algorithms that appear to meet these characteristics, at least for some types of microsimulations.

1.5 Structure of this dissertation

This is a thesis by publication. Table 1 lists the publications included in it.

Chapters 2 to 5 examine modelling with respect to the policy issues affecting the South African HIV epidemic.

Chapter 2 describes theoretical aspects of modelling and also various models used to analyse aspects of the South African HIV epidemic. It primarily consists of an article accepted for publication in the Southern African Journal of HIV Medicine. Because of the word count limitation of the article, it also contains an afterword that describes models omitted from the article.

Chapter 3 describes HIV policy controversies and the role of models in them. It primarily consists of an article that will be included as supplementary online material to the article in chapter two.

Chapter 4 focuses on one salient policy debate: the question of when to start treatment. It primarily consists of an article accepted for publication in the Southern African Journal of HIV Medicine. Here models were just one of the pieces of evidence used by protagonists in the debate to advocate various positions. Although discussion of models forms only part of the article included in this chapter, it helps clarify how models fit into the wider consideration of a complex policy issue.

The next section of the dissertation deals with pair-matching algorithms, a particular technical aspect of microsimulations, the type of model this author has worked most closely with.

Chapter 5 presents several pair-matching algorithms, and analyses their efficiency and effectiveness using two very simple microsimulations, defined specifically for this purpose. It primarily consists of an article published in the Journal of Artificial Societies and Social Simulation.

Chapter 6 presents results of a more complex microsimulation on a fictitious sexually transmitted infection using three of the algorithms. The data is from the German population because we collaborated with a German researcher, Stefan Scholz, to work on the pair-matching problem. However the insights gained from this work are as relevant to microsimulations of the South African HIV epidemic and other STIs. The chapter primarily consists of an article submitted for publication to the Journal of Infectious Disease Modelling. A disquieting result of this article is that, under certain condi-

tions, the incidence estimated by microsimulations that use a sophisticated pair-matching algorithm decreases as the number of agents in the model increases. This raises a question concerning what knowledge we actually derive from models, and we discuss this briefly in an afterword.

The concluding chapter surveys what the preceding chapters have shown and considers what further research arises from it. [8]

[8]Besides the publications presented in this article, an article directly relevant to this work is Johnson and Geffen (2016) (discussed above). The article has been omitted from this dissertation because Geffen is the second author.

Chapter 2

Modelling the South African HIV epidemic

2.1 Foreword

The following article *Modelling the HIV epidemic: A review of the substance and role of models in South Africa* describes the history of HIV modelling in South Africa, starting with the first substantive model, published in 1990. It then discusses most of the major models through to 2016. It constitutes a literature review of HIV models. Both deterministic compartmental and microsimulation models are described. It is the latter that are of particular interest in this thesis, and for which the pair-matching algorithms of Chapter 5 have been developed.

Because of space limitations, a few important models are not mentioned in the article. Brief notes on these are provided in the afterword. It also contains further discussion on models and the pair-matching problem.

An important point we make in this article is that although understanding the inner workings of models requires technical expertise, the assumptions, inputs and outputs of a model should be well articulated by the model authors, and therefore understandable to researchers, activists and policy makers interested in HIV. Models are well-described if non-experts can read the descriptions of two or more models and understand how they differ and why their outputs are different.

Southern African Journal of HIV Medicine
ISSN: (Online) 2078-6751, (Print) 1608-9693

Modelling the human immunodeficiency virus (HIV) epidemic: A review of the substance and role of models in South Africa

Authors:
Nathan Geffen[1]
Alex Welte[2]

Affiliations:
[1]Department of Computer Science, Centre for Social Science Research, University of Cape Town, South Africa

[2]South African Centre for Epidemiological Modelling and Analysis (SACEMA), University of Stellenbosch, South Africa

Corresponding author:
Nathan Geffen,
nathangeffen@gmail.com

Dates:
Received: 29 Mar. 2017
Accepted: 07 Aug. 2017
Published: 21 Feb. 2018

How to cite this article:
Geffen N, Welte A. Modelling the human immunodeficiency virus (HIV) epidemic: A review of the substance and role of models in South Africa. S Afr J HIV Med. 2018;19(1), a756. http://doi.org/10.4102/sajhivmed.v19i1.756

We review key mathematical models of the South African human immunodeficiency virus (HIV) epidemic from the early 1990s onwards. In our descriptions, we sometimes differentiate between the concepts of a model world and its mathematical or computational implementation. The model world is the conceptual realm in which we explicitly declare the rules – usually some simplification of 'real world' processes as we understand them. Computing details of informative scenarios in these model worlds is a task requiring specialist knowledge, but all other aspects of the modelling process, from describing the model world to identifying the scenarios and interpreting model outputs, should be understandable to anyone with an interest in the epidemic.

Introduction

No epidemic has received the attention of the ongoing human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS) pandemic, and no matter of public health concern has been the subject of so much controversy and policy debate. Scenario modelling has been widely employed in attempts to better understand the demographic, health and economic impacts of the epidemic under various interventions, for example, antiretroviral treatment, pre-exposure prophylaxis and condom use.

Despite modelling being ubiquitous and some models generating intense public debate, with consequences, for example, on World Health Organization (WHO) treatment and prevention guidelines, it remains poorly understood by non-specialists. Even modellers themselves hold differing views about the principal uses and limitations of models.

This article reviews the evolution of models, and their applications, in the context of the South African HIV epidemic. We describe, in terms aimed at a wider audience than just modellers, the basic structure of the modelling process, challenges that modellers face and how this has affected policy debate. In Appendix 1, we explore in more detail the social complexities of the particular issues and controversies.

Basic modelling concepts

Books, tutorials and reviews of epidemiological modelling are plentiful, including guidance for working with models in the context of policy debate.[1,2,3,4] Nevertheless, it is useful to review some essential aspects of all scenario modelling. The aim of mathematical modelling is to first identify the key rules that govern the behaviour of a natural world phenomenon, and then to implement those rules in mathematical relations, so that we can learn more about the phenomenon.

For models of the HIV epidemic, this may mean understanding how gender, age, location and other sociological factors influence fertility, exposure to 'infectious contacts', access to healthcare and mortality. What determines whether and what kind of model is feasible or useful are the questions we want to answer about the 'real world' epidemic, coupled with the data available to justify assumptions about precisely stated rules driving critical processes (dynamical rules). Mathematical and computational challenges may be substantial and sometimes curtail the ambitions of modellers.

It is important to differentiate the specialised technical aspects of model construction and analysis, carried out by mathematical modellers, from the conceptual aspects, which are accessible to anyone with basic insights into the situation being modelled, including doctors, politicians, health

system administrators, biologists and activists. This conceptual and technical distinction helps clarify thinking and reminds us that model building should be an inclusive multidisciplinary process rather than the protected domain of specialists.

For example, in the early 2000s, members of the activist organisation the Treatment Action Campaign approached the developers of the Actuarial Society of South Africa (ASSA) models and asked them to incorporate antiretroviral treatment into their model, which they did. An analysis of the cost of rolling out antiretroviral treatment in the public health system, based on the outputs of the ASSA model, was featured on the front page of the *Mail & Guardian*. The ASSA models were explained by demographers in affidavits in litigation by activists advocating for treatment. The scenarios, assumptions and outputs of the model were debated and understood by a broad range of people: politicians, activists, lawyers, etc. The actual equations in the model spreadsheet were likely of interest to, and understood by, only a handful of specialists.[5,6,7,8]

To maintain the distinction between concepts and techniques, we use concepts popularised by ecological modeller Tony Starfield: *model world* versus *model implementation*.[9] A Model World is the conceptual realm in which we explicitly declare the rules – usually some simplification of 'real world' processes as we understand them. Model Implementation then refers to the mathematical and computational details.

For example, a model world may be conceptually inhabited by genderless people between the ages of 15 and 49 who all have exactly the same behaviours and mortality. We may declare that in our model world each day brings the same risk of infection or death as the day before, without any notion of individual age, the mechanisms of infection or death. A related *model* implementation of such a world may consist of some mathematical equations or computer programme.

Model worlds capture the essential ideas which we then formally analyse and explore in technical investigations, using mathematical and computational tools. A model world has abstracted entities and rules, but no particular history. When we set up initial conditions in a model world, like winding up a clock set to midnight, and then let it run, we produce scenarios – particular realisations of processes and events consistent with the assumptions of the model world. A full-fledged investigation may involve many scenarios located in several model worlds.

There are typically two kinds of variables in a model: (1) *state variables*, that is, scenario-specific accounting indicators, such as the size of population, number of infections and number of deaths, and (2) *parameters*, that is, model world defining metrics such as, most critically, rates of infection, rates of death and other state transition rules. When the model executes (e.g. as a stand-alone piece of software or as a spreadsheet), state variables evolve over time from given initial conditions, but for this to happen, parameter values must actively be chosen. Sometimes parameters are chosen based on pre-existing knowledge or estimates. Sometimes they are chosen entirely heuristically, just to see what is implied by their values lying here or there within some plausible range. Another option is *model calibration*, by which parameters are chosen in such a way that the emergent behaviour of the model is consistent with some data. For example, we can try different values of an infectious 'contact rate' (how frequently people become infected), and then see whether a suitably narrow range of this parameter produces a time-varying prevalence that is consistent with survey data.

For sexually transmitted infections, a key aspect of model worlds is how infections occur. Infection can happen for a population group at some rate, without any concern for sexual interactions. There can be a single rate across the population or it could be differentiated by age, risk group and gender. Alternately, infection can be conceptualised at a very fine level of detail: a model world could track sexual relationships – or even sexual acts – per individual, with each individual having their own risk of contracting or transmitting the infection.

In model worlds, there are no grey areas of the kind we find in the real world, no hidden unknown rules, factors and entities – although the interplay of components may be complex and may require some sophistication to implement, or conceptually untangle. Modelling then might be seen as teasing out the implications of hypothetical claims about how the world is composed and governed. If done skilfully, this helps explain some aspect of the real world. It informs real world choices that need to be made, even if the full underlying truth in the real world is much more elusive and ambiguous than in any model world we may have constructed. Table 1 highlights key features of the model worlds implemented in the models we review here.

The crucial point is as follows: everyone with a legitimate interest in the situation being modelled is entitled to a comprehensible description of the model world. They should expect to be part of the model world construction, critique and interpretation processes. Modellers need to talk in conceptual terms about this model world, without resorting to jargon or specialised techniques.

The core demographic models

Padayachee and Schall, working for Johannesburg's City Health Department, published the first serious model of the whole South African HIV epidemic in April 1990.[10] They cited two earlier models that estimated the number of gay men and antenatal care attendees in what was then southern Transvaal with HIV. They also mentioned a WHO model that estimated the number of AIDS cases in South Africa but noted that the model was based on 'very little, if any, supporting evidence from South Africa' (p. 330).

TABLE 1: Examples of models of the South African HIV epidemic.

Model	Model world				Scenarios	Implementation
	Population	Transmission	Mortality	Interventions		
Padayachee and Schall 1990	Black people aged 15 to 49	HIV incidence and prevalence estimated from blood transfusion, antenatal and clinic infection numbers.	Not applicable	None	Used data sources to estimate number of black people aged 15 to 49 with HIV from 1989 to 1991	Three simple models using straightforward calculations
Doyle 1990	Population divided by sex, 5-year age intervals and four HIV risk groups.	Mainly a function of risk group and the proportion of infected people, but 'some allowance' for 'sexual activity according to age and sex'.	Age-related non-HIV mortality. Additional risk of mortality for people with HIV	None	Many. Doyle used it to estimate South African population, while Lee et al. used it to estimate infections in Soweto. The initial HIV-positive population is 'imported' into the model.	Macro
Padayachee 1992	Individuals have age and sex.	Each person, adjusted for age and sex, has a probable number of sexual partners with whom they have sex a probable number times, each of whom has HIV with a specified probability.	Mortality not explicitly discussed, but number of AIDS cases calculated based on infection period	None	From 1985 a prespecified number of immigrants with HIV 'seed' the model. Number of HIV and AIDS cases estimated until 2000	Micro
ASSA (various)	Population divided by sex, province, 5-year age intervals and four HIV risk groups. Infants enter the population annually. People with HIV at various clinical stages of progression.	Function of risk group, proportion of infected people, age and sex. Mother-to-child transmission also modelled.	Age and sex-related non-HIV mortality. Additional risk of mortality for people with HIV	From ASSA2002, antiretrovirals, mother-to-child transmission prevention, condoms, etc.	Calibrated to available data sources up to the year of the model suffix, and then projected forward.	Macro (originally as spreadsheets, then as C++ code)
Granich deterministic 2009	People of no sex or specific age, except that they are 15 to 49 years. People with HIV are assigned to a WHO stage.	Homogenous: no risk groups, single incidence rate for the whole population.	Single mortality rate for people without HIV. Additional risk of mortality for people with HIV	Scaled-up universal test-and-treat versus treating at CD4 count of 350 versus no treatment	Calibrated to South African adult HIV epidemic.	Macro (the authors also did a stochastic model)
Hontelez 2015	In the most complex model of their nine models, people are differentiated by age and sex.	Heterogeneous sexual behaviour. People are part of sexual networks and people at different stages of HIV infection have different degrees of infectiousness.	Age and sex-related non-HIV mortality. Additional risk of mortality for people with HIV	Similar to Granich et al.	Calibrated to South African adult HIV epidemic.	Micro
THEMBISA 2014–2016	Population divided by sex, province, 5-year age intervals and HIV risk groups. Infants enter the population annually. People with HIV at various CD4 count based stages of progression.	Function of risk group, proportion of infected people, age and sex. Mother-to-child transmission also modelled.	Age and sex-related non-HIV mortality. Additional risk of mortality for people with HIV	Antiretrovirals, PMTCT Option B+, condoms, etc.	Calibrated similarly to ASSA but also includes additional data on marriages and partnerships.	Macro

Padayachee and Schall actually implemented three simple models, which used antenatal clinic, blood transfusion, sexually transmitted infection and family planning clinic data and population estimates by province to estimate infections for the whole country up to 1992. Their model worlds consist of adult (aged 15 to 49 years) black people, possibly living in a particular province or urban or rural area, but with no other identifiable characteristics. Their first model fitted clinic and blood transfusion data to estimate a rate at which the epidemic was growing. They extrapolated this to calculate national prevalence and the rate at which it was growing up to 1992. Their second model, whose method they called 'direct', used various data sets to estimate the number of people with HIV in each province, which they then aggregated for the whole country. Their third model, whose method they called back calculation, used the number of known AIDS cases and an assumption about the time from HIV infection to AIDS to back-calculate the number of HIV cases, derive an incidence rate and then use this to project the number of HIV cases in the future.

They estimated the number of black South Africans, aged 15 to 49 years, with HIV for the end of 1989, 1990 and 1991. Their model calculated between 45 000 and 63 000 infections by end of 1989, rising to between 317 000 and 446 000 at the end of 1991. Clearly there were problems with their methodology: for one thing blood donor HIV prevalence rates were not representative. However, their models provided some idea of the extent of the epidemic using the limited data available then. They wrote: 'Because of the lack of basic data, these forecasts are tentative, but they nevertheless indicate the great seriousness of the HIV epidemic in South Africa' (p. 329).

In October 1990, Doyle and Millar, working for the Metropolitan Life Insurance Company, published one of the most influential models of the epidemic.[11] They constructed a model world with an adult population comprising four risk groups: (1) people having no sexual contact, or in long-term monogamous relationships, who are not at risk of HIV, (2) people at some risk, conceived as being in stable relationships but with one or the other partner having more than one sexual relationship, (3) people with higher levels of risk, such as those with other sexually transmitted infections, and (4) sex workers and people with large numbers of sexual partners. These four risk groups remained a part of highly

cited models derived from or based on the Doyle model (as it came to be known) until the late 2000s.

Model world inhabitants were assigned rates for forming new relationships, within and across risk groups, and rates of transmission within relationships. It allowed for 5-year age groups to be defined, with different levels of HIV prevalence at the beginning of a scenario. It also had parameters for fertility, mother-to-child transmission rate and HIV and non-HIV mortality rates. It could be used for heterosexual or homosexual populations, adapted as needed to populations of interest. Doyle applied the model to South Africa, leading to prescient predictions that were not obvious in the early 1990s, for example, that the epidemic would kill many young adults, but that the population would not decline (although the growth rate would slow).[12] Lee et al. applied the model to Soweto, estimating that by 2010 it would account for 28% – 52% of all deaths there.[13]

The model implementation of Doyle and Millar's model world was in the form of population counts at discrete time steps, deterministically updated according to the expected values emerging from statistical rules (like probability of infection or death). Doyle cites other models developed by the Institute of Actuaries and Society of Actuaries at the time but points out that these 'considered one small homogeneous risk group' and were inappropriate for modelling the South African epidemic.[12]

These model implementations are often called *deterministic compartmental, frequency-dependent* or *macro*. By contrast, a *microsimulation, network* or *agent-based* implementation, possibly of the same underlying model world, proceeds by explicitly tracking a large number of identifiable individual model world inhabitants and subjecting them, usually stochastically, to the different events to which they are exposed. The first microsimulation of the South African epidemic that we can find is a Medical Research Council lecture cited in Doyle and Millar's 1990 paper. Unfortunately, we can find no further references to this particular one in the literature.

Doyle's model was a proprietary one used by Metropolitan Life, primarily for the purpose of making decisions about employee benefits (pers comm. Stephen Kramer). It was the progenitor of other models, including those of the ASSA. Given the computing power at the time, the level of detail is impressive, in most respects exceeding the complexity of a widely cited and highly impactful model, published as late as 2009[14] that stimulated the debate on early treatment as a means of reducing new transmissions.

Groeneveld and Padayachee used a microsimulation implementation of a model world in which each person, based on their age and gender, has an expected number of sexual partners per year, with a specified proportion of 'short' relationships, and an estimate of the frequency of sexual contacts with partners who are infected with HIV with a probability dependent on age and gender. The authors also estimated the annual number of immigrants with HIV who entered South Africa annually. Their goal was to 'to estimate the extent of HIV infection among black heterosexual South Africans'. They attempted to predict new HIV infections for the period 1985–2000 and concluded that there would be 5.7 million people with HIV in South Africa by 2000. By comparison, the most comprehensive up-to-date current model of the epidemic, THEMBISA, estimates that there were 3.3 million people infected in 2000.[15]

Brophy adapted a World Bank model for the South African epidemic and investigated demographic effects on the black population under various scenarios.[16] The model world divided the population by sex and 5-year age groups. There were also partially overlapping groups: blood transfusion recipients, heterosexual females, heterosexual males and bisexual males. It considered fertility rates, the age pattern of fertility and mortality levels by male and female. The model population was matched to the sex and age structure of the 1985 census. Various data sources were used to estimate fertility and life expectancy. Some parameters, such as the number of sexual partners, coital frequency, condom use, as well as fertility and life expectancy from 2005 to 2010, were essentially guessed (and various scenarios were tried). They calibrated the model so that it estimated the middle estimate of the number of infections in 1990 of the model by Padayachee and Schall (described above). Brophy predicted substantial reductions in the population and life expectancy in 2000, 2005 and 2010 under three AIDS scenarios of increasing severity versus a no-AIDS scenario. In the bleakest scenario, the model estimated that there would be just about 1.9 million people with HIV in the adult black population in 2000 (the actual number was about 3 million).

Dorrington described the origins of the ASSA models.[17] The first ASSA model was developed on a spreadsheet by a team led by Alan Whitelock-Jones. It was titled ASSA500 and was similar to the Doyle model with some simplifications. Dorrington explains that the motivation for ASSA to develop a model when the Doyle model already existed was that the latter was proprietary and there was a need for a 'program which the user could alter to his or her needs' (p. 99). Consequently, the model was placed on ASSA's website and Dorrington wrote: 'the reader is encouraged to download and play with it' (p. 101).

Dorrington also wished to improve the model world of Doyle, specifically because it assumed constant fertility and non-HIV mortality over time; the ASSA model world would include decreasing fertility rates and improving non-HIV mortality. Using the same risk groups as the Doyle model, it additionally accounted for 'net national in-migration' (p. 100). While users of the Doyle model needed to set the parameters for the community they were modelling, the ASSA models are explicitly aimed at modelling the South African epidemic, with later models disaggregating the outputs by province.

The starting point of the ASSA600 model was the 1985 South African population, known from a census conducted that

year. The model was calibrated to reported AIDS cases in 1995 and antenatal HIV prevalence, derived from annual surveys by the Department of Health for 1994–1997. Dorrington described the calibration of the model as 'perhaps inevitably a little more art than science'. The ASSA modellers aimed to produce a population estimate for 1996, national mortality rates for 1998, a projection of antenatal clinic HIV prevalence rates and a projection of national fertility rates.

Over the next decade, ASSA600 had several successors: ASSA2000, ASSA2002 and ASSA2008. From 2000, the suffix indicates the latest year of the empirical, primarily antenatal survey, data against which the models were calibrated (i.e. not the year they were published). The goal was to fit the known empirical data, estimate past unknown and project future, demographic and HIV outputs, such as population size, non-HIV and HIV mortality, HIV prevalence and incidence. From ASSA2002, the effects of antiretroviral treatment were incorporated.[5,6]

The ASSA models are widely cited. Besides being comprehensive, they have also been open and easily accessible. As Johnson explains, the 'Excel interface of the publicly-available model is appealing to many non-modellers'.[18]

The latest ASSA model has been calibrated with data only as far as 2008. In recent years, the ASSA models too have been superseded, most notably by the THEMBISA model.[18] This combines the features of three other models, besides the ASSA model. The model world complexity is substantial, including more realistic sexual behaviour 'calibrated to marriage data and cross-sectional data on numbers of partners', 'more determinants of mother-to-child transmission' and 'most of the new strategies for preventing and treating paediatric HIV'. For example, it features CD4 count staging instead of clinical staging as in the ASSA models and allows for earlier antiretroviral initiation. It includes newer prevention interventions such as male medical circumcision, pre-exposure prophylaxis and 'WHO options B and B+ for prevention of mother-to-child transmission'. In contrast to the ASSA models, it takes into account change in risk behaviour by people over time.

The Joint United Nations Programme on HIV and AIDS (UNAIDS) has also produced widely used models. In the 1990s, UNAIDS used Epimodel – developed in 1987 by the Global Programme on AIDS – for its global, regional and country HIV projections.[19] This was eventually replaced by Spectrum, developed by the erstwhile Futures Group (now Avenir Health), and the Estimation and Projection Package (EPP), since combined into one programme.[20,21,22]

The model provides a user interface that takes a range of inputs, for example, base year population by age and sex, fertility rates, life expectancy (AIDS and non-AIDS), migration rates, number of people on antiretrovirals, number of people on cotrimoxazole and about a dozen or so more (see Table 1).[22] It then aggregates all cases in the population aged 15 to 49 years and fits a non-age-structured population model to the historical aggregates, thereby inferring incidence and projecting outputs such as HIV infections and deaths.

Johnson[18] writes:

> The Spectrum/EPP model is used … in producing estimates of the global distribution of HIV, and therefore has the advantage of benefiting from a substantial body of international expertise in HIV epidemiology. However, the separation of the modelling of HIV incidence and demographic impact in this model does limit the ability of the model to make use of age-specific data in model calibration. [p. 6]

Spectrum/EPP is used to estimate official estimates for every country in the world every two years for the United Nations Population Division; it serves an important purpose, providing rough estimates of HIV prevalence and mortality where none would otherwise be available. The model is also used to analyse the long-term impact and cost of interventions, though as Johnson says, it is 'limited in its ability to evaluate the impact of HIV prevention strategies and make long-term projections'. Where countries have developed high-quality specialised models, such as the THEMBISA model for South Africa, it makes more sense to use these.

Modelling when to start treatment

In 2009, Granich et al. at the WHO presented two models.[14] The first model is a population-level transmission model (implemented deterministically) that calculated the long-term dynamics of the HIV epidemic based on different treatment strategies. The second model (implemented stochastically) investigated the effect on R0 – 'the number of secondary infections resulting from one primary infection in an otherwise susceptible population' – of different treatment strategies applied to an hypothetical person.

The paper argued that, in South Africa, a policy of universal testing coupled with immediate treatment for adults found to be HIV-positive would effectively eliminate the epidemic. In particular, they estimated that HIV incidence could drop to less than 0.1% per year by 2016. They also costed the strategy.

The paper caused great excitement and controversy. It has been cited, according to Google Scholar, 1640 times (as of 11 March 2017). We know of no other HIV model that has been cited as often, which is extraordinary considering the simplicity of the models: there is no gender or age structure. Perhaps this simplicity, coupled with the strongly stated message the authors conveyed, engaged readers across multiple disciplines and accounted for much of the interest taken in the paper. The paper also encouraged a flurry of other models that looked at the same question.[23]

Even 4 years later, a detailed set of microsimulation models by Hontelez et al. was published, trying to answer the same question as Granich et al.[14] The modellers developed 'nine structurally different mathematical models of the

South African HIV epidemic in a stepwise approach of increasing complexity and realism'.

The simplest resembled the Granich model. The most complex included 'sexual networks and HIV stages with different degrees of infectiousness'. Hontelez et al.[24] defined 'universal test-and-treat' as annual screening and immediate treatment for all HIV-positive adults, starting at 13% in January 2012 and scaling up to 90% coverage by January 2019. Elimination of the HIV epidemic was defined as incidence below 1 per 1000 person-years.

It is controversial whether addition of complexity to models improves them. For example, one of the authors of the Granich et al. paper, Brian Williams, has written:

> Hontelez et al. suggest that the [then] current scale-up of ART at CD4 cell counts less than 350 [cells/mm^3] will lead to elimination of HIV in 30 years. I disagree … and believe that their more complex models rely on unwarranted and unsubstantiated assumptions.[25]

The Granich model and the ensuing attempts by other modellers to verify, refute or improve upon it raise important questions about what we are trying to achieve with modelling. The original paper is the one that was widely debated. Even though it could be improved, it answered the question of whether a test-and-treat policy had the potential to massively reduce incidence. Most subsequent models agreed with that of Granich et al. that universal test-and-treat would substantially reduce new infections but not as quickly as they proposed. The assumption of rapid scale-up of treatment coverage and significant viral suppression in those failing treatment were, perhaps, too optimistic.

Models targeting particular policy conundrums

Interventions other than antiretroviral treatments have also been modelled. There are numerous such models, and here we briefly note some without describing their model worlds.

The results of a randomised controlled trial that compared infection rates in circumcised versus uncircumcised men in Orange Farm[26] were used to calculate that this intervention could prevent between 1.1 and 3.8 million infections as well as 0.1 to 0.5 million deaths over a 10-year period in sub-Saharan Africa.[27] A comparison of the cost-effectiveness of treatment as prevention, treatment (solely for the benefit of the patient) and circumcision concluded that although treatment as prevention was cost-effective, it was less so than treatment or circumcision.[28,29]

Modelling the introduction of pre-exposure prophylaxis (PrEP), researchers found that it could avert 30% of new infections in 'targeted age groups of women at highest risk of infection'. However, they also found that the cost-effectiveness of PrEP relative to treatment would decrease rapidly as treatment coverage increased.[30] Another group had more optimistic results modelling PrEP in serodiscordant couples (although it is unclear how a model can address whether antiretrovirals should be given to the HIV-negative or HIV-positive partner in a relationship).[31] They concluded:

> Although the cost of PrEP is high, the cost per infection averted is significantly offset by future savings in lifelong treatment, especially among couples with multiple partners, low condom use, and a high risk of transmission. [p. 1]

Another model found that treatment plus PrEP was more effective than either strategy alone but would also produce high prevalence of drug resistance.[32] Hallett et al. investigated the use of PrEP for seronegative partners in stable serodiscordant partnerships, as an alternative or adjunct to treatment for the HIV-positive partner.[31]

Sexual behaviour – such as condom use, number of partners, concurrency, and transactional sex – has been widely modelled.[33,34,35,36,37,38] Models developed by the ASSA researchers, for example, estimated that HIV incidence in South Africa dropped during the period from 2000 to 2008 and that increased condom use was the 'most significant factor explaining' this decline.[39] The role of concurrency has however been contentious, with conflicting findings.[33,34,40]

Experimental interventions such as microbicides[41] and vaccines have also been considered,[42] and so has the role of treating sexually transmitted infections.[43] For further references, see Johnson.[18]

Currently, models such as THEMBISA and Spectrum are being used to track progress towards national and global objectives, such as the UNAIDS 90-90-90 targets (90% of people with HIV diagnosed, 90% of people diagnosed on treatment and 90% of people on treatment virally undetectable),[44] as well as elimination of mother-to-child transmission.[45]

Discussion

The distinction between model worlds and the technical implementation of models is useful for demystifying modelling and perhaps allows more people to participate in model construction and critique, and hence reach better informed decisions on the policy implications of models. While models, with their complex equations and computer code, might be impenetrable to all but specialists, the conceptual ingredients – the model world – should be accessible to a wide audience.

The earliest models of the South African HIV epidemic projected prevalence and mortality over time, a task that remains useful today. New models were subsequently developed to estimate the effects of interventions, for example, how antiretroviral treatment would reduce mortality (ASSA2002 interventions model) or how it would reduce new infections (the Granich model).

The challenge facing modellers was summarised by Dorrington[5]:

> Estimating the exact impact of HIV/AIDS on mortality is not a simple task since there are many uncertainties surrounding the dynamics of the spread of the virus and subsequent passage to death. In addition there are difficulties in deciding on the level of overall mortality in South Africa since not all deaths are registered. However, determining an *order of magnitude of the impact* is well within the capabilities of a trained demographer. (our emphasis) (para. 7)

Models, even simple ones, can shed light on 'big picture' questions. They cannot be used to provide precise predictions of the long-term future. Models can also provide plausible estimates of unobserved epidemic indicators and assist with planning for the short-term future. These benefits and limitations of models should be kept in mind before deciding to add complexity to model worlds, and consequently model implementations.

Modelling is still an evolving component of biomedical science. Perhaps, as we argue in Appendix 1, a key factor in advancing consensus in how models are assessed, especially with societal implications, is a more inclusive interdisciplinary approach to defining and debating 'model worlds', and 'model world scenarios', the conceptual aspects of modelling that should be accessible to everyone with an interest in the HIV epidemic. This should lead to improved models that contribute more robustly to policy discussions.

Acknowledgements

Competing interests

Nathan Geffen occupied senior positions, both as a staff and executive member, in the Treatment Action Campaign (TAC) from 2000 to 2013. Both he and the TAC were active participants in the debates described in this article. He has received bursary contributions towards his PhD research from the Centre for Social Science Research and the South African Centre for Epidemiological Modelling and Analysis. Alex Welte has no disclosures.

Authors' contributions

N.G. conceived and drafted the article. A.W. proposed including the terminology on model world, model and scenario, and drafted this section of the article. The authors reviewed and edited the article together.

References

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 Jul 10;9(7):e1001239. https://doi.org/10.1371/journal.pmed.1001239
2. Vynnycky E, White R. An introduction to infectious disease modelling. 1st ed. New York: Oxford University Press; 2010. 400 p.
3. Garnett G. An introduction to mathematical models in sexually transmitted disease epidemiology. Sex Transm Infect. 2002 Feb;78(1):7–12. https://doi.org/10.1136/sti.78.1.7
4. Johnson L. Notes: An introduction to the mathematics of HIV/AIDS modelling. [publisher unknown]; 2004.
5. Dorrington RE. Expert affidavit in TAC Competition Commission complaint [homepage on the Internet]. 2002 [cited 2016 May 18]. Available from: http://www.tac.org.za/Documents/DrugCompaniesCC/TAC_Dorrington_Final.doc
6. Johnson LF, Dorrington R. Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. Demogr Res. 2006 Jun;14(22):541–574. https://doi.org/10.4054/DemRes.2006.14.22
7. Nattrass N, Geffen N. The impact of reduced drug prices on the cost-effectiveness of HAART in South Africa. Afr J AIDS Res. 2005 May;4(1):65–67. https://doi.org/10.2989/16085900509490343
8. Deane N. The price of three million lives. Mail Guard. 2016 Mar;2.
9. Starfield T. Tony starfield shares his thoughts on modeling [homepage on the Internet]. 2012 [cited 2017 Aug 17]. Available from: http://www.uvm.edu/~tdonovan/modeling/NiceToKnow/00_ModelThoughts_transcript.pdf
10. Padayachee GN, Schall R. Short-term predictions of the prevalence of human immunodeficiency virus infection among the black population in South Africa. S Afr Med J Suid-Afr Tydskr Vir Geneeskd. 1990 Apr;77(7):329–333.
11. Doyle P, Millar D. A general description of an actuarial model applicable to the HIV epidemic in South Africa. Trans Actuar Soc South Afr. 1990;VIII(II):561–593.
12. Doyle P. The demographic impact of AIDS on the South African population. In: Cross S, Whiteside A, editors. Facing up to AIDS [homepage on the Internet]. Palgrave Macmillan UK; 1993 [cited 2016 Feb 19]. p. 87–112. Available from: http://link.springer.com/chapter/10.1007/978-1-349-22597-2_5
13. Lee T, Esterhuyse T, Steinberg M, Schneider H. Demographic modelling of the HIV/AIDS epidemic on the Soweto population–results and health policy implications. S Afr Med J Suid-Afr Tydskr Vir Geneeskd. 1996 Jan;86(1):60–63.
14. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. Lancet. 2009 Jan;373(9657):48–57. https://doi.org/10.1016/S0140-6736(08)61697-9
15. Groeneveld H, Padayachee N. A stochastic model for medium-term estimation of the prevalence of HIV infection in a South African heterosexual population. S Afr Med J Suid-Afr Tydskr Vir Geneeskd. 1992 Jan;81(2):67–70.
16. Brophy G. Modelling the demographic impact of AIDS: Potential effects on the black population in South Africa. In: Cross S, Whiteside A, editors. Facing up to AIDS [homepage on the Internet]. Palgrave Macmillan UK; 1993 [cited 2016 Feb 19]. p. 113–134. Available from: http://link.springer.com/chapter/10.1007/978-1-349-22597-2_5
17. Dorrington RE. ASSA600: An AIDS model of the third kind? Trans Actuar Soc South Afr. 1998;13(1):99–153.
18. Johnson L. THEMBISA version 1.0: A model for evaluating the impact of HIV/AIDS in South Africa. Centre for Infectious Disease Epidemiology and Research working paper. Thembisa Project; 2014. p. 6. [cited 2017 Dec 18] Available from: http://www.publichealth.uct.ac.za/sites/default/files/image_tool/images/108/THEMBISA%20version%201.0.pdf
19. UNAIDS, World Health Organisation. Report on the global HIV/AIDS epidemic [homepage on the Internet]. 1998 [cited 2016 May 18]; p. 50–51. Available from: http://data.unaids.org/pub/Report/1998/19981125_global_epidemic_report_en.pdf
20. UNAIDS. UNAIDS quarterly update on HIV epidemiology/1Q 2010: Methods for estimating HIV incidence [Internet]. 2010 [cited 2016 May 18]. Available from: http://www.unaids.org/sites/default/files/media_asset/epi_alert_1stqtr2010_en_3.pdf
21. Stover J, Brown T, Marston M. Updates to the Spectrum/Estimation and Projection Package (EPP) model to estimate HIV trends for adults and children. Sex Transm Infect. 2012 Dec;88 Suppl 2:i11–16. https://doi.org/10.1136/sextrans-2012-050640
22. Stover J, Andreev K, Slaymaker E, Gopalappa C, Sabin K, Velasquez C, et al. Updates to the spectrum model to estimate key HIV indicators for adults and children. AIDS Lond Engl. 2014 Nov;28(4):S427–S434. https://doi.org/10.1097/QAD.0000000000000483
23. Eaton JW, Johnson LF, Salomon JA, et al. HIV treatment as prevention: Systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. PLoS Med. 2012 Jul 10;9(7):e1001245. https://doi.org/10.1371/journal.pmed.1001245
24. Hontelez JAC, Lurie MN, Bärnighausen T, et al. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: A model comparison study. PLoS Med. 2013 Oct;10(10):e1001534. https://doi.org/10.1371/journal.pmed.1001534
25. Williams B. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: Cautions, caveats and the importance of parsimony. ArXiv14037104 Q-Bio [Internet]. 2014 Mar 26 [cited 2015 Nov 16]; Available from: http://arxiv.org/abs/1403.7104
26. Auvert B, Sobngwi-Tambekou J, Cutler E, et al. Effect of male circumcision on the prevalence of high-risk human papillomavirus in young men: Results of a randomized controlled trial conducted in Orange Farm, South Africa. J Infect Dis. 2009 Jan 1;199(1):14–19. https://doi.org/10.1086/595566
27. Williams BG, Lloyd-Smith JO, Gouws E, et al. The potential impact of male circumcision on HIV in sub-Saharan Africa. PLOS Med. 2006 Jul 11;3(7):e262. https://doi.org/10.1371/journal.pmed.0030262
28. Verguet S. Efficient and equitable HIV prevention: A case study of male circumcision in South Africa. Cost Eff Resour Alloc. 2013;11:1. https://doi.org/10.1186/1478-7547-11-1
29. Bärnighausen T, Bloom DE, Humair S. Economics of antiretroviral treatment vs. circumcision for HIV prevention. Proc Natl Acad Sci USA. 2012 Dec;109(52):21271–21276. https://doi.org/10.1073/pnas.1209017110
30. Pretorius C, Stover J, Bollinger L, Bacaër N, Williams B. Evaluating the cost-effectiveness of pre-exposure prophylaxis (PrEP) and its impact on HIV-1 transmission in South Africa. PLoS One. 2010;5(11):e13646. https://doi.org/10.1371/journal.pone.0013646

31. Hallett TB, Baeten JM, Heffron R, Barnabas R, de Bruyn G, Cremin Í, et al. Optimal uses of antiretrovirals for prevention in HIV-1 serodiscordant heterosexual couples in South Africa: A modelling study. PLoS Med [Internet]. 2011 Nov [cited 2016 May 12];8(11). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3217021/

32. Abbas UL, Glaubius R, Mubayi A, Hood G, Mellors JW. Antiretroviral therapy and pre-exposure prophylaxis: Combined impact on HIV transmission and drug resistance in South Africa. J Infect Dis. 2013 Jul;208(2):224–234. https://doi.org/10.1093/infdis/jit150

33. Coffee M, Lurie MN, Garnett GP. Modelling the impact of migration on the HIV epidemic in South Africa. AIDS. 2007 Jan;21(3):343–350. https://doi.org/10.1097/QAD.0b013e328011dac9

34. Mah TL, Halperin DT. Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward. AIDS Behav. 2008 Jul;14(1):11–16. https://doi.org/10.1007/s10461-008-9433-x

35. Johnson L, Dorrington R, Bradshaw D, Pillay-Van Wyk V, Rehle T. Sexual behaviour patterns in South Africa and their association with the spread of HIV: Insights from a mathematical model. Demogr Res. 2009 Sep;21(11):289–340. https://doi.org/10.4054/DemRes.2009.21.11

36. Rehle TM, Hallett TB, Shisana O, et al. A decline in new HIV infections in South Africa: Estimating HIV incidence from three national HIV surveys in 2002, 2005 and 2008. PLoS One. 2010 Jun;5(6):e11094. https://doi.org/10.1371/journal.pone.0011094

37. Sawers L, Stillwaggon E. Concurrent sexual partnerships do not explain the HIV epidemics in Africa: A systematic review of the evidence. J Int AIDS Soc. 2010;13(1):34. https://doi.org/10.1186/1758-2652-13-34

38. Nyabadza F, Mukandavire Z, Hove-Musekwa SD. Modelling the HIV/AIDS epidemic trends in South Africa: Insights from a simple mathematical model. Nonlinear Anal Real World Appl. 2011 Aug;12(4):2091–2104. https://doi.org/10.1016/j.nonrwa.2010.12.024

39. Johnson LF, Hallett TB, Rehle TM, Dorrington RE. The effect of changes in condom usage and antiretroviral treatment coverage on human immunodeficiency virus incidence in South Africa: A model-based analysis. J R Soc Interface R Soc. 2012 Jul;9(72):1544–1554. https://doi.org/10.1098/rsif.2011.0826

40. Tanser F, Bärnighausen T, Hund L, Garnett GP, McGrath N, Newell M-L. Effect of concurrent sexual partnerships on rate of new HIV infections in a high-prevalence, rural South African population: A cohort study. Lancet. 2011 Jul 22;378(9787):247–255. https://doi.org/10.1016/S0140-6736(11)60779-4

41. Vickerman P, Watts C, Delany S, Alary M, Rees H, Heise L. The importance of context: Model projections on how microbicide impact could be affected by the underlying epidemiologic and behavioral situation in 2 African settings. Sex Transm Dis. 2006 Jun;33(6):397–405. https://doi.org/10.1097/01.olq.0000218974.77208.cc

42. Gumel AB, McCluskey CC, van den Driessche P. Mathematical study of a staged-progression HIV model with imperfect vaccine. Bull Math Biol. 2006 Nov 1;68(8):2105–2128. https://doi.org/10.1007/s11538-006-9095-7

43. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The role of sexually transmitted infections in the evolution of the South African HIV epidemic. Trop Med Int Health TM IH. 2012 Feb;17(2):161–168. https://doi.org/10.1111/j.1365-3156.2011.02906.x

44. UNAIDS. 90-90-90 An ambitious treatment target to help end the AIDS epidemic [homepage on the Internet]. 2014 [cited 2016 Apr 4]. Available from: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en_0.pdf

45. South African National AIDS Council. South African National Strategic Plan on HIV, TB and STIs 2017–2022 (Draft 2) [homepage on the Internet]. 2017 [cited 2017 Jul 18]. Available from: http://sanac.org.za/wp-content/uploads/2017/03/NSP-Draft-2_24-February-2017.pdf

2.2 Afterword: Important models not described in the article

Because of limitations on space in the preceding article, a number of important models were left out. These models were located in a literature review of models (Johnson, 2014) and two article that did head-to-head comparisons of multiple models (Eaton et al., 2012, 2015). Here are brief notes on them.

The article notes that the Thembisa model integrates four pre-existing models, including ASSA2008 (Actuarial Society of South Africa, 2011; Johnson, 2014; Johnson et al., 2016). These are the STI-HIV, UCT Paediatric HIV and NSP ART Need models. These are all macro models that essentially implement advances on features of the ASSA model.

In STI-HIV sexual behaviour is calibrated to marriage data and cross-sectional data on numbers of partners. It was developed to "simulate changes in numbers of sexual partners, changes in marital status, changes in commercial sex activity and changes in the frequency of unprotected sex over the life course. This is extended to allow for the transmission of HIV, and the model is fitted to South African HIV prevalence data and sexual behaviour data. Results suggest that concurrent partnerships and other non-spousal partnerships are major drivers of the HIV/AIDS epidemic in South Africa" (Johnson et al., 2009).

UCT Paediatric HIV models children under 15 years. It includes prevention of mother-to-child transmission and antiretroviral treatment. One of the four models integrated into Thembisa. It was developed to estimate the effect of a change in treatment guidelines that provided for children to be

given antiretrovirals as soon as they were diagnosed with HIV, rather than based on CD4 cell percentage or counts or onset of AIDS. It concluded that the guideline change would significantly reduce paediatric AIDS mortality at young ages, "but further efforts are required to reduce the substantial growing AIDS mortality in older children" (Johnson et al., 2012).

The NSP ART Model uses CD4 staging instead of clinical staging that ASSA2008 uses, and allows for earlier treatment initiation. Its purpose was to assess rates of progression through CD4 stages based on ART coverage (Johnson, 2012).

Bacaër et al. (2010) developed a similar model to Granich et al. (2009) but included age structure, which the Granich model lacked. They found that because of high reported condom use, lower testing coverage than suggested by the Granich model would also lead to a long-term decline in HIV incidence in South Africa.

Eaton and Hallett (2014) implemented a macro model that estimates how HIV transmissions by people in the early phase of infection, usually before they are aware of their own status, affects incidence, even after the introduction of treatment as prevention (i.e. everyone who tests positive for HIV is offered ART). The "model includes stages of HIV infection, flexible sexual mixing, and changes in risk behavior over the epidemic. The model was calibrated to HIV prevalence data from South Africa using a Bayesian framework." The authors found: "Immediately after ART was introduced, more early transmission was associated with a smaller reduction in HIV incidence rate—consistent with the concern that a large amount of early transmission reduces the impact of treatment on incidence. However, the proportion of

early transmission was not strongly related to the long-term reduction in incidence."

EMOD is a microsimulation in which individuals have age and gender and there is heterogeneous sexual pairing through a complex pair-matching algorithm (Klein, 2012). It allows various assumptions to be made about how ART is provided. The model has been used to "estimate the potential impact of expanding treatment guidelines to allow earlier initiation of ... ART ... in sub-Saharan Africa with current or improved treatment coverage" (Klein et al., 2014).

The re-PopART model aimed to estimate the effect on incidence of interventions in a clinical trial called PopART (Cori et al., 2014). It is a macro model of heterosexual HIV transmission. Individuals do not have age (they're assumed to be over 15), but they do have gender. The model predicted that the trial intervention (ART offered to everyone who tested positive, plus some additional interventions) could reduce HIV population-level incidence by 60% over three years. The study is expected to complete in December 2017 (Network, 2016).

The Synthesis microsimulation is used to estimate the effects of long-term drug resistance on HIV mortality (Phillips et al., 2011; Cambiano et al., 2013, 2014). Agents have age and gender and are paired with other agents as the simulation proceeds. Agents can be in multiple relationships at a time.

The above is not intended to be a complete survey of HIV models. However, it is, to the best of our knowledge, a description of the most influential or important ones, especially with respect to the South African epidemic.

Chapter 3

Debates related to models of the South African HIV epidemic

3.1 Foreword

The following article, *A history of controversies involving HIV models in South Africa*, describes the often-heated life-and-death policy debates that surrounded HIV models from the 1990s through to 2015. While the article in Chapter 2 described how modelling is done, the purpose of this article is to show how models have been used and debated in the context of HIV policy-making.

A view underlying this dissertation is that technical work in the absence of context is an impoverished way of approaching modelling. Constructing models in a policy void risks being out of touch with societal needs and

therefore being an irrelevant, albeit intellectually stimulating, pastime. Examining the politics of modelling, i.e. the role of models in changing views and policies, is crucial for a rounded, richer understanding of them.

Appendix 1: A history of controversies involving human immunodeficiency virus (HIV) models in South Africa

1. Introduction

Disputes over models of the human immunodeficiency virus (HIV) epidemic have been public, vociferous, fraught and acrimonious. In this article, which supplements our article reviewing the main models of the South African HIV epidemic, we describe these debates. It is perhaps strange that the equations of obscure spreadsheets or highly technical computer code, understood only by a few specialists, should occupy so much attention in the media. But the stakes have been high, with millions of lives at risk. Also, as we explain in the main article, while, on the one hand, the technical workings of models might be understood by a few, on the other hand, the *model world*, that is, the conceptual realm of explicitly declared rules that match, in a simplified way, some aspect of the real world, can be understood by many.

2. The effect of models on policy in the 1990s

In 1990, although less than 1% of pregnant women attending public antenatal facilities in South Africa tested positive for HIV, modellers had predicted explosive growth during the subsequent decade, unless significant interventions were mounted. Indeed, by 2000, antenatal prevalence had increased to over 24%.[1] This failure to control the epidemic might suggest that modelling had no useful influence on policy, but in fact the situation was more complex.

Concern about the growing HIV epidemic, based on modelling data and antenatal survey results, led to the formation of a body called the Networking HIV, acquired immune deficiency syndrome (AIDS) Community of South Africa (NACOSA) in 1991.[2] In the transition to democracy in 1994, the African National Congress released a national health plan. It cites demographic modelling projections, stating that credible predictions 'indicate that by the year 2005, between 18% and 24% of the adult population will be infected with HIV' and that the 'cumulative death toll will be 2.3 million, and that there will be about 1.5 million AIDS orphans' (p. 30).[3] But the plan's recommendations, while cognisant of the rights of people with HIV, were mostly broad and vague, which, to be fair, was largely a reflection of the lack of effective interventions available at the time.

The small but growing AIDS activist movement was also aware of the model projections. The Treatment Action Campaign (TAC) was formed in late 1998. There are few early documents of the organisation's work still available, but a letter to the Pharmaceutical Manufacturers Association on 22 September 1999 notes that the HIV epidemic is an 'unprecedented health crisis in South Africa' and '3.5 million people are already infected with HIV and it is estimated that 150,000 people die of AIDS related illnesses every year'.[4] These figures were likely obtained from a model by the Actuarial Society of South Africa, such as ASSA600.[5]

3. The Medical Research Council report

A controversy with a model at its centre erupted in 2001. Dorrington et al. produced a technical report under the auspices of the South African Medical Research Council (MRC) that analysed mortality data collected by government agencies. The researchers also compared the mortality data with the projections of the ASSA600 model.[6]

The preface written by the president of the MRC at the time, Malegapuru Makgoba, stated there had been a shift in the age pattern of mortality in the country 'from the old to the young over the last decade particularly for young women – this is a unique phenomenon in biology', and 'this shift in mortality fits several AIDS models'. Makgoba wrote that the 'future burden' of the epidemic was 'broadly predictable from the models with reasonable confidence over the next decade' (p. 4).

The study investigated trends in reported deaths until 1996 based on data from Statistics South Africa (Stats SA), compared with more recent data (mid-1997 to September 2000) from the population register of the Department of Home Affairs. These empirical data were then compared with the outputs of the ASSA600 demographic model 'to assess the consistency of the empirical data with the model projections' (pp. 8–9). The authors used standard techniques for adjusting the data to take into account under-reporting of deaths.

The data showed a 'steady increase in adult mortality in the 1990s'. Women aged between 25 and 29 years had a 3.5 times higher death rate in 1999/2000 than in 1985 (p. 5).

The authors compared the empirical data with the projections of the ASSA600 demographic model, which they described as a 'behavioural demographic component projection model, which models the heterosexual epidemic for the country as a whole, ignoring race and geographical heterogeneity' (p. 19). The model was calibrated to reproduce the results of antenatal HIV surveys up to 1997.

The model projected antenatal infections between the values found for the 1999 and 2000 antenatal clinic surveys. It estimated that there would be between 4 and 7 million AIDS deaths from 2000 to 2010 in the absence of any interventions (behavioural change or treatment). The authors wrote that: 'given the pattern of deaths exhibited by the ASSA600 model … the … estimate of non-AIDS deaths is probably a little on the low side … and the AIDS deaths a little exaggerated.' (p. 24)

They also considered and critiqued the outputs of three other models: the Doyle one discussed in our main article,[7] the one by the United Nations and the one by the US Bureau of Census (pp. 24–25).

The report noted limitations of both the available data and models. Its recommendations included proposals for improving both. It stated: 'Considering these different sources of information, it seems highly probable that about 40% of the adult South African 1999/00 mortality in the 15–49 age group is due to HIV/AIDS.' (p. 37).

It briefly considered interventions to mitigate the effect of the epidemic: AZT for mother-to-child transmission prevention, promoting increased use of condoms and a national campaign to treat sexually transmitted infections. The authors concluded that the interventions 'can make a significant difference to the course of epidemic, although it will still exact a heavy toll' (p. 38).

They also wrote: 'Unfortunately the ASSA600 model was not designed to model the impact of antiretroviral therapies. Provided these drugs could be implemented successfully they could have a significant impact on the future prevalence levels.' (p. 38)

(Note: the interaction between antiretrovirals, incidence, prevalence and mortality is complex and still not fully resolved by today's models.)

The report was written against the background of the acrimonious debate in South Africa over the cause of AIDS, the size of the epidemic and whether antiretrovirals should be introduced in the public health system, both for mother-to-child transmission prevention as well as treatment (for a history of the AIDS denialist era, see Cameron).[8] In particular, in March 2001, the Presidential AIDS Advisory Panel Report had been released.[9]

This panel, constituted by President Mbeki, consisted of a roughly equal number of AIDS denialists and conventional scientists. It was criticised for promoting AIDS denialism.[10] The panel's report contained a very short section on modelling (p.44), essentially noting a fundamental disagreement on their utility. It also stated in a section on epidemiology that repeated requests for 'reliable data and statistics on the magnitude of the AIDS problem or even HIV prevalence' (p. 45) had not been provided to the panel. Yet, the MRC report did provide this, as did many other reports available during the time of the panel's deliberations, such as the Department of Health's annual antenatal clinic studies.

The MRC report was carefully researched and showed the growing impact of the epidemic on adult deaths. However, the MRC board, together with the Minister of Health, stopped or delayed its publication, possibly 'because it contradicted President Thabo Mbeki's view that the epidemic was being vastly exaggerated and that there were other, larger causes of death'.[11] However, findings from the report were leaked to the media. It was subsequently officially released in October 2001.[12] The tensions surrounding the MRC study are illustrated by a news report: '[*The study*] prompted a whole new furore around AIDS statistics and the reliability of MRC research. Stats SA was then used by government to rubbish the MRC report, a move that was yesterday slammed by one of the authors of the MRC report, University of Cape Town Actuarial Science professor Rob Dorrington. Pointing out that he was speaking in his personal capacity, Dorrington said it was a great shame that Stats SA had decided to trash the report. "It is clear that they have a limited understanding of the estimation process and model. Their (Stats SA) presentation was riddled with half truths and misunderstandings".'[9]

Instead of acting on the report's concerning findings, the state's response was to try to determine who leaked it. On 17 April 2002, Independent Newspapers published quotations from a letter obtained by reporter Lynne Altenroxel and written by the Minister of Health, Manto Tshabalala-Msimang, to the chair of the MRC board, Taole Mokoena, on 17 September 2001.[13] Tshabalala-Msimang wrote, 'this is not the first time that the MRC president has acted against government'.

She continued: 'You will recall that when the president of South Africa established a website for the members of the Presidential Advisory Council on AIDS to debate their different points of view, the MRC president was instrumental in establishing a separate website for the orthodox scientists, under the umbrella of the MRC.'[13]

She further wrote: 'The [*health department*] director-general [*Ayanda Ntsaluba*] advised the MRC president and his team not to release the report until the report had been presented to the minister of health and the cabinet.'[13]

Tshabalala-Msimang called for 'corrective action' to be taken. Makgoba was accused of being the source of the leak, but a subsequent investigation cleared him and three other MRC members. An earlier report by Independent Newspapers alleged that a private consultancy had been paid to find out the source of the leak.[13] Makgoba soon resigned from the MRC and became the vice-chancellor of the University of KwaZulu-Natal.

Dorrington would later write that the only institution that seriously questioned the finding by the MRC study that AIDS was the largest cause of mortality in South Africa by 2000, responsible for 25% of all deaths, was Stats SA. But, wrote Dorrington, '[*Stats SA*] have not produced any statistics of their own and have not claimed that the figure should in fact be lower'.[14]

Stats SA officials attempted to discredit the Actuarial Society of South Africa (ASSA) model results. The institution released a press statement claiming that the model gave lower

projected AIDS mortality by 2010 (1 to 2 million deaths vs. 5 or 6 million, in the absence of antiretroviral treatment) simply by changing the model's assumptions. But as Dorrington pointed out, Stats SA failed to calibrate the model under their assumptions to known prevalence data.[14]

4. Mother-to-child transmission prevention court case

When the TAC launched with a small protest at St George's Cathedral in Cape Town, one of the protesters' demands was for the ministers of health and finance to meet with AIDS organisations to 'plan for resources to introduce free AZT for pregnant mothers with HIV/AIDS'.[15]

Over the next 4 years, the TAC tried to convince the South African government to implement a countrywide mother-to-child HIV transmission prevention programme. The organisation, along with several others, proceeded with litigation against the national and provincial health ministers, eventually winning a seminal judgement at the Constitutional Court in July 2002. The court ordered the state to: 'devise and implement within its available resources a comprehensive and co-ordinated programme to realise progressively the rights of pregnant women and their newborn children to have access to health services to combat mother-to-child transmission of HIV.'[16]

In this and subsequent TAC litigation, we find clear examples of modelling being used to make an argument for policy changes to increase access to antiretroviral medicines.

Evidence put before the court by the TAC included an affidavit by Nicoli Nattrass, an economist at the University of Cape Town. She concluded that the: 'cost to the health sector of [*mother-to-child transmission prevention*] programmes ... is less than the costs of treating all children born HIV+ in the absence of a … programme. This is true for all … of the … programmes discussed here.'[17]

Nattrass performed costing analyses in her affidavit, and cited similar work by other researchers. Her model was relatively simple compared with most of those discussed here, but Nattrass's affidavit offered compelling arguments in favour of implementing mother-to-child transmission prevention. Although the role of Nattrass's affidavit in the court's decision is not mentioned explicitly in the court's judgement, her submission made it practically impossible for the state to offer a coherent financial argument against implementing the programme (p. 31).[16]

5. Competition commission complaints

In 2002, the TAC lodged a complaint with the Competition Commission against two pharmaceutical companies, GlaxoSmithKline and Boehringer Ingelheim, over what the organisation called the excessive pricing of the antiretroviral medicines zidovudine (AZT), lamivudine and nevirapine.[18]

Here again an expert affidavit describing the results of several models was placed on record.[14] The affidavit, written by Dorrington, explained the impact of HIV: 'According to the models referred to above, well over five million people are currently infected with the virus and, unless they receive treatment that would increase their life expectancy, most of these people will die within the next 10 years. It is clear that HIV/AIDS is estimated by all demographers outside government to be having a devastating effect on the population and is undoubtedly the leading cause of death these days in South Africa.' (p. 7)

In 2003, the TAC reached a settlement with the two companies that allowed generic manufacturers to sell the drugs in competition with them, not only in South Africa but also in sub-Saharan Africa.

The TAC followed up with other successful complaints and actions to lower antiretroviral prices. The effect on drug prices was profound: when Judge Cameron began taking antiretrovirals in 1997, the monthly cost of his regimen was R3419. By 2008, the standard regimen in the private sector cost under R240 per month.[11] The Dorrington affidavit played a small but significant role in this.

6. Pushing for the state to treat

After the TAC won the mother-to-child transmission prevention court case, the organisation stepped up its demand for antiretrovirals to be made generally available in the public health system for the treatment of HIV.

To make the case for this much more expensive and vast programme, a TAC researcher, with the assistance of the ASSA model developers, Nattrass and others, calculated the cost of a countrywide treatment programme using the outputs of the ASSA2000 demographic model. The article concluded that implementing treatment would incur substantial direct costs but potentially provide long-term savings from reduced hospitalisations and treatment of opportunistic infections. The publication of this article in 2003 was at the height of the conflict between the TAC and government over antiretroviral treatment.[19] The lead story in one of the country's leading weekly newspapers at the time, *Mail & Guardian*, describing this work, was 'Counting the cost of three million lives'.[20] Its findings were debated in subsequent issues of the *Mail & Guardian*.[21]

This was not the first such costing model. In October 2002, Boulle et al. modelled eight scenarios of a limited antiretroviral rollout.[22] This research did 'not explicitly link their numbers on treatment to an external demographic model', but they estimated the number of people needing treatment in their model would be about 10% of new HIV cases in an ASSA model.[23]

Another case of a model being used to advocate for treatment arose in mid-2003. The government had established the Joint Health & Treasury Technical Team, which used a model to

estimate the cost of implementing an antiretroviral treatment programme in the public health system, and the number of deaths that would be averted by such a programme. The Director-General of Health, Dr Ayanda Ntsaluba, presented the team's findings to the Health MinMec (the national and provincial ministers of health) on 9 May 2003.[24,25] The presentation considered an antiretroviral treatment programme scaled up over three years in public hospitals. It showed three scenarios: treating 20%, 50% and 100% of AIDS cases. In the 50% scenario, 600 000 people would be on treatment by 2008 at a cost of about R10 billion (about $1.3 billion at the 2003 exchange rate). According to slides not shown at the meeting, this scenario would 'defer' 733 000 deaths until after 2010, assuming that treatment led to '4–5 additional years of relatively illness-free life' (an extremely conservative assumption, on hindsight).

What model was used and how the results were calculated remains out of the public domain. The presentation was supposed to be a secret. However, based on a review of costing models by Boulle et al.,[23] it appears that this prescient costing model was likely developed by Fareed Abdullah, an official at the time in the health department, in March 2003.

The TAC obtained the presentation and leaked it to the media in July 2003. Accusations and counter-accusations followed. At the time, following pressure from TAC, including a civil disobedience campaign, negotiations for a treatment plan were taking place, at the National Economic Development and Labour Council (NEDLAC), between the state, labour, business and civil society organisations. Advocate Rams Ramashia, the Director-General of the Department of Labour, accused the TAC of breaching 'state security' and 'undermining and possibly de-railing the NEDLAC process'.[24]

The TAC responded: 'On a matter of such fundamental importance to millions of people's lives, the Constitutional right of access to information and the Constitutional duties that govern public administration are paramount. The notion that state security has been breached is ludicrous: in fact it is the personal security of millions affected and infected with HIV that is threatened by government procrastination.'[24]

By April 2004, the state began providing antiretroviral treatment to people with AIDS in the public health system. The programme stuttered in its first few years as Mbeki and Tshabalala-Msimang continued to undermine it, for example, by promoting untested remedies as alternatives. Eventually, the programme scaled up rapidly and is the largest of its kind in the world.

Since 2008, there has been considerable debate about the optimal clinical stage at which to start antiretroviral treatment. As described in our main article, models, such as the one by Granich et al., have played a central role in this debate.[26] Following the results of a clinical trial in 2015, the World Health Organization has recommended, and the South African government has adopted, a policy of universal access to treatment for all people with HIV.

7. AIDS denialists attack models

South Africa's period of state-supported scepticism of the link between HIV and AIDS under President Thabo Mbeki lasted from the late 1990s until Mbeki was removed from power by his own party in 2008. It's important to note that the governing party was not unified in its dismissal of the epidemic, and Mbeki's position ultimately delayed rather than entirely prevented the implementation of the public sector antiretroviral treatment programme, which began in 2004. Throughout this period, there was a constant battle between the supporters of the scientific position and the AIDS denialists, with the former slowly becoming ascendant until AIDS denialism ceased to be a relevant political force in South Africa.[11]

Mathematical models were at the centre of this conflict. The AIDS denialist attack on modelling did not come from a scientist but from a journalist, Rian Malan. Well-known for his best-selling non-fiction book *My Traitor's Heart*,[27] Malan wrote an article in *Rolling Stone* in 2001, disputing that there was a large HIV epidemic in Africa including South Africa. The article questioned HIV testing methodology on the continent and essentially accused UNAIDS of cynically exaggerating the size of the epidemic.[28] He followed this with articles in the British magazine *The Spectator*[29] and the South African magazine *Noseweek*.[30] A large part of the latter two articles was aimed at mathematical modelling of the epidemic, particularly the ASSA models.

A TAC researcher published a detailed rebuttal of Malan.[31] Interestingly, no rebuttal approaching the detail of the TAC's response was written by scientists with recognised expertise in demography, although Leigh Johnson, one of the main producers of the ASSA and subsequent models, assisted the TAC's researcher. It appears that academics found Malan's arguments so absurd that they were not worth more than a cursory occasional response in newspaper articles. This despite the fact that his three articles, which appeared in large-circulation popular publications, almost certainly were more widely read than the peer-reviewed publications on AIDS demographics in sub-Saharan Africa.

Malan's articles had numerous errors. In *Noseweek*, he miscalculated the number of South African HIV deaths from a Stats SA report, ignoring that many deaths owing to AIDS were not officially classified as AIDS deaths. He therefore reached the incorrect conclusion that they were a small fraction of the ASSA estimates.

The Stats SA report in fact made it clear that if physicians wrote the cause of death as, for example, tuberculosis, then this was not classified as an AIDS death, even though many such deaths are AIDS-related. The report stated that extricating the HIV-related deaths from the other death categories is where 'official statistics stop and research

begins' (p. 28).[32] The TAC added: 'Malan has not bothered with such research, which would be a very complex undertaking'.[31]

Mbeki mentioned Malan favourably in his 2004 State of the Nation speech, which a TAC researcher characterised thus: 'It was not explicitly about HIV, but to anyone following the debate at the time, it was clear that Mbeki was grateful for Malan's support on AIDS'.[33]

Running battles in print between Malan, activists and, to a lesser extent, scientists continued through the 2000s. In 2007, Malan published again in Noseweek, suggesting that the rise in recorded deaths was primarily owing to improved registration.[34] Grebe[35] pointed out on a website dedicated to refuting AIDS denialism, https://www.aidstruth.org/, that Malan continued to ignore the age pattern of deaths in South Africa, in which most recorded deaths were among young adults, as well as 'the increase in the recorded deaths resulting from causes typically associated with AIDS'.

In the post-Mbeki, and thus post-denialist, period, Minister of Health Aaron Motsoaledi delivered a presentation in which one of his slides had erroneously substantially overstated the number of 2008 deaths. Malan pounced on this in the online news site Politicsweb. Besides correcting the error, Malan wrote, '[T]here is no apocalypse. No massive AIDS related death surge. If anything, death registrations are stable'.[36] Malan's point was petty: Motsoaledi's slide had mistakenly transposed two digits, reporting 756 062 instead of 576 062 AIDS deaths.

The argument continued when Malan published his book, *Resident Alien*, in 2009.[37] It contained a chapter that reaffirmed his position, disputing there was a large HIV epidemic. The Daily Maverick, a popular South African news site, gave it a favourable review, and then published a critical reply by the TAC.[38,39]

After the removal of Mbeki from office, AIDS denialism no longer had any political force. Public debates over the size of the HIV epidemic receded.

8. The impact of AIDS denialism

After the rollout of antiretroviral treatment and the termination of Thabo Mbeki's presidency, two studies were conducted that calculated the loss of life owing to AIDS denialist policies.

Nattrass[40] used the ASSA2003 demographic model to estimate that if the national government had used antiretrovirals for mother-to-child transmission prevention and treatment of people with HIV at the same rate as the Western Cape province 'which defied national policy on ARVs', then 171 000 HIV infections and 343 000 deaths could have been prevented, just between 1999 and 2007.

A few months later, Chigwedere and colleagues at Harvard University used a different methodology but reached similar conclusions.[41] They considered what the South African government could have achieved had it scaled up treatment coverage from 5% in 2000 to 50% in 2005 instead of 3% to 23%. These estimates were actually more modest than what was achieved by Botswana or Namibia. Using a UNAIDS model accounting for the period 2000 to 2005, they concluded that delayed treatment caused 2.2 million lost person-years and over 330 000 deaths, and delayed mother-to-child transmission prevention caused over 35 000 excess infections and 1.6 million lost person-years. Both studies have been cited in an argument that Mbeki and the late South African Health Minister Manto Tshabalala-Msimang should have been prosecuted.[42]

9. Conclusion

This article has shown that mathematical models have been at the centre of policy debates and decision-making in the context of the South African HIV epidemic. In the early 1990s, models acted as a warning sign of the pending mortality that would be caused by the disease. In the late 1990s until the mid-2000s, modelling was a key point of discussion in the AIDS denialist controversy that characterised the government's response to the epidemic. Modelling also informed discussions on the relative efficacy of treatment and prevention options.

But, these examples also show the limitations of the influence of modelling over public policy. Despite the warnings of the early 1990s, little was done to stem the rise of HIV infections. And in the 2000s, the models were simply disputed by the AIDS denialists, so that antiretroviral treatment was delayed until 2004, and then only after an immense conflict between the state and AIDS activists, of which the dispute over modelling results was but one aspect. Also, no AIDS denialists have been held accountable for their role in hundreds of thousands of avoidable deaths, despite the estimates of Nattrass and Chigwedere et al.

Even the Granich model, cited an order of magnitude more often than any other model, had a limited effect on public policy. It was not until the publication of the HPTN 052 trial that there was consensus that test-and-treat would be effective at reducing new infections,[43] and it was not until the results of the START randomised control clinical trial in 2015 that there was consensus that antiretrovirals should be provided to all with HIV irrespective of CD4 count.[44] Although models have informed these debates, they do not carry the same weight as other forms of evidence in medicine, especially randomised controlled trials.

Perhaps, if consensus is reached on which modelling techniques produce the most robust estimates of past outputs and future projections – in other words if the science of modelling improves greatly – future models will be more effective at changing policy. We are aware of no developments that suggest this is likely to happen. Nevertheless, as a means

of exploring future scenarios and understanding the epidemic better, models have played a vital role.

10. Acknowledgements

Competing interests

Geffen was with the Treatment Action Campaign from 2000 to 2013 and an active participant in some of the debates discussed in this article.

Authors' contributions

N.G. conceived and drafted the article. N.G. and A.W. reviewed and edited the article.

11. Summary of Appendix 1

Mathematical models have helped describe and project the South African human immunodeficiency virus (HIV) epidemic. They have also informed, and been the subject of, public debates. We describe the main policy debates in which models had a crucial role, explaining how they were used to inform these debates, and we discuss the limits of how they influence policy. In the early 1990s, models were used to warn of the impending epidemic. The models of the early 2000s informed debates on treatment for people with acquired immune deficiency syndrome (AIDS) and prevention of mother-to-child transmission. Models were also at the centre of the AIDS denialist controversy. In more recent years, models have played a key role in the debate on when to start treatment.

12. References

1. Department of Health. The National HIV and Syphilis Prevalence Survey South Africa 2007 [homepage on the Internet]. 2008 [cited 2016 May 19]. Available from: https://dl.dropboxusercontent.com/u/193052/PrimaryDocuments/GovernmentDocs/DOH/antenatalSurveyfor2007Released2008WithIncorrectCalculations.pdf
2. McNeil J. A history of HIV/AIDS in South Africa [homepage on the Internet]. South Africa History Online; 2016 [cited 2016 May 19]. Available from: http://www.sahistory.org.za/topic/history-official-government-hivaids-policy-south-africa
3. African National Congress. A national health plan for South Africa. African National Congress Johannesburg; 1994.
4. Jara M. Open letter to the Pharmaceutical Manufacturers Association of South Africa: Join the partnership against AIDS [homepage on the Internet]. 1999 [cited 2016 May 19]. Available from: http://www.tac.org.za/community/node/2455
5. Dorrington RE. ASSA600: An AIDS model of the third kind? Trans Actuar Soc South Afr. 1998;13(1):99–153.
6. Dorrington RE, Bourne D, Bradshaw D, Laubscher R, Timaeus IM. The impact of HIV/AIDS on adult mortality in South Africa [homepage on the Internet]. 2001 [cited 2016 May 22]. Available from: http://www.mrc.ac.za/bod/
7. Doyle P. The demographic impact of AIDS on the South African population. In: Cross S, Whiteside A, editors. Facing up to AIDS [homepage on the Internet]. Palgrave Macmillan UK; 1993 [cited 2016 Feb 19]. p. 87–112. Available from: http://link.springer.com/chapter/10.1007/978-1-349-22597-2_5
8. Cameron E. Edwin Cameron: Witness to AIDS. 1st ed. Cape Town: Tafelberg; 2005.
9. South African Government. Presidential AIDS advisory panel report [homepage on the Internet]. 2001. [cited 2017 Dec 05]. Available from: https://www.gov.za/sites/www.gov.za/files/aidspanelpdf_0.pdf
10. Geffen N. Echoes of Lysenko: State-sponsored pseudo-science in South Africa. Soc Dyn. 2005;31(2):183–210. https://doi.org/10.1080/02533950508628713
11. Geffen N. Debunking delusions: The inside story of the treatment action campaign. 1st ed. Auckland Park, South Africa: Jacana Media; 2010. 256 p.
12. Health-e News. One in four deaths in SA related to HIV/AIDS – MRC report [homepage on the Internet]. 2001 [cited 2016 May 24]. Available from: https://www.health-e.org.za/2001/10/16/one-in-four-deaths-in-sa-related-to-hivaids-mrc-report/
13. Altenroxel L. Minister demanded 'corrective action' at MRC [homepage on the Internet]. IOL; 2002 [cited 2016 May 24]. Available from: http://www.iol.co.za/news/south-africa/minister-demanded-corrective-action-at-mrc-85131
14. Dorrington RE. Expert affidavit in TAC competition commission complaint [homepage on the Internet]. 2002 [cited 2016 May 18]. Available from: http://www.tac.org.za/Documents/DrugCompaniesCC/TAC_Dorrington_Final.doc
15. Treatment Action Campaign. HIV/AIDS treatment action campaign 10 December 1998 – Day of Action International Human Rights Day [homepage on the Internet]. 1998 [cited 2016 Feb 11]. Available from: http://www.tac.org.za/community/node/2454
16. Constitutional Court. Minister of health v treatment action campaign (TAC) (2002) 5 SA 721 (CC) [homepage on the Internet]. 2002 [cited 2016 Feb 11]. Available from: http://www.tac.org.za/Documents/MTCTCourtCase/ConCourtJudgmentOrderingMTCTP-5July2002.pdf
17. Nattrass N. Affidavit in treatment action campaign and others versus minister of health and others [homepage on the Internet]. Treatment Action Campaign; 2001 [cited 2014 Sep 04]. Available from: http://www.tac.org.za/Documents/MTCTCourtCase/affidavit/Nattrass.txt
18. Tau H, Zwedala NP, Godwana S, et al. Statement of complaint in terms of Section 49B(2)(b) of the Competition Act 89 of 1998 [homepage on the Internet]. Treatment Action Campaign; 2002 [cited 2016 May 29]. Available from: http://www.section27.org.za/wp-content/uploads/2010/10/TauvGSKevidenceAndLegalSubmissions.pdf
19. Geffen N, Raubenheimer C, Nattrass N. The cost of HIV prevention and treatment interventions in South Africa [homepage on the Internet]. Report No.: 28. Centre for Social Science Research; 2003 [cited 2017 Dec 05]. Available from: https://open.uct.ac.za/handle/11427/19732
20. Deane N. The price of three million lives. Mail & Guardian 2016 Mar 7;2.
21. Nattrass N. We need to fight aids and poverty [homepage on the Internet]. The M&G Online. 2003 [cited 2016 May 31]. Available from: http://mg.co.za/article/2003-03-25-we-need-to-fight-aids-iandi-poverty/
22. Boulle A, Kenyon C, Skordis J, Wood R. Exploring the costs of a limited public sector antiretroviral treatment programme in South Africa. S Afr Med J. 2002;92(10):811–817.
23. Boulle A, Kenyon C, Abdullah F. A review of antiretroviral costing models in South Africa. In: Moatti J-P, editor. Economics of AIDS and access to HIV/AIDS care in developing countries: Issues and challenges. Paris: ANRS; 2003. p. 293–309.
24. Treatment Action Campaign. TAC releases details of government costing study [homepage on the Internet]. 2003 [cited 2016 Jun 04]. Available from: http://www.tac.org.za/newsletter/2003/ns14_07_2003.htm
25. Treatment Action Campaign. Report of the joint health and treasury task team charged with examining options to supplement comprehensive HIV/AIDS care in the public health sector [homepage on the Internet]. 2003 [cited 2016 Jun 04]. Available from: http://www.tac.org.za/Documents/TreatmentPlan/slideshowongovernmentcostingstudy.pdf
26. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. Lancet. 2009;373(9657):48–57. https://doi.org/10.1016/S0140-6736(08)61697-9
27. Malan R. My traitor's heart: A South African exile returns to face his country, his tribe, and his conscience. Reprint edition. New York: Grove Press, 2000; 368 p.
28. Malan R. HIV & AIDS – AIDS in Africa – In search of the truth [homepage on the Internet]. 2001 [cited 2016 Jun 04]. Available from: http://www.virusmyth.com/aids/hiv/rmafrica.htm
29. Malan R. Africa isn't dying of AIDS [homepage on the Internet]. The Spectator 2003 Dec. [cited 2017 Dec 05]. Available from: https://www.spectator.co.uk/2003/12/africa-isnt-dying-of-aids/
30. Malan R. Apocalypse when [homepage on the Internet]. Noseweek 2003 Dec. [cited 2017 Dec 05]. Available from: https://www.noseweek.co.za/article.php?current_article=0629
31. Geffen N. Rian Malan spreads confusion about AIDS statistics [homepage on the Internet]. 2004 [cited 2016 Jun 04]. Available from: http://www.tac.org.za/newsletter/2004/ns20_01_2004.htm
32. Statistics South Africa. Causes of death in South Africa 1997–2001: Advance release of recorded causes of death. Pretoria: Statistics South Africa; 2002.
33. Geffen N. The shameless Rian Malan [homepage on the Internet]. 2009 [cited 2016 Jun 04]. Available from: http://www.tac.org.za/community/node/2778
34. Malan R. Rian Malan's last ever Aids piece – Or so he says... Noseweek [homepage on the Internet]. 2007 [cited 2016 Jun 04];(88). Available from: http://www.noseweek.co.za/article/1364/Rian-Malans-last-ever-Aids-piece--or-so-he-says
35. Grebe E. A response to Rian Malan's 'last ever AIDS piece' [homepage on the Internet]. 2007 [cited 2016 Jun 04]. Available from: https://old.aidstruth.org/features/malan
36. Malan R. Did Zuma & Motsoaledi get their AIDS stats wrong? [homepage on the Internet]. 2009 [cited 2016 Jun 04]. Available from: http://www.politicsweb.co.za/opinion/did-zuma--motsoaledi-get-their-aids-stats-wrong
37. Malan R. Resident alien. 1st ed. Johannesburg: Jonathan Ball Publishers SA, 2009; 380 p.
38. Bloom K. Review: Return of Rian Malan, the aids bore [homepage on the Internet]. 2009 [cited 2016 Jun 04]. Available from: http://www.dailymaverick.co.za/article/2009-12-23-review-return-of-rian-malan-the-aids-bore/#.V1Kr2p6fFyQ
39. Geffen N. TAC replies to our review: 'Rian Malan is shameless' [homepage on the Internet]. 2009 [cited 2016 Jun 04]. Available from: http://www.dailymaverick.co.za/article/2009-11-19-tac-replies-to-our-review-rian-malan-is-shameless/#.V1KqOZ6fFyQ
40. Nattrass N. AIDS and the scientific governance of medicine in post-apartheid South Africa. Afr Aff. 2008;107(427):157–176. https://doi.org/10.1093/afraf/adm087

41. Chigwedere P, Seage GR, Gruskin S, Lee T-H, Essex M. Estimating the lost benefits of antiretroviral drug use in South Africa. J Acquir Immune Defic Syndr. 2008;49(4):410–415. https://doi.org/10.1097/QAI.0b013e31818a6cd5

42. Geffen N. Justice after AIDS denialism: Should there be prosecutions and compensation? J Acquir Immune Defic Syndr. 2009;51(4):454–455. https://doi.org/10.1097/QAI.0b013e3181ab6da2

43. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. 2011;365(6):493–505. https://doi.org/10.1056/NEJMoa1105243

44. INSIGHT START Study Group. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med. 2015;373(9):795–807. https://doi.org/10.1056/NEJMoa1506816

Chapter 4

The "when to start" debate

4.1 Foreword

The following article, *When to start antiretroviral treatment? A history and analysis of a scientific controversy*, is a detailed description of one of the seminal HIV debates. It began with the trial of the first ARV, zidovudine, in the 1980s. It concluded with the publication of results of a randomised controlled clinical trial in 2015. Official treatment guidelines and prevailing views fluctuated dramatically over this three-decade period. As explained in the article, as it became clear that initiating people with HIV onto ART reduced their infectiousness, it raised the issue of what was in the best interest of patients versus public health.

The debate is particularly interesting from a modelling perspective. A model by Granich et al. (2009) [1] reignited the debate in the late 2000s. The authors' model estimated that providing treatment to everyone with HIV

[1]Strictly speaking, the authors published two different models in one article.

could eradicate the epidemic. No model has been cited as often, generated as much discussion, debate, praise and annoyance. The Granich et al. model was a very simple one, consisting of a homogeneous population of adults 15 years and above; there was no differentiation based on age, sex or risk group. A flurry of models followed in the wake of Granich et al. that attempted to give more sophisticated estimates. Whether they succeeded remains a matter of debate (Williams, 2014).

After the article the section titled *Understanding the importance of pair-matching in these models* (4.2) further discusses the role of modelling in the debate about when to start ART, and hence its role in this dissertation.

Southern African Journal of HIV Medicine
ISSN: (Online) 2078-6751, (Print) 1608-9693

When to start antiretroviral treatment? A history and analysis of a scientific controversy

Authors:
Nathan Geffen[1]
Marcus O. Low[2]

Affiliations:
[1]Department of Computer Science and Centre for Social Science Research, University of Cape Town, South Africa

[2]Spotlight Magazine, Cape Town, South Africa

Corresponding author:
Nathan Geffen,
nathangeffen@gmail.com

Dates:
Received: 28 Jan. 2017
Accepted: 18 Aug. 2017
Published: 05 Dec. 2017

How to cite this article:
Geffen N, Low MO. When to start antiretroviral treatment? A history and analysis of a scientific controversy. S Afr J HIV Med. 2017;18(1), a734. https://doi.org/10.4102/sajhivmed.v18i1.734

Background: Since 1987 HIV scientists and activists have debated the optimal point to start antiretroviral treatment. Positions have varied between treating people with HIV as soon as they are diagnosed, based on biological, modelling and observational evidence, versus delaying treatment until points in disease progression at which clinical trial evidence has shown unequivocally that treatment is beneficial.

Objectives: Examining the conduct and resolution of this debate may provide insight into how science works in practice. It also documents an important part of the history of the HIV epidemic.

Method: We describe clinical trials, observational studies, models and various documents that have advanced the debate from 1987 to 2015.

Results and conclusion: Evidence accumulated over the past decade, especially from randomised controlled clinical trials, has shown that immediate treatment both reduces the mortality and the risk of HIV transmission; it benefits both public health and the individual patient. By mid-2015, the debate was resolved in favour of immediate treatment.

Introduction

Since the publication of the first antiretroviral trial in 1987, scientists and patient advocates have debated the optimal time for people with HIV to start antiretroviral treatment. Guidelines, both national and international, have changed back and forth on this question, reflecting changes in expert opinion and new scientific developments.

We describe the when-to-start debate and its resolution in mid-2015. This debate exemplifies the problem of deciding policy when the evidence is still being collected, or how 'technological decision making' is done when there is 'scientific uncertainty'.[1] While much has been written about the debate over the cause of AIDS, a consequence mainly of former South African President Thabo Mbeki's views, little has been written on the when-to-start debate. Yet, there is more to be learnt about how science works from the when-to-start debate. This is because it was genuinely hard to determine public health policy from the limited evidence. By contrast, the science that HIV is the cause of AIDS was clear, and that debate was fuelled not by legitimate scientific disagreements, but by politics and ideology.

The question of when-to-start treatment was contested not only between scientists, but also between AIDS activists. Participants with reasonable claims to expertise who for the most part were familiar with the same scientific literature reached opposing conclusions on what treatment guidelines should recommend.

The participants in the debate differed in their assessments of the value of observational versus clinical trial data. They also differed on whether the public health benefits of reducing HIV transmission by treating people earlier outweighed the unknown harms to individual patients because of side effects of drugs, difficulties with adherence to lifelong medication and the development of drug resistance. And they differed on how much value to assign mathematical models and observational data. The stakes were high: the contestants understood that settling the question of when-to-start treatment might have considerable effects on life expectancy and the incidence of HIV.

If we think about the when-to-start debate as a court case, then the main exhibits were a mathematical model by Granich et al.[2] which showed that a policy of universal testing followed by immediate treatment of people with HIV would lead to the eradication of the disease; a clinical

trial that showed that people with HIV on antiretroviral treatment are unlikely to transmit the virus[3]; several observational studies, with inconsistent results, which compared what happened to patients who started treatment at different stages of HIV infection; and a massive multinational clinical trial called Strategic Timing of Antiretroviral Treatment (START).[4] Besides these exhibits, there were many others that either supported or contested some of the main ones.

The publication of results of the START trial in August 2015, a year-and-a-half ahead of schedule, effectively resolved the question of when-to-start treatment, generating broad scientific consensus on the question. But it did not and could not resolve differences in values and methodologies of the debate's participants. These differing values and methodologies in the approach to resolving medical science questions will continue, perhaps indefinitely, to be the subject of sociological and philosophical enquiry.

Background

The results of the first randomised controlled antiretroviral clinical trial, BW002, were published in 1987.[5] For 24 weeks, people with AIDS received azidothymidine (now better known as AZT or zidovudine) or placebo. Of the 145 participants who received AZT, one died, compared to 19 out of 137 who received placebo.

Despite this promising result, the trial was too short to show that monotherapy soon results in drug resistance followed by most patients developing AIDS illnesses again. New combination treatments were needed to reduce the risk of resistance.

New antiretrovirals went to trial and were approved by the United States Food and Drug Administration (FDA) through the 1990s: didanosine (1991), zalcitabine (1992), stavudine (1994) and lamivudine (1995). It was, however, the development of protease inhibitors and non-nucleoside reverse transcriptase inhibitors – such as saquinavir (1995), ritonavir (1996), indinavir (1996) and nevirapine (1996) – which changed the nature of HIV treatment.[6] Arts and Hazuda[6] write, 'The advent of combination therapy, also known as HAART, for the treatment of HIV-1 infection was seminal in reducing the morbidity and mortality associated with HIV-1 infection and AIDS'. People with HIV on combination therapy, typically three antiretrovirals taken daily for life, who adhere to their regimen have a very small risk of resistance. The virus can remain suppressed indefinitely restoring near-normal life expectancy.[7]

Today there are about 25 individual antiretroviral drugs spread over six different classes (i.e. differing modes of action) approved by the FDA.[8] But it was only in the second half of 2015, 28 years after the completion of the first randomised controlled antiretroviral trial that the answer to the when-to-start question was settled.

Changing guidelines

The two main criteria in treatment guidelines for determining when to start treatment have been symptoms of AIDS and CD4 T-lymphocyte count. The 'to and fro' of treatment guideline changes has previously been described.[9] When AZT was approved in 1987, the US Department for Health and Human Services (DHHS) set the CD4 threshold at 500 cells/μL. In April 2001, it was reduced to 350, and then to 200 in 2003. In 2007, it was raised to 350, and then 500 in 2009. In 2013, CD4 count was removed as a criterion for determining when-to-start treatment. In 2003, the World Health Organization (WHO) guidelines – produced for resource-limited settings – set the CD4 threshold at 200 cells/μL. This increased to 350 in 2010 and then 500 in 2013, with a recommendation that some groups of patients start irrespective of CD4 count. Changes over time in the CD4 initiation threshold can be found in the South African Department of Health's, British HIV Association's and European AIDS Clinical Society's guidelines. And often, they were not in sync with each other. For example, in 2012, these differed from the DHHS guidelines by retaining the 350 threshold. South Africa's guidelines have changed from 200 cells/μL to 350 to 500, followed by treatment irrespective of CD4 count.

One of the reasons why many scientists, clinicians and activists in the late 1990s and the early 2000s were reluctant to endorse early treatment for people was the surprising results of the Concorde trial.[10] Symptom-free people with HIV were enrolled in the trial from 1988 to 1991. Follow-up of the patients continued until they died or end of 1992, whichever came first. When the trial began, AZT was the only antiretroviral available. Participants were randomly assigned either to receive AZT immediately or to defer treatment until they developed AIDS symptoms or had persistently low CD4 counts. The trial was blinded: the deferred group received placebo, but upon developing signs of AIDS, participants were unblinded and offered AZT if they were on placebo.

There was no statistical difference in the primary outcome between the two arms: on the immediate arm, 176 of 877 people died or progressed to AIDS versus 171 of the 872 on the deferred arm. By starting treatment before they were ill, the immediate arm participants found no more benefit from AZT than those who deferred, and they were more likely to have become resistant to the drug so that by the time they did become ill, it was no longer beneficial.

That it was disadvantageous to start early was confirmed by a long-term follow-up of the trial participants who showed statistically significant worse survival in the immediate arm. But even then matters were not straightforward, because by pooling the results of a similar trial that was conducted at about the same time as Concorde, there was no significant difference between the deferred and immediate strategies.[11]

Even though these discouraging results were based on monotherapy, and the drug resistance this approach caused,

Concorde was a warning about jumping to the conclusion that early combination treatment would be beneficial.

With the growing success of combination therapy, Ho published an article in the *New England Journal of Medicine (NEJM)* in 1995 provocatively titled 'Time to Hit HIV, Early and Hard'. He wrote that recent scientific findings and therapeutic development favoured an 'aggressive interventional strategy early in the course of HIV-1 infection'.[12]

But the scientific findings Ho referred to were based on improved understanding of the pathogenesis of the disease. For some, this was unconvincing because it was not based on clinical data and did not consider drug side effects and long-term adherence challenges. In an article published in *The Lancet* entitled 'Hit HIV-1 hard, but only when necessary' by Harrington and Carpenter,[13] the authors argued for caution and a CD4 threshold of 350 cells/μL. They stated that:

> [N]o available regimen can eradicate HIV-1; all currently effective regimens may cause undesirable, sometimes life-threatening, toxic effects; and, unless regimens are strictly adhered to, multidrug resistance can develop, limiting future treatment options.

Through the 2000s, as various randomised controlled clinical trials were conducted, the when-to-start debate became increasingly nuanced. A trial showed that treating infants upon diagnosis reduced mortality by 76% and HIV progression by 75%.[14] Two trials in adults showed that a threshold of 350 cells/μL resulted in better outcomes than 250 or 200.[15,16] But the question of whether to treat adults irrespective of CD4 cell count, or to wait until it declined to some optimal value remained unanswered, at least in clinical trials.

A mathematical model causes a stir

Studies of antenatal transmission of HIV as well as observational data showing that sexual transmission was more likely if the infected partner's viral load was higher suggested that antiretroviral treatment could be used to reduce new infections.[17] Based on these findings, Granich et al.[2] published results of two mathematical models. They found that if a policy of universal testing coupled with the offer of immediate treatment to people who were found to be HIV-positive was introduced in South Africa, incidence and mortality because of the disease could be reduced to 'less than one case per 1000 people per year by 2016, or within 10 years of full implementation of the strategy'. They wrote that the prevalence of HIV could be 'less than 1% within 50 years'.

The authors included leading WHO researchers, including Kevin de Cock, the director of its HIV department. Its publication, while not responsible for starting the discussion on whether the CD4 count initiation criterion should be dispensed with and a policy of universal testing and immediate treatment should be pursued, certainly escalated the intensity of the debate. At the time of writing the article has been cited over 1600 times according to Google Scholar. This is extraordinarily high for mathematical models, the details of which most scientists, activists and policymakers are unlikely to understand, even though these were relatively simple models, which was part of their appeal.

The article's findings were first presented ahead of World AIDS Day in 2008, and the response to it was divided. Email correspondence at the time by leaders of the Treatment Action Campaign (TAC), the leading AIDS activist organisation in South Africa, conveyed both the excitement and scepticism the article generated. The organisation's leader, Zackie Achmat wrote, 'This is going to overwhelm us with calls. [Our policy department] will draft a statement. The heavens are opening up' (Achmat Z, personal communication, n.d.).

Another leader of the organisation, Mark Heywood, wrote:

> I heard Kevin de Cock present this paper in Geneva … I have serious concerns about it, as does Peter Piot and most at UNAIDS! It has the potential to create a great deal of confusion, so our statement will have to be very careful. You should also be aware that in meetings to justify the paper de Cock is also claiming it has the support of activists... (Heywood M, personal communication, n.d.).

Heywood was a co-signatory on a statement by a group of 'independent experts advising UNAIDS on HIV and human rights' published on World AIDS Day 2008. While welcoming 'a model that proposes the attainment of universal access to HIV treatment and HIV testing', that 'confirms the critical link between HIV prevention and HIV treatment', the authors wrote the study did not 'really address' the problems of stigma and discrimination which could be exacerbated by potentially coercive approaches. They wrote:

> To be both effective and just, programmes to scale-up HIV testing and treatment must be based on evidence and must protect the human rights of both the non-infected and the infected.

They cautioned about 'the application of theoretical models to fictitious populations'.[18]

The publication of the Granich et al. article was accompanied by letters from accomplished researchers in various fields of HIV who criticised various aspects of the model:

- The 'hypothesis that suppressive antiretroviral therapy can reduce HIV transmission within a sexual relationship is plausible, but unproven', wrote Cohen et al.,[19] scientists who within a few years would indeed prove the protective effect of treatment within a sexual relationship.
- They underestimated infectiousness in early infection and overestimated the number of partners South Africans report having, wrote Harvard demographers.[20]
- Harold Jaffe, who was at the forefront of the discovery of the AIDS epidemic, and his colleagues pointed out that the risks and benefits of treating people with a CD4 count above 350 cells/μL were unknown. They wrote, 'Trials of therapy for patients with higher counts are yet to begin. Within the field of communicable diseases, we are aware

of little precedent for the approach of "treating for the common good".'[21]

- Ethiopian public health officials described the difficulties of implementing mass testing in a resource-limited setting.[22]

More complex models were developed in the aftermath of the Granich et al. article, though none achieved as much public discussion. Twelve models, including one of the Granich et al. ones, were described in an article by Eaton et al.[23] The model results were compared under a set of similar assumptions about how universal testing and treatment would be carried out versus if the South African treatment guideline criteria at the time (with a CD4 initiation threshold of 350 cells/μL) were used.

The authors concluded that although the models evaluating the impact of treatment 'vary substantially in structure, complexity, and parameter choices', all suggested that treatment at 'high levels of access and with high adherence' would reduce new infections. Although there 'was broad agreement regarding the short-term epidemiologic impact of ambitious treatment scale-up', the models varied on their 'longer term projections' and 'in the efficiency with which treatment can reduce new infections'.

One of the most sophisticated set of models aimed at determining the effect of universal testing and treatment on the epidemic was published by Hontelez et al.[24] Explaining the motivation for their study, they wrote:

> there are as many different conclusions as there are models that investigated the issue. As models are profoundly different in many aspects – structure, parameterization, and assumptions about the intervention – it is difficult to determine which factors are responsible for the differences in the model predictions. (p. 2)

The period since the publication of the Granich et al. model had also produced new evidence that the authors relied upon.

The authors developed nine structurally different models of increasing complexity, starting with one that resembled that of Granich et al. In contrast to the set of relatively simple differential equations that characterised the Granich et al. model, their most complicated models simulated people (usually referred to as agents in simulation literature) with complex algorithms for choosing sexual partners. Their results confirmed that 'universal testing and immediate treatment at 90% coverage' would eliminate the HIV epidemic in South Africa. But they also found that their models, which they claimed were more realistic, 'show that elimination is likely to occur at a much later point in time than the initial model suggested'. They also found that universal testing and treatment is cost-effective, but less so than calculated by Granich et al. Most interestingly, they found that 'the current South African … treatment policy alone could already drive HIV into elimination'.[24]

However, it is controversial whether adding complexity to models improves them. One of the authors of the Granich et al.'s article, Brian Williams, a leading figure in mathematical modelling of infectious diseases, has written a response questioning their methodology. He writes:

> Hontelez et al. suggest that the current scale-up of ART at CD4 cell counts less than 350 [cells/μL] will lead to elimination of HIV in 30 years. I disagree … and believe that their more complex models rely on unwarranted and unsubstantiated assumptions.[25]

Williams' view was that there was already sufficient evidence to make treatment universally available. He wrote:

> the challenge now is to mobilize the political will and the financial support to make early treatment available to all that want it in order to save lives, save money and stop AIDS.[25]

Treatment as prevention

Observational studies published between 2006 and 2011 showed that people with HIV on antiretroviral treatment were likely less infectious.[26,27,28,29] But a clinical trial was needed to remove the possibility of confounding factors and estimate the magnitude of the effect.

In July 2011, the results of the HPTN 052 study were presented to a standing ovation at the meeting of the International AIDS Society in Rome. A month later the results were published. In this multinational randomised controlled trial of 1700 sero-discordant couples, the partner with HIV was randomly assigned to receive treatment immediately or to delay until 250 cells/μL. This partner also had to have a CD4 count between 350 cells/μL and 550 cells/μL at enrolment, which took place between 2007 and 2010.

Using genetic analysis, the authors found that in the immediate group, there was only one transmission to the HIV-negative partner. In the deferred group, there were 27 such transmissions, meaning that the transmission rate in the immediate group was 96% lower.[3] To date, this remains the most beneficial HIV sexual transmission prevention effect found in any randomised controlled clinical trial.

The study also found that there were clinical benefits for patients who started earlier, but as the initiation threshold was 250 cells/μL, a point already known to be lower than optimal (although the most convincing clinical trial showing this had not yet completed at the time HPTN 052 enrolled), it did not resolve the when-to-start debate, at least not from the perspective of the individual patient. However, it did result in the WHO publishing guidelines that recommended immediate treatment – for the purpose of prevention – for HIV-positive people with HIV-negative sexual partners.[30]

The debate on when-to-start swung noticeably towards earlier treatment after the publication of HPTN 052. Here is some of the discussion that followed.

Joseph Sonnabend, a physician, wrote a blog expressing caution against immediate treatment of anyone who tested positive and had a CD4 count above 350 cells/μL:

> The recent demonstration that antiretroviral treatment can prevent transmission of HIV among sero-discordant heterosexual couples is great news. However, when the person offered treatment has not yet been shown to personally benefit from it, an ethical issue needs to be addressed.[31]

In an interview, the study's principal investigator, Myron Cohen, stated his support for the earlier treatment recommendations made to the US guidelines following HPTN 052. 'That's a pretty big change', he said, 'and it respects the accrued benefits, which are very, very strong'.[32]

In a critical response to Cohen, AIDS activist Simon Collins[33] wrote:

> a radical public health approach to HIV care is presented as self evident, while neglecting to discuss the lack of important data or presence of contradictory evidence. This is a serious omission in an historical context of guideline recommendations that have been wrong on this question more often than they have been right.

He further wrote:

> Even with the best intentions, guidelines produced by experts, can be wrong. The limited evidence and lack of randomised data, restricts the ability to know the risks as well as the benefits.[33]

Given the state of uncertainty about the optimal initiation threshold and that many sexually active people would want to start treatment to reduce their infectiousness, Collins and Geffen wrote:

> the decision of when to start must be taken by the HIV-positive person in consultation with their health worker based on accurate information. That choice will vary depending on a person's individual health, their reason to want to treat and the resources of the health-care facility.[9]

Observational data

In April 2009, two large studies were published that had a considerable impact on the when-to-start debate. Both used observational data to calculate the effect on mortality of different CD4 count initiation thresholds.[34,35]

Kitahata et al.[34] studied over 17 000 Canadian and US patients. They found a substantial increase in the risk of death for people who deferred treatment below a CD4 count of 500 cells/μL. Those who deferred to below 350 cells/μL had the highest risk of death. However, the study used novel methods that introduced bias in favour of earlier treatment. The authors were criticised for this in several subsequent letters to the editor. They responded that even taking these concerns into account, their data still supported earlier treatment.

An accompanying editorial pointed out that the strengths of this study:

> included its relatively large size, the use of advanced statistical methods that attempted to analyze the data in a fashion similar to that of a randomized trial, and the use of survival … as the end point.[36]

Nevertheless:

> the results of the … study cannot be considered definitive evidence that everyone with HIV should start receiving antiretroviral therapy. This was not a randomized trial, and the patients who chose to begin therapy early might have differed in other important ways from those who chose to defer therapy – ways that improved survival but were not measured.[36]

The editorial concluded that if five years previously an asymptomatic patient with HIV with a CD4 cell count above 500 cells/μL wished to start treatment, most experienced clinicians 'could have made an excellent case' for deferring treatment.

> Today, if a similar patient were eager to start, we should be ready and willing to prescribe therapy – with ongoing careful monitoring of toxic effects that could arise during decades of treatment.[36]

But the UK funded when-to-start consortium,[35] which looked at over 21 000 patient records, had less convincing results with less controversial methods. The authors found that deferring therapy to a CD4 count of 250–350 cells/μL was associated with higher rates of a composite endpoint of AIDS or death than deferring to 351–450. However, when mortality alone was considered, there was no statistical significance. And at stepwise comparisons of higher CD4 count ranges, they could find no significant difference in the primary outcome. The authors noted, 'The evolution of guidelines has been compared to the swings of a pendulum'. They motivated for a 350 cells/μL threshold.

Subsequently, UK and US guidelines diverged, with the latter taking steps in subsequent editions that promoted earlier treatment.

Jain and Deeks[37] summarised the situation at the time:

> Although the debate regarding when to start antiretroviral therapy has been present for over two decades, consensus on this question has been hard to achieve. This lack of clarity continues in the current era, with major guidelines recommending very different treatment strategies. All agree, however, that the pendulum has swung back in favor of more aggressive approaches to therapy. The philosophy of delaying potentially toxic medications as long as possible has increasingly shifted toward a philosophy of initiating therapy as soon as possible.

This shift was evident when UNAIDS published its 90–90–90 strategy in October 2014.[38] The second of the three 90s referred to having 90% of people diagnosed HIV-positive on sustained antiretroviral treatment by 2020 – a target that amounts to an endorsement of test and treat. But this was at odds with the WHO's treatment guidelines, which at the time only recommended treatment initiation at CD4 counts of 500 cells/μL or below.[39]

In his budget vote speech in July 2014, shortly after returning from the 20th International AIDS Conference in Melbourne Australia, South Africa's Minister of Health endorsed the 90–90–90 targets and treatment irrespective of CD4 count. While he endorsed test and treat, he only went as far as announcing that the treatment initiation threshold would be raised from 350 cells/μL to 500 cells/μL.[40] In response, the TAC's policy director criticised Motsoaledi for recommending earlier treatment initiation without consulting activists.[41]

Strategic Timing of Antiretroviral Treatment (START)

Members of the trial's community advisory board wrote that the Strategic Timing of Antiretroviral Treatment (START) trial:

> is a study that has been driven by community demand that the optimal clinical initiation threshold for [antiretroviral treatment] be determined by clinical trial evidence rather than expert opinion informed primarily by observational data.[42]

START was conceived in the mid-2000s to resolve definitively the question when it would be best to start treatment from the perspective of a patient with HIV. The trial was randomised but open-label because a placebo arm would have created insurmountable practical and ethical problems.[4]

Nearly 4700 people enrolled in the trial at 215 sites in 35 countries between April 2009 and December 2013. To participate, patients had to be antiretroviral treatment naive, and have a CD4 cell count greater than 500 cells/μL. Participants were randomised either to begin treatment immediately or to wait until their CD4 counts dropped to 350, or treatment was clinically indicated. The primary endpoint was a composite of any serious AIDS- or non-AIDS-related event, or death.[43]

Support for the trial was not universal. Franco and Saag[44] wrote that the balance of data strongly supported starting treatment in nearly everyone regardless of CD4 count. They cited the availability of better drugs that were now available, the current understanding of HIV biology and pathogenesis and evidence from observational data. They conceded that a small group of people who 'have undetectable virus in the absence of antiretroviral therapy' might be exceptions. But for:

> everyone else, to wait on randomized clinical trial data could well be doing harm. The time spent waiting is time that the patients cannot get back and the long-term damage associated with waiting could well be irreversible.[44]

By the time they wrote this, however, START was well underway.

Contrast this with a view expressed by some of the main researchers involved in the publication of the observational data, after the US guidelines changed. Phillips et al.[45] wrote 'We are concerned that some may interpret the new recommendations as implying that the deferral group of this trial [START] is no longer ethical. Such an interpretation would endanger the future of the trial in the USA'. After explaining the problems with the observational data, they concluded:

> We therefore do not believe that there is convincing evidence to conclude that deferral of initiation of ART to a CD4 count of 350 causes net harm, particularly in terms of mortality, compared with starting at any higher level. We strongly support continued enrolment into START. Large randomised studies represent the only means of eventually obtaining the definitive result we need to properly inform future patient care.

The trial was only expected to produce results in late 2016 or early 2017. But in May 2015, the trial's independent data and safety monitoring board informed the main sponsor that the question had been answered. It recommended that the findings be 'immediately disseminated'. The primary endpoint occurred in 42 people in the immediate arm versus 96 in the deferred one, meaning the risk of serious illness or death was less than half in the immediate one (though for an HIV cohort, patient outcomes were good in both arms). Even at high CD4 counts, treatment reduced the risk of AIDS illnesses.[43]

At about the same time as START, a smaller trial (2076 participants) called TEMPRANO was run in Côte d'Ivoire. Primarily concerned with the effect of earlier treatment in an area with high tuberculosis, it had a similar primary endpoint to START. Its findings, which also showed the benefit of immediate treatment, were made public shortly before START's were. However, the details of the study were presented at the International AIDS Conference in Vancouver on the same day as START's and both studies were published in the same journal on the same day.[46]

The last outstanding piece of evidence in the when-to-start controversy had now been answered.

Discussion

There were two main points of contention in the when-to-start debate.

First, there was disagreement over what constituted sufficient evidence that early treatment was beneficial. There was an implicit hierarchy of evidence, with biological plausibility and mathematical models constituting the lowest evidence, followed by observational data. For many, this was sufficient to make the case for immediate treatment.

But those who considered observational data to be too uncertain and prone to confounding demanded a randomised clinical trial. They were concerned by the additional adherence demanded of patients who started treatment early in their HIV infection, the side effects primarily associated with earlier generations of antiretroviral drugs, the unfortunate experience of Concorde and the fluctuations of guideline recommendations in the absence of compelling evidence. There was also concern that the public health concern of the reduced infectiousness of people with HIV

was taking precedence over the uncertainty about the benefit of immediate treatment for the health of individuals with HIV. The publication of the START results resolved these points of contention.

Conclusion

There are no accepted criteria for resolving scientific debates with policy repercussions when the evidence is still being gathered. In contrast to the destructive and irrational debate on the cause of AIDS that took place in South Africa in the 2000s, the protagonists on both sides of the when-to-start debate included leading experts in HIV science who could draw on substantial evidence to make their arguments. In the case of AIDS denialism, one side of the debate shunned the immense body of evidence, preferring conspiracy theories instead, whereas on the question of when-to-start, the science truly was in dispute (see[1,47,48,49] for further discussion on this). Now that all the data, including the gold standard of medical science – a randomised controlled trial – support immediate treatment, continuing to advocate for delayed treatment on medical grounds would be irrational.

Acknowledgements

Thank you to Nicoli Nattrass and Martin Weinel for comments.

Competing interests

N.G. is a member of the Community Advisory Board of INSIGHT which runs the START trial. Both authors have been members of the Treatment Action Campaign which has been involved in the debates described here.

Authors' contributions

N.G. conceived and drafted the article. Both authors reviewed and edited the article. N.G. is the corresponding author.

References

1. Weinel M. Technological decision making under scientific uncertainty [homepage on the Internet] [PhD thesis]. Cardiff: Cardiff University; 2010 [cited 2015 Nov 10]. Available from: http://orca.cf.ac.uk/55502/1/U564796.pdf
2. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. Lancet. 2009;373(9657):48–57. https://doi.org/10.1016/S0140-6736(08)61697-9
3. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. 2011;365(6):493–505. https://doi.org/10.1056/NEJMoa1105243
4. Babiker AG, Emery S, Fatkenheuer G, et al. Considerations in the rationale, design and methods of the Strategic Timing of AntiRetroviral Treatment (START) study. Clin Trials Lond Engl. 2013;10(10):S5–S36. https://doi.org/10.1177/1740774512440342
5. Fischl MA, Richman DD, Grieco MH, et al. The efficacy of azidothymidine (AZT) in the treatment of Patients with AIDS and AIDS-related complex. N Engl J Med. 1987;317(4):185–191. https://doi.org/10.1056/NEJM198707233170401
6. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med [serial online]. 2012 [cited 2015 Nov 11];2(4). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312400/
7. Sabin CA. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy? BMC Med. 2013;11:251. https://doi.org/10.1186/1741-7015-11-251
8. Food and Drug Administration. FDA-approved HIV medicines [homepage on the Internet]. AIDSinfo. 2015 [cited 2015 Nov 11]. Available from: https://aidsinfo.nih.gov/education-materials/fact-sheets/21/58/fda-approved-hiv-medicines
9. Collins S, Geffen N. Community views: Balancing the public health benefits of earlier antiretroviral treatment with the implications for individual patients – Perspectives from the community. Curr Opin HIV AIDS. 2014;9(1):4–10. https://doi.org/10.1097/COH.0000000000000024
10. Concorde Coordinating Committee. Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. Lancet. 1994;343(8902):871–881. https://doi.org/10.1016/S0140-6736(94)90006-X
11. Joint Concorde and Opal Coordinating Committee. Long-term follow-up of randomized trials of immediate versus deferred zidovudine in symptom-free HIV infection. AIDS Lond Engl. 1998;12(11):1259–1265.
12. Ho DD. Time to hit HIV, early and hard. N Engl J Med. 1995;333(7):450–451. https://doi.org/10.1056/NEJM199508173330710
13. Harrington M, Carpenter CC. Hit HIV-1 hard, but only when necessary. Lancet. 2000;355(9221):2147–2152. https://doi.org/10.1016/S0140-6736(00)02388-6
14. Violari A, Cotton MF, Gibb DM, et al. Early antiretroviral therapy and mortality among HIV-infected infants. N Engl J Med. 2008;359(21):2233–2244. https://doi.org/10.1056/NEJMoa0800971
15. SMART Study Group. CD4+ count–guided interruption of antiretroviral treatment. N Engl J Med. 2006;355(22):2283–2296. https://doi.org/10.1056/NEJMoa062360
16. Severe P, Jean Juste MA, Ambroise A, et al. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. N Engl J Med. 2010;363(3):257–265. https://doi.org/10.1056/NEJMoa0910370
17. Montaner JSG, Hogg R, Wood E, et al. The case for expanding access to highly active antiretroviral therapy to curb the growth of the HIV epidemic. Lancet Lond Engl. 2006;368(9534):531–536. https://doi.org/10.1016/S0140-6736(06)69162-9
18. UNAIDS Reference Group on HIV and Human Rights. Time for action towards universal access to prevention, treatment, care and support: Beyond theory towards practice and protection [homepage on the Internet]. 2008 [cited 2015 Nov 15]. Available from: http://www.hivhumanrights.org/statements/time-for-action-towards-universal-access-to-prevention-treatment-care-and-support/
19. Cohen MS, Mastro TD, Cates Jr, W. Universal voluntary HIV testing and immediate antiretroviral therapy. Lancet. 2009;373:1080.
20. Ruark A, Shelton JD, Halperin DT, Wawer MJ, Gray RH. Universal voluntary HIV testing and immediate antiretroviral therapy. Lancet. 2009;373(9669):1078. https://doi.org/10.1016/S0140-6736(09)60642-5
21. Jaffe H, Smith A, Hope T. Universal voluntary HIV testing and immediate antiretroviral therapy. Lancet. 2009;373:1080. https://doi.org/10.1016/S0140-6736(09)60646-2
22. Assefa Y, Lera M. Universal voluntary HIV testing and immediate antiretroviral therapy. Lancet. 2009;373:1080. https://doi.org/10.1016/S0140-6736(09)60647-4
23. Eaton JW, Johnson LF, Salomon JA, et al. HIV treatment as prevention: Systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. PLoS Med. 2012;9(7):e1001245. https://doi.org/10.1371/journal.pmed.1001245
24. Hontelez JAC, Lurie MN, Bärnighausen T, et al. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: A model comparison study. PLoS Med. 2013;10(10):e1001534. https://doi.org/10.1371/journal.pmed.1001534
25. Williams B. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: Cautions, caveats and the importance of parsimony. ArXiv14037104 Q-Bio [homepage on the Internet]. 2014 [cited 2015 Nov 16]; Available from: http://arxiv.org/abs/1403.7104
26. Bunnell R, Ekwaru JP, Solberg P, et al. Changes in sexual behavior and risk of HIV transmission after antiretroviral therapy and prevention interventions in rural Uganda. AIDS Lond Engl. 2006;20(1):85–92. https://doi.org/10.1097/01.aids.0000196566.40702.28
27. Del Romero J, Castilla J, Hernando V, Rodríguez C, García S. Combined antiretroviral treatment and heterosexual transmission of HIV-1: Cross sectional and prospective cohort study. BMJ. 2010;340:c2205. https://doi.org/10.1136/bmj.c2205
28. Donnell D, Baeten JM, Kiarie J, et al. Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: A prospective cohort analysis. Lancet Lond Engl. 2010;375(9731):2092–2098. https://doi.org/10.1016/S0140-6736(10)60705-2
29. Reynolds SJ, Makumbi F, Nakigozi G, et al. HIV-1 transmission among HIV-1 discordant couples before and after the introduction of antiretroviral therapy. AIDS Lond Engl. 2011;25(4):473–477. https://doi.org/10.1097/QAD.0b013e3283437c2b
30. World Health Organisation. Guidance on couples HIV testing and counselling – Including antiretroviral therapy for treatment and prevention in serodiscordant couples [homepage on the Internet]. WHO. 2012 [cited 2015 Nov 18]. Available from: http://www.who.int/hiv/pub/guidelines/9789241501972/en/
31. Sonnabend J. Treatment as prevention: Protecting patient autonomy [homepage on the Internet]. 2011 [cited 2015 Nov 16]. Available from: http://blogs.poz.com/joseph/archives/2011/06/treatment_as_prevent_1.html
32. Smith M. HIV/AIDS in practice: When to start therapy, a clinical context report [homepage on the Internet]. 2013 [cited 2015 Nov 16]. Available from: http://www.medpagetoday.com/clinical-context/HIVAIDS/36634
33. Collins S. Why the 'when to start' question is complex and informed by limited evidence: A response to Dr Myron Cohen. HIV Treat Bull [serial online]. 2013 [cited 2015 Nov 16]. Available from: http://i-base.info/htb/21038
34. Kitahata MM, Gange SJ, Abraham AG, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. N Engl J Med. 2009;360(18):1815–1826.

35. When To Start Consortium, Sterne JAC, May M, et al. Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: A collaborative analysis of 18 HIV cohort studies. Lancet. 2009;373(9672):1352–1363. https://doi.org/10.1056/NEJMoa0807252

36. Sax PE, Baden LR. When to start antiretroviral therapy – Ready when you are? N Engl J Med. 2009;360(18):1897–1899. https://doi.org/10.1056/NEJMe0902713

37. Jain V, Deeks SG. When to start antiretroviral therapy. Curr HIV/AIDS Rep. 2010;7(2):60–68. https://doi.org/10.1007/s11904-010-0044-6

38. UNAIDS. 90-90-90 An ambitious treatment target to help end the AIDS epidemic [homepage on the Internet]. 2014 [cited 2016 Apr 4]. Available from: http://www.unaids.org/sites/default/files/media_asset/90-90-90_en_0.pdf

39. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infections: Recommendations for a public health approach [homepage on the Internet]. 2013 [cited 2016 Apr 4]. Available from: http://apps.who.int/iris/bitstream/10665/85321/1/9789241505727_eng.pdf

40. Motsoaledi A. Health budget vote speech by the Minister of Health [homepage on the Internet]. 2014 [cited 2016 Apr 4]. Available from: http://www.health-e.org.za/wp-content/uploads/2014/07/Minister-of-Health-Budget-Vote-Speech-2014-15-.pdf

41. Low M. HIV fight requires wisdom [homepage on the Internet]. M&G Online. 2014 [cited 2015 Nov 18]. Available from: http://mg.co.za/article/2014-11-27-comment-hiv-fight-requires-wisdom/

42. Geffen N, Aagaard P, Corbelli GM, et al. Community perspective on the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial. HIV Med. 2015;16(Suppl 1):10–13. https://doi.org/10.1111/hiv.12228

43. INSIGHT START Study Group. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med. 2015;373(9):795–807. https://doi.org/10.1056/NEJMoa1506816

44. Franco RA, Saag MS. When to start antiretroviral therapy: As soon as possible. BMC Med. 2013;11(1):147. https://doi.org/10.1186/1741-7015-11-147

45. Phillips A, Costagliola D, Sabin C, Sterne J. Early initiation of treatment for HIV infection. Lancet. 2010;375(9715):639. https://doi.org/10.1016/S0140-6736(10)60269-3

46. The TEMPRANO ANRS 12136 Study Group. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. N Engl J Med. 2015;373(9):808–822. https://doi.org/10.1056/NEJMoa1507198

47. Geffen N. Debunking delusions: The inside story of the treatment action campaign. 1st ed. Auckland Park, South Africa: Jacana Media; 2010. 256 p.

48. Nattrass N. Defending the boundaries of science: AIDS denialism, peer review and the medical hypotheses saga. Sociol Health Illn. 2011;33(4):507–521. https://doi.org/10.1111/j.1467-9566.2010.01312.x

49. Nattrass N. The AIDS conspiracy: Science fights back. 1st ed. New York, NY: Columbia University Press; 2012. 240 p.

4.2 Understanding the importance of pair-matching in these models

The debate on the optimal point following HIV infection to begin treatment exemplifies the role of modelling in policy debates, and allows for an explanation on the role of pair-matching. Although the when-to-start debate long preceded Granich et al. (2009), their article, cited over 1,600 times, brought the question to the fore of public discussion. Yet it is an extremely simple model. The model world consists of adults, without gender or age, either on or not on ART, and with a reduced mortality rate and infectiousness if they are on ART.

The transmission function works as follows: On each time step of the model, additional people become infected as a function of the proportion of people in the population who are infected and not on treatment and the proportion of people who are infected and are on treatment. The greater the latter proportion relative to infected people not on treatment, the lower the proportion of new infections (or incidence). Heterogeneity is also accounted for very simply: as a greater portion of the population becomes infected, it is reasonable to assume the more saturated is the pool of people at risk. Therefore a factor is included in the transmission function so that incidence reduces as prevalence increases. See Figure 1.2 for details.

By contrast, in more complex macro models, such as Thembisa (Johnson, 2014), the population is divided into many more compartments (usually based on age, gender and risk group) and each compartment has a transmission function, with higher risk groups having higher transmission rates. This

accounts for heterogeneity in a more sophisticated way than Granich et al.

In microsimulations this can be taken further with agents representing individual people being paired in partnerships. The risk of infection is simulated per partnership, or even per sexual act. The key challenge in these microsimulations is how to pair agents into relationships. We could randomly pair agents, but then that defeats the purpose of using a microsimulation in the first place, because we are failing to account for heterogeneity. Alternatively we could pair agents according to a distribution of relationships in the population we're modelling as in Emod (Klein, 2012), STDSIM (Hontelez et al., 2013) and Johnson and Geffen (2016). A problem that then arises is that pair-matching algorithms can be very slow. The technical computer science work in this dissertation is to develop pair-matching algorithms that are fast and reasonably approximate the distribution of sexual relationships in the population being modelled.

4.2.1 Critique of additional complexity

In the article *When to start antiretroviral treatment? A history and analysis of a scientific controversy* the criticisms we describe by Williams (2014) — one of the authors of the Granich et al. model — of the STDSIM microsimulation by Hontelez et al. (2013) raises an important question about the additional complexity introduced by complex models such as microsimulations. Williams, among other criticisms, states that the authors have relied on unrealistic assumptions about heterogeneity in sexual behaviour. Indeed, little is understood about the distribution of sexual relationships and it is a

valid criticism that over-defining that distribution in a model doesn't add to our knowledge of the epidemic. On the other hand, the Granich model goes in the opposite direction, with its extremely simple assumptions of homogeneous sexual behaviour.

The question about what it makes sense for models to include needs to be kept in mind when implementing pair-matching in a microsimulation. On the one hand more fine-grained simulation is taking place that apparently accounts for heterogeneous sexual behaviour, but on the other hand it should be asked if this additional complexity is based on quality data about what is actually happening in the real-world population that we are modelling.

Chapter 5

Pair-matching Algorithms

5.1 Foreword

The articles in this chapter and the next constitute the original technical work of this PhD thesis.

The following article, *Efficient and effective pair-matching algorithms for microsimulations*, presents and analyses algorithms for matching agents into pairs (Geffen and Scholz, 2017). Six pair-matching algorithms, conceived and implemented by this author, are described and compared: Brute force, Distribution counting, Cluster shuffle, Weighted shuffle, Random k and Random. Cluster shuffle, Weighted shuffle, Distribution counting and Random k pair-matching are novel. However, Brute force and Random pair-matching are too obvious to be considered original, and are no doubt implemented in other work without being named as such. The article in Chapter 6 implements some of the algorithms in an actual microsimulation of a fictitious STI.

Chapter 2 described most of the important models of the South African HIV epidemic. It differentiated between two types of models: deterministic compartmental (or macrosimulation) versus agent-based models (or microsimulation). The former type of model divides the population into compartments and then uses differential equations to estimate outputs. By contrast in microsimulations, each individual in a population is represented by an agent. The potential advantage of microsimulations is that they allow for fine-grained modelling that allows us to capture much more heterogeneous behaviour. For STIs, microsimulations provide the opportunity for much more sophisticated modelling of individuals starting and ending sexual relationships. However, it is difficult to do this in a way that captures the distribution of the relationships in a population and does so speedily enough to make the microsimulation practical. The following article presents algorithms we have developed that aim to achieve this.

The algorithms are implemented in discrete microsimulations, There appears to be no barrier to implementing them in event-driven microsimulations in continuous time, but this needs to be tested in future research.

The algorithms presented in the paper require a domain-specific function, called a distance function, that measures the suitability of two agents for pairing. Although the distance functions implemented in the paper are quite simple, they can in theory be made as complex as desired. They can even account for a situation in which the agent attributes are multivariate with complex covariance structures. But a practical barrier to this is that the distance functions may then become too slow, rendering the model impractical beyond small numbers of agents.

5.1.1 Reassessment of claim about Blossom V's practicality for some simulations

Between the following article being accepted for publication and the submission of the article presented in Chapter 6, further research caused the authors to tone down one of the (minor) claims made in the following article. In this chapter's article the authors state that the Blossom V algorithm is not usually practical for microsimulations because it is too slow and not stochastic.

In fact, the algorithm may be practical for some simulations. The authors successfully executed it on a population of 100,000 agents in the article presented in Chapter 6 (the mating pool on each time step was considerably smaller than this, approximately 400 agents each day of the ten-year simulation on average). It took about 50 minutes for a single simulation. [1] Also, although the Blossom V algorithm is not stochastic, the selection of agents for the mating pool is stochastic, and this may be sufficient.

5.1.2 Errata

Since the paper is published, errors detected post-publication are described here:

Table 1 states "90% of males are men who have sex with men, 90% of females are women who have sex with women". It should instead state "90% of males are men who have sex with women, 90% of females are women who have sex with men".

[1] This is far too slow however if we wish to run thousands of simulations in order to fit parameters or do sensitivity analysis.

Efficient and Effective Pair-Matching Algorithms for Microsimulations

Nathan Geffen[1] **and Stefan Scholz**[2]

[1]*Department of Computer Science, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa*
[2]*Center for Health Economics Research Hannover (CHERH), Leibniz University of Hannover, Universitätstr. 25, 33615 Bielefeld, Germany*
Correspondence should be addressed to *nathangeffen@gmail.com*

Journal of Artificial Societies and Social Simulation 20(4) 8, 2017
Doi: 10.18564/jasss.3485 Url: http://jasss.soc.surrey.ac.uk/20/4/8.html

Received: 16-11-2016 Accepted: 11-05-2017 Published: 31-10-2017

Abstract:

Microsimulations and agent-based models across various disciplines need to match agents into relationships. Some of these models need to repeatedly match different pairs of agents, for example microsimulations of sexually transmitted infection epidemics. We describe the requirements for pair-matching in these types of microsimulations, and present several pair-matching algorithms: Brute force (BFPM), Random (RPM), Random k (RKPM), Weighted shuffle (WSPM), Cluster shuffle (CSPM), and Distribution counting (DCPM). Using two microsimulations, we empirically compare the speeds, and pairing quality of these six algorithms. For models which execute pair-matching many thousands or millions of times, BFPM is not usually a practical option because it is slow. On the other hand RPM is fast but chooses poor quality pairs. Nevertheless both algorithms are used, sometimes implicitly, in many models. Here we use them as yardsticks for upper and lower bounds for speed and quality. In these tests CSPM offers the best trade-off of speed and effectiveness. In general, CSPM is fast and produces stochastic, high quality pair-matches, which are often desirable characteristics for pair-matching in discrete time step microsimulations. Moreover it is a simple algorithm that can be easily adapted for the specific needs of a particular domain. However, for some models, RKPM or DCPM would be as fast as CSPM with matches of similar quality. We discuss the circumstances under which this would happen.

Keywords: Agent-Based Modelling, Pair-Matching, Partner Matching, Sexually Transmitted Infections, HIV

Introduction

1.1 Microsimulations and agent-based models (ABMs) are increasingly used across a broad area of disciplines, i.e. biology (Gras et al. 2009), sociology (Macy et al. 2002), economics (Deissenberg et al. 2008) and epidemiology (Gray et al. 2011). In many of these applications an artificial society of agents, usually representing humans or animals, is created, and the agents need to be paired with each other to allow for interactions between them.

1.2 Zinn (2012) addresses some of the conceptual challenges of finding suitable pairs of agents, particularly with respect to closed continuous ABMs. The author differentiates between stochastic versus stable matching rules, discusses different measures of compatibility between agents (which we call distance functions) and which agents choose their partners and which only get chosen. While the paper provides a conceptual framework of matching procedures, computational algorithmic aspects are left out and only little research can be found on this topic. Bouffard et al. (2001) presents an algorithm for finding suitable pairs for marriage. However this algorithm would be too slow for the repeated pair-matching required in many microsimulations of sexually transmitted infections (STIs).

1.3 The identification of suitable pairs of agents for partnerships may be difficult to compute efficiently. If a simulation does not measure the compatibility of two agents, i.e. it assumes that agents behave uniformly and match randomly with one another, then it risks failing to model essential features that determine the outcomes being studied.

1.4 On the other hand if a model attempts to measure the compatibility of agents, so that it produces sets of agent partnerships that very closely match the population being studied, it may become too computationally slow to study large populations; it may even be too slow for small populations where agents seek new partners repeatedly. The choice of pair-matching algorithm therefore needs to consider the trade-off between accuracy and speed to ensure the feasibility of the microsimulation for the research question at hand. This is especially true if the model is stochastic and needs to be run many times for the construction of confidence intervals or sensitivity analyses.

1.5 Two algorithms are commonly used for pair-matching in microsimulations. One randomly matches agents in a mating pool into pairs. While this does not usually yield good quality matches with respect to the actual distribution of partnerships in the population being studied, it is fast: the time of the pair-matching procedure is linear with the number of agents being matched.

1.6 The second algorithm is a brute force approach (using common computer science terminology, e.g. Levitin (2011)). This algorithm compares each agent to every other unmatched agent in the mating pool, using a compatibility or distance function, in order to choose appropriate partnerships. While this method usually yields good quality matches, it is slow, with execution time quadratic with the number of agents being matched.

1.7 The aim of our study is to present several alternative pair-matching algorithms for microsimulations and ABMs, which allow for a better trade-off between good quality matches and computation time than the random and brute-force methods. After the general problem statement and a description of general characteristics of pair-matching algorithms, we provide a detailed description of the matching algorithms. As this paper focuses on the computational features, we analyse the respective relative efficiency and effectiveness of the algorithms compared to brute force and random matching algorithms for one abstract microsimulation, and one microsimulation that models some features of a natural population in the field of epidemiology.

1.8 We have provided open-source C++ implementations of the algorithms, although modellers will likely wish to adapt them for their domains. We hope that these algorithms will lead to faster more accurate, easier-to-implement microsimulations, especially, but not only, of STI epidemics.

Background

General problem statement

2.1 The typical structure of a discrete microsimulation or ABM is that the simulation is divided into time steps, and in each time step events execute over all the agents. For example, see Algorithm 1.

Algorithm 1 Structure of a discrete microsimulation

```
1: for each time step do
2:     for each event e do
3:         for each agent a do
4:             if e has to be applied to a then
5:                 Apply e to a
6:             end if
7:         end for
8:     end for
9: end for
```

2.2 The practicability of a simulation is dependent on the efficiency of the events. A particularly simple event is to age each agent. This merely requires incrementing an age property for each agent. The execution time is linear with the number of agents. We usually want the efficiency class of events to be linear, or at worst linearithmic (i.e. the execution time is proportional to $n \log n$, where n is the number of agents being matched). An event whose time efficiency is quadratic (n^2) with the number of agents being matched will slow simulations with large numbers of agents to the point that it may become unfeasible to generate confidence intervals or conduct sensitivity testing.

2.3 Pair-matching represents a more complex event involving the properties of more than one agent and can be stated formally as follows: We have a set of n agents eligible for pairing $a_1, a_2 ... a_n$. We have a distance function $distance$ which takes two agents as its parameters such that if $distance(a, b) < distance(a, c)$ then b is

a more likely, compatible, suitable or appropriate match for a than c. (If it is possible that $distance(a,b) = distance(a,c)$ the modeller must decide on a tie-breaking mechanism.) A pair-matching solution is a set of matches such that every agent is paired with exactly one other agent. This can be recast as a fully-connected graph problem such that every vertex is an agent and every edge is a distance between two agents.

2.4 There exists for such a graph with n vertices multiple sets of distinct pair-matchings. One or more of these is a minimum-distance — or perfect — set in that the sum of all the distances between the pairs is less than or equal to every other pair-matching set. There are algorithms that find the minimum-distance set of pair-matchings, such as Blossom V (Kolmogorov 2009; Cook & Rohe 1999). However, Blossom V suffers from two serious problems: it is far too slow for most microsimulations, and it is not stochastic — it always produces the same set of pairs, unless there are multiple perfect pair-matching sets, in which case the algorithm could be modified to produce a random tie-breaking mechanism. Usually, though, more stochasticism than this is required. Nor is reproducing the expected value of a probabilistic distribution with certainty usually a desirable statistical attribute of a microsimulation.

2.5 Instead we are interested in pair-matching algorithms that approximate the underlying distribution, and do so quickly. This paper presents several such pair-matching algorithms, some of them novel. They are tested, compared and analysed, including against the Blossom V algorithm.

Characteristics of pair-matching algorithms

2.6 Pair-matching algorithms may be described and categorized in general by the following characteristics:

Distance function If agents are not matched randomly, a distance function must be defined. This function indicates the compatibility of a pair of agents for a partnership based on the distribution of partnerships in the population being modelled, as described in 2.3.

Exact vs approximate In some applications, agents are only paired if they are an exact match, i.e. for two agents a and b, they are paired if and only if $distance(a,b)$ equals a defined value. In other applications, only an approximate match is necessary. In this paper the algorithms are compared using approximate matching applications. Nevertheless, the algorithms can all be adapted to do either approximate or exact matching.

Stochasticism Stochasticism in partner choice is usually desired in microsimulations, for example so that multiple executions of the simulation differ from each other. As agents are usually stored in an array data structure, or similar, pair-matching algorithms will process the array from front to back. If the agents that are processed first are more likely to find compatible partners than those processed last, as the pair-matching event is repeated in subsequent time-steps, the partner selections will become increasingly biased. An easy way to avoid this "storage" bias is through the introduction of randomness by shuffling agents using the technique described in Knuth (1997, pp.142–146) and that is implemented in most modern programming language standard libraries. Further randomness can be introduced if desired. For example, the maximum number of evaluated potential matching candidates or the threshold of the value of the distance function for accepting partners may be stochastic.

Effectiveness We call an algorithm's success at generating matches its effectiveness. We evaluate this using two measures. When exact matching is used, the first effectiveness measure is calculated as the proportion of the agent population, who are supposed to be in relationships, that are actually in relationships after the algorithm is executed across all agents. For approximate matching, effectiveness is the mean or median distance between all paired agents. The second effectiveness measure is defined as follows: For any agent a we can rank all other agents in order of distance from a. Then the effectiveness is the mean or median rank of all agent partners. Both measures are relative, if the best possible matching result is unknown (because it is too computationally demanding to calculate).

Initialisation vs in-simulation All the pair-matching algorithms discussed here are intended for execution as events during a simulation as in Algorithm 1. However, in some simulations, it is necessary to pair agents before the simulation starts. An example of an algorithm to do this is discussed in Scholz et al. (2016).

Distance functions

2.7 Euclidean distance would be a convenient measure of the suitability of pairing two agents a and b with m appropriately scaled properties $a_1, b_1, a_2, b_2 ... a_m, b_m$. For multi-dimensional space, the equation for Euclidean distance is:

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + ... + (a_m - b_m)^2}$$

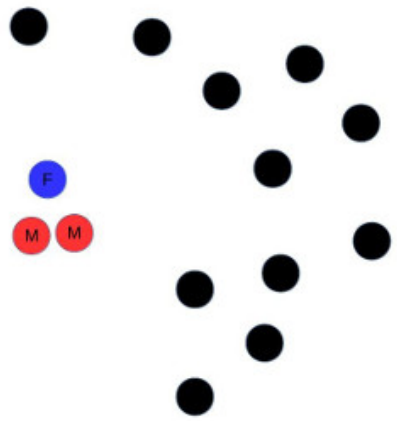

Figure 1: Heterosexual pair-matching violates the Triangle Inequality and does not map to a Euclidean plane. The two male agents, marked M, are closer on the Euclidean plane than the female agent, marked F, but if we are modelling heterosexual pair-matching then the distance function will record the two males as being further apart than the agent marked F.

2.8 Algorithms exist that efficiently but approximately find the nearest neighbour to a point in high-dimensional Euclidean space (Indyk & Motwani 1998; Beis & Lowe 1997; Arya et al. 1998). It is unclear how easily these can be adapted to microsimulations, but perhaps if we could map the properties of agents to Euclidean space, we could use one of these algorithms (or a variation thereof) to find suitable partners.

2.9 To map the matching properties to metric space, of which Euclidean space is one example, the distance function would have to satisfy the triangle inequality (Weisstein 2016). For agents a, b and c this is:

$$distance(a, b) + distance(b, c) \geq distance(a, c)$$

2.10 In some simulations the agent properties violate the triangle inequality. An example is the primarily heterosexual HIV epidemic in South Africa. Consider the agent properties we are interested in for measuring pairing suitability:

- age
- sex
- desire for a new partnership at this time
- risk behaviour propensity (including whether or not the agent is a sex worker)
- relationship status (including whether or not the agent is married)
- location

2.11 The more similar two individuals the closer they are to each other in Euclidean space, but this is not the case for heterosexual agents in an HIV epidemic: If two people share the same sex they are not suitable partners in this model. See Figure 1 for a graphical depiction of how this violates the triangle inequality.

2.12 The problem is not confined to modelling heterosexual relationships. Even if we want to model only men who have sex with men (MSM) with respect to an STI, it would be difficult to map our agents to Euclidean space. We might want our microsimulation to make it less likely for married agents seeking a new partnership to partner with other married agents, or with other agents currently in a relationship. When searching for a new relationship, we might want the distance to be great between otherwise well-matched agents who have been in a partnership with each other previously. We might also want to extend our model to account for people looking for partners of different age, wealth or education from themselves.

2.13 In our model's set of agent properties, age, risk behaviour propensity and geographical location may be attributes that map well to Euclidean space; we could define our distance function so that the more similar these are between agents, the more likely they are to form partnerships. However, sex for heterosexual agents does not map to Euclidean space, and, depending on the model's assumptions, neither might several other agent

attributes. (While it might be argued that sexual orientation is a categorical variable, a distance function simply returns a real number, and therefore has to represent incompatible categories of agents by calculating a very high value for such pairs of agents.)

2.14 We call a model's agent properties which do map well to Euclidean space *attractors*, and attributes that do not *rejectors*. Two of the algorithms we present (CSPM and WSPM) work by clustering agents together based on their attractors. Therefore clustering functions have to be defined by modellers who use these algorithms. However, when the rejectors dominate the distance function, the cluster function is less useful and the effectiveness of these two algorithms degrades.

Pair-matching algorithms

2.15 Besides the above mentioned Brute force pair-matching (BFPM) and Random pair-matching (RPM) algorithms, we describe the following four pair-matching algorithms.

- Random k pair-matching (RKPM)
- Weighted shuffle pair-matching (WSPM)
- Cluster shuffle pair-matching (CSPM)
- Distribution counting pair-matching (DCPM)

Brute force pair-matching (BFPM)

2.16 This algorithm works as follows: For every agent a in a set of agents, it calculates the distance to every remaining unpaired agent after a in the set. The agent b with the smallest distance to a is marked as a's partner (and vice-versa).

2.17 This is a naive algorithm that is very effective. However, unless pair matching only needs to be executed a few times, or on small populations, it is impractically slow — quadratic with the number of agents being paired. For example to run one simulation of a subset of the South African population to model the HIV epidemic for five years, we iterate over tens of thousands of agents daily or weekly, executing pair-matching on each iteration. Numerous simulations need to be run in order to build confidence intervals or to perform sensitivity testing (see for example Hontelez et al. (2013) and Johnson & Geffen (2016)). The *BFPM* algorithm is usually too slow for this purpose.

2.18 As BFPM, or variations thereof, is widely implemented in microsimulations and ABMs, we use it here as a reference against which to measure the effectiveness and speed of the other algorithms.

2.19 The pseudocode for *BFPM* is provided in Algorithm 2.

Algorithm 2 Brute force pair-matching (BFPM)

Parameters: $Agents$, an array of agents, with subscripts $0..n-1$, where n is the number of agents. If n is uneven, one agent will remain unmatched.

```
1:  function BruteForceMatch(Agents)
2:      shuffle(Agents)                          ▷ So that the algorithm is stochastic
3:      for each unmatched agent a in Agents do
4:          best ← ∞
5:          for each unmatched agent b after a in Agents do
6:              d ← distance(a, b)
7:              if d < best then
8:                  best ← d
9:                  bestPartner ← b
10:             end if
11:         end for
12:         Make a and bestPartner partners
13:     end for
14: end function
```

Random pair-matching (RPM)

2.20 Many deterministic models that use differential equations to model population dynamics or epidemiology, as well as simple microsimulations make the simplifying assumption that people are randomly paired with their sexual partners. This corresponds to a random matching algorithm using no distance function. Although this is extremely fast, with execution speed linear with the number of agents being matched, by ignoring the compatibility of matched agents it may result in the simulation producing outputs that differ wildly from the population being modelled.

2.21 As with BFPM, this algorithm's primary purpose is as a yardstick against which to measure the other algorithms. A good algorithm will execute much faster than BFPM but perhaps considerably slower than RPM. The effectiveness of a good algorithm should far exceed RPM. Algorithm 3 presents the pseudocode for this algorithm.

Algorithm 3 Random pair-matching (RPM)

Parameters: $Agents$, an array of agents, with subscripts $0..n-1$, where n is the number of agents. For simplicity assume n is even.

```
1: function RANDOMMATCH(Agents)
2:     shuffle(Agents)
3:     for i ∈ 0, 2, 4.., n − 4, n − 2 do
4:         make Agents[i] and Agents[i + 1] partners
5:     end for
6: end function
```

Random k pair-matching (RKPM)

2.22 This algorithm is a small conceptual advance on RPM and BFPM (Pelillo 2014; Larose & Larose 2014). Instead of only considering all agents following a, the one under consideration (as in BFPM), or matching with the adjacent agent (as in RPM), we consider the k adjacent agents after a in the array of agents. We partner a with the agent with the smallest distance. Assuming k is a constant (which it is in our implementation), the efficiency of this algorithm is linear with the number of agents.

2.23 The pseudocode for *RKPM* is provided in Algorithm 4.

Algorithm 4 Random k pair-matching (RKPM)

Parameters:
$Agents$, an array of agents, with subscripts $0..n-1$, where n is the number of agents. For simplicity assume n is even.
k, the number of adjacent agents to consider when finding a suitable partner

```
1: function RANDOMKMATCH(Agents, k)
2:     shuffle(Agents)
3:     for each unmatched agent a in Agents do
4:         best ← ∞
5:         for each unmatched agent b in one of up to k positions in the array after a do
6:             d ← distance(a, b)
7:             if d < best then
8:                 best ← d
9:                 bestPartner ← b
10:            end if
11:        end for
12:        Make a and bestPartner partners
13:    end for
14: end function
```

2.24 Surprisingly, unlike RPM the effectiveness of this simple algorithm is quite good in our applications. In effect, assuming n agents in the mating pool, the algorithm randomly draws k elements without replacement from an

ordered list of n elements, and selects l, the lowest of the k elements in the ordering. The mean position of l is $\frac{1}{k+1} \times n$ in the ordered list, and the median is $1 - e^{-\frac{\ln(2)}{k}} \times n$.

2.25 This means, for example, if we have $k >= 100$, then on average the partner chosen will be ranked in the top one percent in suitability from the remaining available agents.

Weighted shuffle pair-matching (WSPM)

2.26 This algorithm is an extension of RKPM, except that instead of ordering the agents entirely randomly, those with compatible pair-matching attributes are more likely to be clustered together. The algorithm works as follows:

2.27 Agents are assigned a cluster value. Then the cluster value is multiplied by a uniform random number — giving a random value weighted towards the cluster of the agent — to introduce stochasticism into the algorithm. Then the agents are sorted on this weighted value. Finally for each agent a, the agent with the smallest distance to a in the k adjacent agents is selected.

2.28 To calculate the cluster value for each agent, we need a domain-specific clustering function. This would typically be based solely on the attractor attributes, or a subset thereof, of the distance function. Despite being domain specific, a programmer without knowledge of the application can work out a cluster function solely by examining the distance function. However, it is likely that with greater domain knowledge, a more sophisticated cluster function can be defined.

2.29 Algorithm 5 is an example of a simple cluster function using some of the attractor attributes described above.

Algorithm 5 Example of a cluster function

1: **function** CLUSTER(a) ▷ a is an agent
2: **return** $AGE\_FACTOR * a.age + ORIENTATION\_FACTOR * a.orientation + RISK\_FACTOR * a.riskiness$
3: **end function**

2.30 Algorithm 6 provides pseudocode for WSPM.

Algorithm 6 Weighted shuffle matching (WSPM)

Parameters:
$Agents$**, an array of agents, with subscripts** $0..n-1$**, where** n **is the number of agents. For simplicity assume** n **is even.**
k**, the number of adjacent agents to consider when finding a suitable partner.**
The function `rand()` **returns a random number in the range** $[0..1)$**.**

1: **function** WEIGHTEDSHUFFLEMATCH($Agents$, k)
2: **for** each agent a in $Agents$ **do**
3: $a.weight \leftarrow cluster(a) * rand()$
4: **end for**
5: sort $Agents$ by weight
6: Execute lines 3 to 13 of *RKPM* (Algorithm 4).
7: **end function**

2.31 Assuming k is a constant, which it is in our implementation, then sorting has the slowest efficiency of the steps in this algorithm, linearithmic with the number of agents being matched. Therefore the efficiency of the entire algorithm is linearithmic with the number of agents.

Cluster shuffle pair-matching (CSPM)

2.32 This algorithm is a conceptual advance on WSPM. As with WSPM, it also extends RKPM and uses a cluster function. It works as follows:

2.33 The agents are sorted by the value returned by the cluster function (as opposed to a cluster weighted random number as in WSPM). They are then divided into a user-specified number of clusters. Then each cluster is shuffled to introduce stochasticism. (In contrast to WSPM, agents in the same cluster always remain relatively close

to each other.) Finally, just as with WSPM and RKPM, it finds the best of k agents after the agent under consideration.

2.34 Algorithm 7 provides the pseudocode for this algorithm.

Algorithm 7 Cluster shuffle pair-matching (CSPM)

Parameters:
$Agents$**, an array of agents, with subscripts** $0..n-1$**, where** n **is the number of agents. For simplicity assume** n **is even.**
c**, the number of clusters to divide the agents into. For simplicity assume** c **divides into** n**.**
k**, the number of adjacent agents to consider when finding a suitable partner.**

```
1:  function CLUSTERSHUFFLEMATCH(Agents, c, k)
2:      for each agent, a, in Agents do
3:          a.weight ← cluster(a)
4:      end for
5:      sort Agents by weight
6:      clusterSize ← n/c
7:      i ← 0
8:      for each cluster do
9:          first ← i * clusterSize
10:         last ← first + clusterSize
11:         shuffle Agents[first...last − 1]          ▷ to introduce stochasticism
12:         i ← i + 1
13:     end for
14:     Execute lines 3 to 13 of RKPM (Algorithm listing 4).
15: end function
```

2.35 As with WSPM, the efficiency of the entire algorithm is linearithmic with the number of agents, because the only step that is not linear is the sort, which is linearithmic.

Distribution counting pair-matching (DCPM)

2.36 This algorithm is influenced by sorting by counting (Knuth 1998, p 75-79), also called distribution counting sort (Levitin 2011, p 283). It is only useful if the agents can be distributed into buckets that exactly or approximately correspond to the properties of the partners with whom they need to be matched. Therefore a domain specific function that places a given agent into one of a set of predefined buckets must be defined.

2.37 The algorithm works as follows: Pointers to the shuffled agents are copied into an array. The copy is sorted using distribution counting which makes use of the bucket function (Levitin 2011, p 283). This sorting has linear efficiency — as opposed to the linearithmic efficiency of standard sorting algorithms — because it uses information about the underlying distribution of the objects provided by the bucket function.

2.38 A table with the same number of entries as there are buckets is then constructed so that given the "desired" partner properties of an agent, a, we can directly access the first agent with those characteristics in the copied agents array. We then linearly examine up to k agents, selecting the partner with the closest distance. The algorithm uses a defined $getBucket$ function which assigns an agent, or its "desired" partner to a bucket.

2.39 To understand how agents are assigned to buckets consider the agent properties of age, sex, and sexual orientation. If an agent representing a female "desires" a 25-year-old male, then it is assigned to the bucket of agents looking for 25-year-old heterosexual males. When looking for a partner, it will search up to k agents in this bucket and choose the one with the lowest distance to it. If the model's agents are either male or female, heterosexual or homosexual, and any age from 16 to 50, then there are $2 * 2 * 35 = 140$ buckets.

2.40 One caveat: Because age is a flexible characteristic (e.g. if the desired partner of a is 25 it does not mean that if b is a 24-year-old, it should have no prospect of partnership with a), the algorithm can be adapted so that it searches for a partner in several adjacent buckets, which differ by one or more years of age.

2.41 Assuming k is a constant, which it is in our implementation, the algorithm's execution time is linear with the number of agents being matched.

2.42 The pseudocode for the unadapted algorithm is provided in Algorithm 8.

Algorithm 8 Distribution counting pair-matching (DCPM)

Parameters: $Agents$**, an array of** $0..n-1$ **pointers to agents,**
$buckets$**, the number of buckets the agents can be distributed into**
k**, the number of agents in a bucket to search for a partner.**
There is also a domain-specific $getBucket$ **function which assigns an agent to its correct bucket in the range** $[0..buckets-1]$**.**

function DISTRIBUTIONCOUNTINGMATCH($Agents$, $buckets$, k)
  $shuffle(Agents)$
  $copyAgents \leftarrow Agents$
  Sort $copyAgents$ using sort by counting and the $getBucket$ function
  for $i = 0$ to $buckets - 1$ **do**
    $Table[i].start \leftarrow 0$
    $Table[i].entries \leftarrow 0$
  end for
  for each a in $Agents$ **do**
    $bucket \leftarrow getBucket(a)$
    $table[bucket].entries \leftarrow table[bucket].entries + 1$
  end for
  $lastIndex \leftarrow 0$
  for $i = 0$ to $buckets - 1$ **do**
    $table[i].start \leftarrow lastIndex$
    $lastIndex \leftarrow lastIndex + table[i].entries$
  end for
  for each unmatched agent a in $Agents$ **do**
    $bucket \leftarrow getBucket(a's\ desired\ partner)$
    $startIndex \leftarrow table[bucket].start$
    $endIndex \leftarrow startIndex + table[bucket].entries$
    $lastIndex \leftarrow min(startIndex + k, endIndex)$
    Find agent b that returns the smallest $distance(a, b)$ in $copyAgents[startIndex..lastIndex - 1]$
    Make a and b partners
    $swap(copyAgents[index\ of\ agent\ b], copyAgents[lastIndex - 1]$
    $table[bucket].entries \leftarrow table[bucket].entries - 1$
  end for
end function

Methods

Simulation experiments

General set-up

3.1 To test and analyze the algorithms, two different simulations have been programmed that use the structure depicted by Algorithm 1. The two models differ in the properties of the agents and the distance functions. While the first (ATTRACTREJECT) is somewhat abstract and intended to explore the relative effect of attractor and rejector attributes, the second (STIMOD) resembles a more concrete application in the field of epidemiology taken from research on the HIV-epidemic in South Africa (Eaton et al. 2012; Hontelez et al. 2013; Johnson & Geffen 2016).

3.2 The purpose of these simulations is solely to test and compare the algorithms; the simulations are not intended to model the natural world. They also differ from simulations of the natural world in that every agent on every iteration (or time step) is in the mating pool, so that comparisons of the algorithms remain as fair as possible.

3.3 Both experiments were coded in C++ (ISO C++11) and compiled with the GNU C++ compiler version 4.8.4. Processing of results was conducted using R. The experiments were executed on a machine with an Intel Xeon with 20 cores running at 2.3 GHZ and with 32GB RAM, under the GNU/Linux operating system. The source code is available on Github at `https://github.com/nathangeffen/pairmatchingalgorithms`.

ATTRACTREJECT experiment

3.4 In the first microsimulation, which we call *ATTRACTREJECT*, every agent has two properties: *attractor* and *rejector*. When executing the simulation, the user sets two constants, *ATTRACTOR_FACTOR* and *REJECTOR_FACTOR*, to values between 0 and 1, such that they add to 1. The distance function is simple and the pseudocode for it is provided in Algorithm 9.

Algorithm 9 Example of a distance function

Parameters a and b are agents with properties *attractor* and *rejector, both in the range* $[0, 1]$. *ATTRACTOR_FACTOR* and *REJECTOR_FACTOR* are user defined positive constants whose sum is 1.

1: **function** DISTANCE(a, b)
2: $\quad attraction \leftarrow ATTRACTOR\_FACTOR * |a.attractor - b.attractor|$
3: $\quad rejection = REJECTOR\_FACTOR * |a.rejector - (1 - b.rejector)|$
4: $\quad$ **return** $attraction + rejection$
5: **end function**

3.5 This distance function captures the issue of attractors and rejectors that affect the difficulty of the pair-matching problem when the triangle inequality is violated. The closer to 1 ATTRACTOR_FACTOR is set and the closer to 0 REJECTOR_FACTOR is set, the more clustered agents suitable for pairing will be. Conversely, the closer to 0 ATTRACTOR_FACTOR, the less useful clustering is, because agents that are clustered together will not be suitable matches.

STIMOD experiment

3.6 The *ATTRACTREJECT* model is not intuitive. The model in the second microsimulation (*STIMOD*), is intended to be more so. It is loosely based on previously published microsimulations (Johnson & Geffen 2016; Hontelez et al. 2013) that model the South African HIV epidemic, but much simpler than these. The pair matching characteristics of the agents are typical for a model of STIs: a list of previous partners, age, sex, sexual orientation and a variable that represents at how much risk of infection their sexual behaviour places them. We have also included an additional attractor factor in the distance function: location, described by x and y co-ordinates on a Euclidean plane, representing how far agents live from one another. An overview of the characteristics of the starting population can be found in Table 1.

3.7 We emphasise that the model and distance function used here are solely for comparing the algorithms. "Production" standard models of the natural world would need to be more complex.

3.8 As with the ATTRACTREJECT model there are user defined constants that specify the weighting of each of these factors in the final distance calculation. The pseudocode for this distant function is provided in Algorithm 10.

Variable	Range	Amount that distance function increases by (higher scores are poorer matches)
Age	Uniformly distributed over the ages 15 to 25	AGE_FACTOR x the difference in age
Sex	50% male/50% female	If mismatch on sexual orientation, add
Sexual orientation	90% of males are men who have sex with men, 90% of females are women who have sex with women	ORIENTATION_FACTOR x difference in (which is always 1 sexual orientation
Risk index	Uniformly distributed over [0-1]	RISK_FACTOR x the absolute difference between two agents
Location	Uniformly distributed over a Euclidean plane with co-ordinates (0-10,0-10)	0.1 x the Euclidean distance (DISTANCE_FACTOR)

Table 1: Agent attributes in STIMOD microsimulation

Algorithm 10 Distance function used in *STIMOD*

Parameters a and b are agents.
Upper case names are user defined constants.
The higher PREV_PARTNER_FACTOR the lower the probability of previous sexual partners rematching.
The higher AGE_FACTOR, ORIENTATION_FACTOR and RISK_FACTOR the lower the probability of agents of different age, incompatible sexual orientation and risk behaviour are to match.
The heterosexual property is 1 if the agent is heterosexual else 0.

function DISTANCE(a, b)
  if a and b have been partners before **then**
    $prev\_partner \leftarrow PREV\_PARTNER\_FACTOR$
  else
    $prev\_partner \leftarrow 0$
  end if
  $age\_diff \leftarrow AGE\_FACTOR * |a.date\_of\_birth - b.date\_of\_birth|$
  if $a.sex = b.sex$ **then**
    $sex\_diff \leftarrow ORIENTATION\_FACTOR * (a.heterosexual + b.heterosexual)$
  else
    $sex\_diff \leftarrow ORIENTATION\_FACTOR * ((1.0 - a.heterosexual) + (1.0 - b.heterosexual))$
  end if
  $risk\_diff \leftarrow RISK\_FACTOR * |a.riskiness - b.riskiness|$
  $distance\_diff = EuclideanDistance(a_x, a_y, b_x, b_y) * DISTANCE\_FACTOR$
  return $prev\_partner + age\_diff + sex\_diff + risk\_diff$
end function

3.9 In all tests, we set PREV_PARTNER_FACTOR to 500, ORIENTATION_FACTOR to 100, AGE_FACTOR to 1, RISK_FACTOR to 1, and DISTANCE_FACTOR to 0.1 (see Table 2). Agents with similar ages, riskiness and location are more likely to partner, so these are attractors. Since 90% of agents are set to heterosexual, having the same sex is usually a rejector. Two agents who have been partners are also very likely to reject each other.

Effectiveness and efficiency measures

Measuring speed

3.10 To compare the speed of the algorithms, we ran 10 simulations of the STIMOD model for each algorithm using 20,000 agents. Each simulation was executed for 20 iterations (i.e. time steps). In other words, every algorithm was executed 200 times.

3.11 We are also interested in how algorithms perform as the number of agents increases, or as the size of k increases for the algorithms that depend on this parameter. For the CSPM algorithm we are also interested in how the number of clusters affects speed. Since this algorithm achieved the best balance of effectiveness and speed in our tests, we ran tests on it varying the number of agents up to 10 million. We also ran tests varying k from 50 to 500, and tests varying c from 50 to 500.

Measuring effectiveness

3.12 We ran the following tests to compare the effectiveness of the algorithms in the STIMOD model:

- STIMOD with 5,000 agents, simulated 36 times with different random number seeds, with each simulation having 20 iterations. In other words every algorithm executes 720 times. Algorithms were compared against Blossom V for effectiveness.
- STIMOD with 20,000 agents, simulated 36 times, with each simulation having 20 iterations.
- Using CSPM with 20,000 agents, we varied k, the number of adjacent agents to consider, from 50 to 1,000 stepping up by 50 on each execution of the microsimulation, while holding the number of clusters constant, to see the effect on the mean ranking and distance. We then similarly varied the number of clusters, c, initially setting it to 1 and then from 50 to 1,000 stepping up by 50 on each execution of the microsimulation, while holding k constant at 200.

3.13 We ran the following tests to compare the effectiveness of the algorithms in the ATTRACTREJECT model:

- ATTRACTREJECT microsimulation with 5,000 agents, with each simulation executed 20 times with different random seeds with the following combinations of ATTRACTOR_FACTOR and REJECTOR_FACTOR respectively:
  - 0.0 and 1.0
  - 0.25 and 0.75
  - 0.5 and 0.5
  - 0.75 and 0.25
  - 1.0 and 0.0
- ATTRACTREJECT microsimulation with 20,000 agents, with each simulation executed 20 times with the above combinations of ATTRACTOR_FACTOR and REJECTOR_FACTOR respectively.

Best-possible result

3.14 As mentioned above, the *mean distance* and *mean ranking* measures can only be used for the relative comparison of the different algorithms. Furthermore, the set of partnerships with the minimum distance possible is not known a priori. This can be calculated using the Blossom V algorithm. A prerequisite for the application of Blossom V is the creation of a fully connected undirected graph where the vertices represent the agents and the edges represent the distances between them, a process with a quadratic speed increase with the number of agents being matched. The Blossom V algorithm itself is not stochastic and is very slow, with speed worse than cubic with the number of agents being matched. It is therefore only used for reference purposes in the simulations of 5,000 agents.

Mean of mean distances

3.15 In the first measure we calculate mean distance of all the pairings for a single execution of the pairing algorithm. Since every algorithm is executed multiple times, we calculate the mean of these means. The measure of effectiveness for an algorithm is then the mean of its mean distances divided by the mean of the mean distances calculated by Blossom V.

3.16 However, when we use 20,000 agents Blossom V is too slow to use. On state of the art consumer hardware (an Intel Xeon running 20 i7 cores) a single execution of Blossom V over 20,000 agents, including creating the graph, takes over 2 hours — and we have had to run hundreds of simulations. Instead, effectiveness of the algorithm with the lowest mean of mean distances (usually BFPM) is assigned the value of 1. The effectiveness of the remaining algorithms is the ratio of their mean of mean distances to this algorithm's mean of mean distances.

Mean of mean rankings

3.17 In the second measure for every agent we calculate the distance to every other agent, creating a fully connected undirected graph. For each agent, a, we can then order every other agent by its suitability as a partner to a. The ranking of a pairing for a given agent is its partner's place in this ordering. If an agent chooses its ideal partner the rank is 0. If it chooses the least desirable partner, the rank is $n - 2$, where n is the number of agents being matched. We can then calculate the mean ranking over all the agents being matched for a single execution of an algorithm. We compare the algorithms by calculating the mean of its mean rankings over many executions. The lower the mean of mean ranking the better the algorithm has done.

3.18 Effectiveness is calculated analogously to the way it is done for the mean of mean distances measurement. Likewise with 5,000 agents effectiveness is established with Blossom V (although Blossom V guarantees lowest mean distance and not lowest mean ranking, in practice it generally returns the lowest ranking). With 20,000 agents, the method of assigning 1 to the effectiveness of the best algorithm is used.

Results

Speed tests

4.1 Table 2a lists the results of the speed tests. The fastest algorithm, RPM, has an average speed more than 5,700 times faster than the slowest algorithm, BFPM. The former completes the entire pair-matching event (i.e. match every agent for an iteration) in less than a millisecond on average, while the latter takes over three seconds on average. CSPM, RKPM, WSPM and DCPM have a mean speed between 27 and 39 milliseconds, and are all approximately an order of two magnitudes faster than BFPM.

Algorithm	Mean speed (ms)	Speedup (BFPM ref)
RPM	0.6	5,715
RKPM	27	115
CSPM	29	109
WSPM	29	109
DCPM	39	81
BFPM	3,143	1

(a) Comparison of algorithms by speed. Each algorithm was executed 20 times per simulation, and each simulation was run 10 times. Each simulation consisted of 20,000 agents. For algorithms that take a k parameter, k was set to 200. For CSPM, c was set to 100.

Number of agents	Mean speed (ms)
20,000	29
50,000	53
100,000	109
500,000	608
1,000,000	1,263
5,000,000	6,842

(b) Speed of CSPM with increase in agents. The algorithm was executed once per simulation, and each simulation was run 10 times. The k parameter was set to 200, and c was set to 100.

k	Mean speed (ms)
50	301
100	410
200	606
300	808
400	1,006
500	1,230

(c) Change of speed of CSPM as k increases. The algorithm was executed once per simulation, and each simulation was run 10 times. 500,000 agents were used.

Table 2: Speed comparisons of algorithms

	Distance				Rank				Speed
Algorithm	Mean	SD	Median [IQR]	Effectiv.	Mean	SD	Median [IQR]	Effect.	Time
Blossom V	2.7	14.3	0.32 [0.27;0.41]	1.00	40	225	1.7 [0;5.4]	1.00	6.5 minutes
BFPM	2.8	19.2	0.32 [0.26;0.44]	1.03	63	328	1.6 [0.0;6.7]	1.6	345 ms
CSPM	3.3	17.1	0.46 [0.34;0.84]	1.21	138	507	15.6 [3.1;74.9]	3.5	15 ms
WSPM	3.4	18.3	0.79 [0.48;1.38]	1.25	180	361	64.7 [20.1;187.6]	4.6	14 ms
RKPM	3.4	18.3	0.79 [0.48;1.39]	1.25	181	362	65.5 [20.4;188.7]	4.6	14 ms
DCPM	3.8	20.5	0.57 [0.37;0.98]	1.40	125	347	29.7 [9.4;91.2]	3.2	17 ms
RPM	105.6	93.1	104.17 [4.99;203.29]	38.6	2501	1444	2503.4 [1251.7;3752.2]	63	0.3 ms

Table 3: Distance and rank results on 5,000 agents for STIMOD. The mean, standard deviation, median, IQR and time columns are the means of these values over all executions. The effectiveness is the ratio of the algorithm's mean distance to Blossom V's.

4.2 As the number of iterations increases, thereby creating more previous partners, the mean speed of the algorithms (except RPM) increases slightly, because the distance calculation has to search for more previous partners. For example BFPM increased from a mean of over 2 seconds on the first iteration of a simulation to 4.6 seconds on the 20th, and CSPM increased from 19 to 43 milliseconds.

4.3 The speed of the CSPM algorithm increased linearly with the number of agents, as Table 2b indicates. This is expected because even though our analysis above shows that its speed increases linearithmically with the number of agents being matched, this is due to its sorting operation, which is a small contributor to the overall time of the algorithm with these small numbers of agents.

4.4 Modifying the number of clusters also shows a linear effect on the speed of the algorithm. This is expected because modifying the value of c in Algorithm 7 increases the number of iterations of the for loop at line 8 although it correspondingly decreases the number of agents to be shuffled at line 11.

4.5 The speed of the CSPM algorithm also increased linearly with the size of k, as Table 2c indicates. This too is expected as increasing k proportionately increases the number of comparisons made for each agent a with the following agents in the array.

Effectiveness tests

4.6 Table 3 shows the results of the comparison of the algorithms on the STIMOD microsimulation for 5,000 agents. With this few agents, it is still feasible to identify the theoretically lowest mean distance using the Blossom V algorithm. In these tests CSPM, followed by DCPM algorithm, offers the best trade-off of speed and effectiveness. Nevertheless RKPM and WSPM have similar effectiveness and speeds.

4.7 Table 4 shows the distance and ranking results respectively of the ATTRACTREJECT microsimulation. In contrast to the STIMOD microsimulation the DCPM algorithm is entirely unsuited to this distance function and performs poorly, although much better than RPM. The remaining algorithms improve their relative effectiveness as ATTRACTOR_FACTOR rises to 1 and REJECTOR_FACTOR declines to 0.

4.8 Interestingly CSPM exceeds the effectiveness of BFPM for higher values of ATTRACTOR_FACTOR. A possible explanation for this is that agents at the front of the array processed by BFPM will be well-matched, but agents nearer the back of the array have fewer partner options that give low distance measurements. CSPM, on the other hand, by first clustering, ensures the agents at the back of the array are nearer likely partners.

4.9 This raises a key weakness of BFPM, CSPM, WSPM, RKPM and possibly even DCPM: the distribution of their matches varies greatly in quality between the front and back of their arrays, and there is room for further research as to how they can be adapted to have a more equal distribution of quality.

4.10 Consider Table 6 which shows the mean median ranking, mean interquartile range (IQR) and mean standard deviation across the ATTRACTREJECT simulations with ATTRACTOR_FACTOR set to 0.5. The median ranking for BFPM is much lower than the mean ranking shown in Table 4. It is even better than the median ranking of Blossom V. In fact at least 75% of the rankings are better than the mean ranking, implying that some very poor rankings toward the back of the agent array bring down the mean ranking. This skewed quality is also a problem for CSPM but, as shown by its smaller standard deviation and median ranking closer to its mean ranking, not as profoundly as it is for BFPM, WSPM, RKPM and DCPM (see also Figure 2).

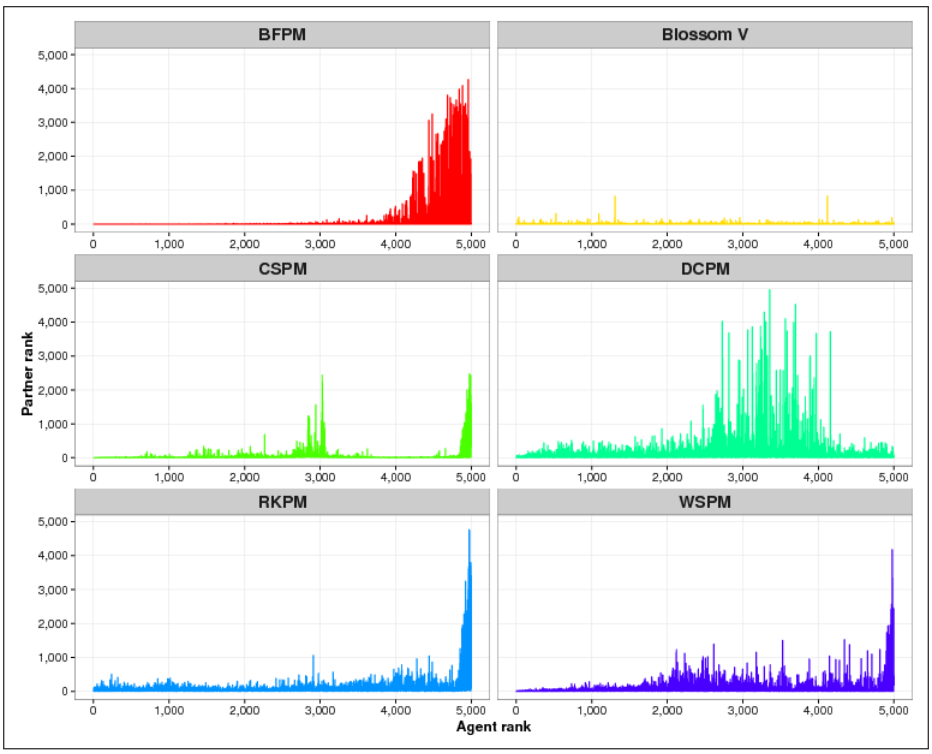

Figure 2: Distribution of rankings for single execution of each algorithm
This figure shows the partner rank for a single run of each algorithm (except RPM). Notice how Blossom V's poorly ranked agents are distributed uniformly across the array, but the poorly ranked agents for the other algorithms are distributed towards the back of the array.

	Attractor:	0	0.25	0.5	0.75	1
	Rejector:	1	0.75	0.5	0.25	0
	Algorithm	Mean (ratio to best)				
Distance	Blossom V	0.006 (1)	0.009 (1)	0.010 (1)	0.008 (1)	0.0002 (1)
	BFPM	0.008 (1.3)	0.015 (1.6)	0.016 (1.6)	0.012 (1.6)	0.0007 (3.5)
	CSPM	0.016 (2.6)	0.018 (1.9)	0.018 (1.8)	0.015 (1.9)	0.0004 (2.0)
	WSPM	0.016 (2.6)	0.032 (3.3)	0.033 (3.4)	0.026 (3.3)	0.0025 (12.2)
	RKPM	0.015 (2.5)	0.036 (3.8)	0.037 (3.9)	0.030 (3.8)	0.0032 (15.8)
	DCPM	0.018 (2.9)	0.041 (4.3)	0.044 (4.6)	0.039 (4.9)	0.0135 (65.9)
	RPM	0.334 (53.6)	0.334 (35.3)	0.334 (34.3)	0.334 (42.3)	0.3337 (1,633)
Rank	Blossom V	61 (1)	8 (1)	5 (1)	4 (1)	1 (1)
	BFPM	63 (1.1)	47 (6.1)	37 (7)	25 (5.8)	5 (5.4)
	CSPM	148 (2.4)	39 (5.1)	22 (4)	15 (3.5)	3 (2.9)
	WSPM	147 (2.4)	82 (10.5)	64 (12)	52 (12.2)	23 (23.3)
	RKPM	139 (2.3)	93 (12)	76 (14)	60 (14.2)	30 (30.5)
	DCPM	161 (2.7)	129 (16.5)	112 (21)	112 (26.4)	127 (127.4)
	RPM	2501 (41)	2502 (321)	2502 (472)	2501 (588)	2499 (2499)

Table 4: Distance and rank results for 5,000 agents for ATTRACTREJECT. Ranks and ratios are rounded.

4.11 Table 4 also shows that when REJECTOR_FACTOR is much larger than ATTRACTOR_FACTOR, then as expected RKPM is as effective as CSPM, DCPM and WSPM.

4.12 When the number of agents in the STIMOD model is increased to 20,000 we do not see any substantial changes in effectiveness across the algorithms compared to when 5,000 agents are used, except that CSPM clearly outper-

	Distance				Rank				Speed
Algorithm	Mean	SD	Median [IQR]	Effect.	Mean	SD	Median [IQR]	Effect.	Time
BFPM	1.7	14.9	0.2 [0.2;0.3]	1	130	971	1.5 [0;6]	1	7 seconds
CSPM	2.0	15.3	0.3 [0.2;0.4]	1.2	276	1,315	14.6 [3.6;62.7]	2.1	51 ms
DCPM	2.7	15.2	0.6 [0.4;1.1]	1.6	520	1,207	141 [44.6;448.9]	4	61 ms
RKPM	2.8	13.6	0.8 [0.5;1.5]	1.7	865	1,504	312 [96.3;910]	6.7	47 ms
WSPM	2.8	13.7	0.8 [0.5;1.5]	1.7	869	1,508	312.6 [95.6;916.9]	6.7	51 ms
RPM	104.7	91.3	104.08 [5;203]	63	10,000	5,773	10,001 [5,002;14,999]	77	2 ms

Table 5: Distance and rank results for 20,000 agents for STIMOD.

Algorithm	ATTRACTREJECT		STIMOD			
	Median rank [IQR]	SD	Rankings			
			Mean	Best	Worst	SD
Blossom V	1.4 [0;5.1]	17	–	–	–	–
BFPM	1 [0;5.3]	225	130	43	254	51
CSPM	7.4 [2.7;16.1]	14	276	66	537	107
WSPM	22.9 [7.2;62.3]	165	869	354	1684	276
RKPM	36 [13.8;80]	190	865	345	1673	275
DCPM	44 [15.5;112]	250	520	283	957	113
RPM	2,502 [1,259;3,746]	1,440	10,000	9,797	10,192	56

Table 6: Selected statistics for 5,000 agents for ATTRACTREJECT with attractor and rejector set to 0.5, and volatility of the results as demonstrated for STIMOD with 20,000 agents

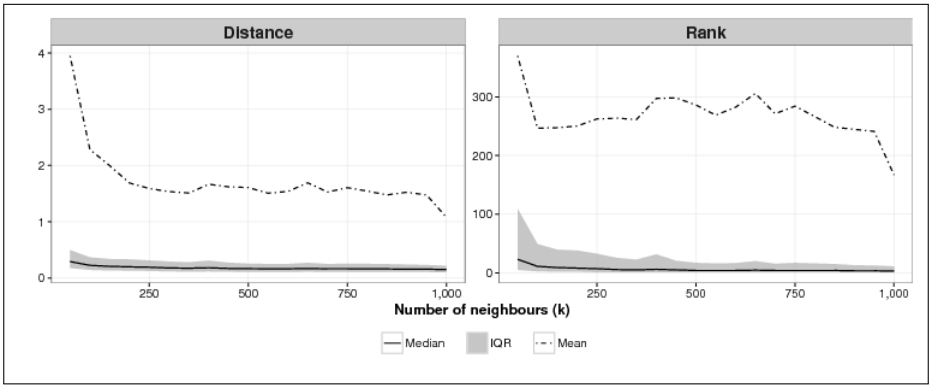

Figure 3: Finding the ideal number of neighbours to minimize distance and rank (displayed on the respective y-axes).

forms DCPM on both mean distance and mean ranking. Whether this is due to something qualitatively different happening as the number of agents increases or due to random fluctuations of the results is unclear. The results from simulation to simulation are indeed volatile as Table 6 shows, with BFPM being the most stable algorithm followed by CSPM.

4.13 We also attempted to find the ideal values of k, the number of neighbours, and c, the number of clusters for the CSPM algorithm. As Figure 3 shows, the relationship between the value of k and effectiveness is unclear for this application. For $k = 50$, the lowest value of k we tried, the mean of the mean rankings over 40 simulations was 370. For $k = 1,000$, the highest value of k we used, the mean ranking was 167, the lowest, but as the graph shows for all values in between, there is no discernible pattern. On other test runs we got different results where $k = 1000$ was not the best and $k = 50$ was not the worst. Results are domain dependent, and modellers who use CSPM will have to experiment to find the best value of k.

4.14 We are unsure what the relationship between c and effectiveness is as Figure 4 shows. For $c = 1$ (essentially an array sorted on the cluster function), the mean of the mean rankings is poor: 791. For all values of c tested from 50 to 1,000 we were unable to identify a discernable difference in effectiveness across the simulations. As with

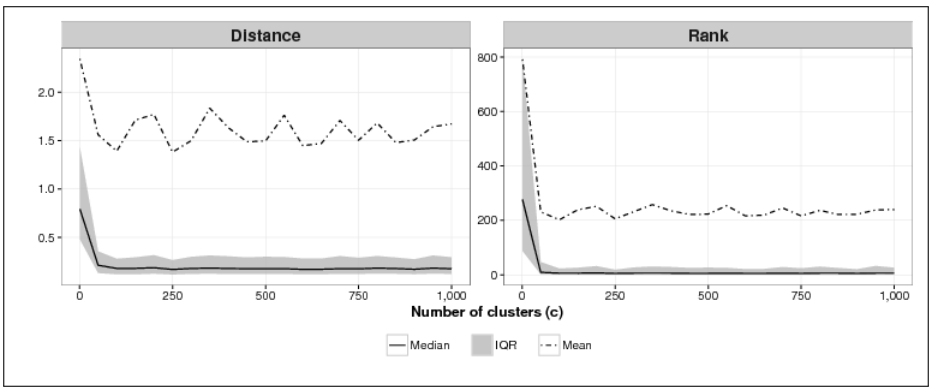

Figure 4: Finding the ideal number of clusters to minimize distance and rank (displayed on the respective y-axes).

the ideal value of k modellers who use CSPM will have to experiment to find the best value of c.

Discussion

5.1 We have used two microsimulations to show how four algorithms CSPM, DCPM, WSPM and RKPM can match agents much better than chance (RPM) yet much faster than the BFPM algorithm and Blossom V — which finds a perfect set of pairs without desired stochasticism.

5.2 CSPM performed consistently well across our tests. For example in the STIMOD microsimulation it only had a mean ranking four times worse than Blossom V, whereas random matching was more than 61 times worse. When the number of agents was increased to 20,000, a number beyond which it is practical, at least on our hardware, to check effectiveness against Blossom V, CSPM's mean ranking was only half as poor as that of BFPM, more than double that of the next best algorithm DCPM and 45 times better than random matching. Yet it executed on average in just over 50 milliseconds versus 7 seconds for BFPM.

5.3 However, while CSPM has done well here, it is likely that there are other applications where the other fast approximation algorithms, DCPM, WSPM and RKPM, will be equally good or better choices. We are currently exploring an application where DCPM appears to be better. In this application, the agents fall neatly into a prespecified set of buckets, which is ideal for DCPM.

5.4 As the tests with the ATTRACTREJECT microsimulation show, when rejector attributes of agents dominate, the RKPM algorithm might as well be used: it is faster than CSPM, trivial to implement and will produce results that are at least as good as CSPM because clustering achieves nothing in such pair-matching scenarios.

5.5 CSPM requires the user to decide the values of two parameters, k and c. We cannot currently offer good heuristics for choosing these values other than to suggest preliminary empirical tests for the specific microsimulation being used.

5.6 We have not yet found a situation where WSPM outperforms all of CSPM, DCPM and RKPM. It is possible that it is simply an inferior algorithm, but it is not inconceivable that a useful microsimulation application exists in which WSPM makes sense as the pair-matching algorithm.

5.7 One serious limitation of BFPM, CSPM, RKPM and WSPM, at least in the applications considered here, is that the quality of their matches are skewed to the front of the array of agents that they process. After shuffling, agents near the front of the array will tend to be matched with partners with smaller distances than the agents at the back of the array. This is reflected in the fact that the mean of the median rankings — and even the 75% quartile — is smaller than the mean of the mean rankings in the simulations where we examined this. Further research is required to improve one or more of CSPM, RKPM or WSPM to reduce this skewed matching.

5.8 While we are interested in microsimulations that simulate sexual pairing, it is possible there are uses of these algorithms in other microsimulations which require agent interaction. However, for such applications the algorithms would need to be evaluated to see if they produce results that compare well to observed data.

5.9 Further research is also needed to see how useful an algorithm such as CSPM is in the context of modelling sexually transmitted infections. For example, the most cited model of the South African HIV epidemic is a deterministic equation-based one that implicitly assumes people are randomly matched (Granich et al. 2009). Other

deterministic models of the epidemic make more complex assumptions, e.g. by dividing the population into a small number of sexual groups (Johnson 2014). It would be interesting to compare the outputs of a full-fledged microsimulation of an STI epidemic using random matching versus CSPM. If the results of such a microsimulation are similar irrespective of the pair-matching algorithm then the speed of random matching means it is the better choice. If however, the results are vastly different, it has implications for how we model, not only using agents, but also with deterministic differential-equation models.

References

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R. & Wu, A. Y. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, *45*(6), 891–923

Beis, J. S. & Lowe, D. G. (1997). Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *CVPR '97 Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*

Bouffard, N., Easther, R., Johnson, T., Morrison, R. J. & Vink, J. (2001). Matchmaker, matchmaker, make me a match. *Brazilian Electronic Journal of Economics*, *4*(2)

Cook, W. & Rohe, A. (1999). Computing Minimum-Weight Perfect Matchings. *INFORMS Journal on Computing*, *11*(2), 138–148

Deissenberg, C., van der Hoog, S. & Dawid, H. (2008). EURACE: A massively parallel agent-based model of the European economy. *Applied Mathematics and Computation*, *204*(2), 541–552

Eaton, J. W., Johnson, L. F., Salomon, J. A., Bärnighausen, T., Bendavid, E., Bershteyn, A., Bloom, D. E., Cambiano, V., Fraser, C., Hontelez, J. A. C., Humair, S., Klein, D. J., Long, E. F., Phillips, A. N., Pretorius, C., Stover, J., Wenger, E. A., Williams, B. G. & Hallett, T. B. (2012). HIV treatment as prevention: Systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. *PLoS medicine*, *9*(7), e1001245

Granich, R. M., Gilks, C. F., Dye, C., De Cock, K. M. & Williams, B. G. (2009). Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: A mathematical model. *The Lancet*, *373*(9657), 48–57

Gras, R., Devaurs, D., Wozniak, A. & Aspinall, A. (2009). An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial Life*, *15*(4), 423–463

Gray, R. T., Hoare, A., McCann, P. D., Bradley, J., Down, I., Donovan, B., Prestage, G. & Wilson, D. P. (2011). Will changes in gay men's sexual behavior reduce syphilis rates? *Sexually transmitted diseases*, *38*(12), 1151–1158

Hontelez, J. A. C., Lurie, M. N., Bärnighausen, T., Bakker, R., Baltussen, R., Tanser, F., Hallett, T. B., Newell, M.-L. & de Vlas, S. J. (2013). Elimination of HIV in South Africa through expanded access to antiretroviral therapy: A model comparison study. *PLoS medicine*, *10*(10), e1001534

Indyk, P. & Motwani, R. (1998). Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, (pp. 604–613). New York, NY, USA: ACM

Johnson, L. (2014). *THEMBISA version 1.0: A model for evaluating the impact of HIV/AIDS in South Africa*. Centre for Infectious Disease Epidemiology and Research working paper

Johnson, L. F. & Geffen, N. (2016). A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. *Sexually Transmitted Diseases*, *43*(3), 139–146

Knuth, D. E. (1997). *The Art of Computer Programming, Volume 2: (2Nd Ed.) Seminumerical Algorithms*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.

Knuth, D. E. (1998). *The Art of Computer Programming, Volume 3: (2Nd Ed.) Sorting and Searching*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.

Kolmogorov, V. (2009). Blossom V: A new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation*, *1*(1), 43–67

Larose, D. T. & Larose, C. D. (2014). *k-Nearest Neighbor Algorithm*, (pp. 149–164). John Wiley & Sons, Inc.

Levitin, A. (2011). *Introduction to the Design and Analysis of Algorithms*. Boston: Pearson, 3 edition edn.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*(1), 143–166

Pelillo, M. (2014). Alhazen and the nearest neighbor rule. *Pattern Recognition Letters*, *38*, 34 – 37

Scholz, S., Surmann, B., Elkenkamp, S., Batram, M. & Greiner, W. (2016). Creating populations with partnerships for large-scale agent-based models – A comparison of methods. In *Proceedings of the Summer Computer Simulation Conference (SCSC 2016)*, Simulation Series, (pp. 351–356). Red Hook, NY, USA: Curran Associates, Inc.

Weisstein, E. W. (2016). *Metric Space*. From MathWorld – A Wolfram Web Resource. `http://mathworld.wolfram.com/MetricSpace.html`. Accessed: 16 November 2016

Zinn, S. (2012). A Mate-Matching Algorithm for Continuous-Time Microsimulation Models. *International Journal of Microsimulation*, *5*(1), 31–51

Chapter 6

Comparison of pair-matching algorithms in a model

6.1 Foreword

The following article, *The influence of design decisions on incidence in microsimulations of sexually transmitted infections* (currently under peer review), extends the work of the previous chapter by comparing the pair-matching algorithms in a microsimulation of a fictitious STI.

We analysed how incidence changed when different pair-matching algorithms were used, as well as the effect of the algorithms on the speed of the microsimulation.

We also varied various aspects of the model, such as population size, risk of infection, and heterogeneity of sexual behaviour, to see how these in conjunction with the pair-matching algorithms affected incidence. Surprisingly, we found that for pair-matching algorithms that account for heterogeneity,

incidence decreased with a rise in population size. The more infectious the STI the more profound this effect.

This is counter-intuitive and also somewhat concerning. If this finding is not simply an artefact of our methodology, i.e. it is an effect present in the natural world, it suggests that agent-based epidemic models may behave differently under different population sizes. [1] Ideally then, the population size should be set to the size of the population being studied. However if that results in a computationally-infeasible model, sensitivity analysis should be conducted to investigate the impact of the assumed population size on the conclusions.

This finding was the impetus for an afterword to this chapter that considers the question: Does STI modelling increase our knowledge?

[1] It has been brought to the author's attention that it is commonly known by most modellers who work with stochastic agent-based models that results differ with different population sizes.

How various design decisions on matching individuals in relationships affects the outcomes of microsimulations of sexually transmitted infection epidemics

Nathan Geffen
Department of Computer Science, University of Cape Town
Stefan Scholz
Department of Public Health, University of Bielefeld

March 4, 2018

Abstract

Objective: Microsimulations are increasingly used to estimate the prevalence of sexually transmitted infections (STIs). These models consist of agents that represent a sexually active population. Matching agents into sexual relationships is computationally intensive and presents modellers with difficult design decisions: how to select which partnerships between agents break up, which agents enter a mating pool, and how to pair agents in the mating pool. The aim of this study was to analyse the effect of these design decision on STI prevalence.

Methods: We compared two strategies for selecting which agents enter a daily mating pool and which agent partnerships break up: random selection in which agents are treated homogenously versus selection based on data from a large German longitudinal data set that accounts for sex, sexual orientation and age heterogeneity. We also coupled each of these strategies with one of several recently described algorithms for pairing agents and compared their speed and outcomes. Additional design considerations were also considered, such as the number of agents used in the model, increasing heterogeneity of agents'

sexual behaviour, and the proportion of relationships that are casual sex encounters.

Results: Approaches that account for agent heterogeneity estimated lower prevalence than less sophisticated approaches that treat agents homogeneously. Also, in simulations that accounted for heterogeneous pairing of agents, as the risk of infection increased, incidence declined as the number of agents increased. Our algorithms facilitate the execution of thousands of simulations with large numbers of agents quickly.

Conclusion: Fast pair-matching algorithms provide a practical way for microsimulation modellers to account for sexual behaviour heterogeneity. For STIs with high infection rates modellers may need to experiment with different population sizes.

Introduction

Microsimulations or agent-based models (ABMs) are increasingly used to simulate incidence and prevalence of STIs as well as to identify the costs and benefits of strategies to contain them. Diseases modelled using this approach include HIV [4, 7, 13], syphilis [12], gonorrhoea [9], HPV [11], herpes, chlamydia and trichomoniasis [16]. The popularity of such stochastic microsimulations may lie in the easier implementation of complex, heterogeneous sexual behaviour when compared to traditional equation-based (compartmental) models.

However, there are many design decisions that must be made for microsimulations, with respect to sexual behaviour modelling. For example, which agents should be considered for relationships (i.e. placed in a mating pool), how agents in the mating pool should be paired with one another, and which relationships should terminate (break up). These decisions may be informed by agent characteristics, such as age, sex and sexual orientation, but also by individual variations from the average behaviour, e.g. sexual risk taking and propensity to remain in relationships. [1]

It is well understood how the design decisions for differential equation models affect outcomes [19], including for specific diseases [15], but much

[1] This has similarities with regression analysis for longitudinal data: Some variation of the dependent variable can be explained by observable, group-level characteristics like age, sex, and sexual orientation., but adding a random effect for unobservable, individual characteristics may explain additional variation.

less so for STI microsimulations. The aim of our research is to fill this gap and to understand how the above-mentioned design decisions affect estimates of disease incidence and prevalence. Specifically we want to analyse the effect of different algorithms for matching and unmatching agents and their interaction with (I) the probability of transmission, (II) the size of the model population and (III) heterogeneity in agent behavior.

To do this we built a microsimulation model with behaviour based on data drawn from the German population, and that simulates the spread of a generic, fictitious STI in a fixed cohort. To isolate the effects of different approaches to partner matching and breakups, and to ensure that prevalence is always cumulative incidence, the STI has no healing rate, births, deaths or migration. The prevalence estimated at the end of our simulations is a function of the risk of infection in serodiscordant partnerships, the number of partnerships over time, and the distribution of partnerships over time. The last of these is especially affected by how the mating pool is chosen, the algorithm that pairs agents in the mating pool, and how breakups are modelled. Moreover the number and distribution of partnerships can also be affected by the likelihood of casual sex encounters (modelled as partnerships lasting one day in our simulations), and the heterogeneity of agent sexual behaviour.[2]

The microsimulation had to be fast and capable of handling large numbers of agents in order to carry out this research. Appendix 1 therefore discusses the implementation details of the microsimulation.

[2]Note on terminology: We use *microsimulation* and *agent based model* synonymously. A single execution of a microsimulation model, usually but not always 10 years in our experiments, is called a *simulation* or *run*. A *time-step* is a single iteration of a simulation, which happens to always be one day in the experiments described here. A *mating pool* is a subset of agents on each day that must be *paired* into relationships. We use *partnership*, *relationship* and *pair* synonymously. The terms *breakup* and *unmatch* describe the termination of a relationship.

Materials and methods

General structure of the model

We implemented a discrete time-step microsimulation which has the structure of Algorithm 1. [3]

Users can specify the number of agents, the time period for which to run the simulation, and the daily risk of infection for a seronegative agent in a serodiscordant partnership. This risk is specified based on the sex of the uninfected agent and whether it is an opposite- or same-sex relationship. The time step of the model is one day in all the experiments described in this paper. On each day a series of events is executed on all or a subset of agents. These events, executed in the following order, are:

Age Each agent becomes a day older.

Infect Each uninfected agent in a relationship with an infected one may become infected.

Breakup All the relationships are traversed and some of them are terminated.

Select A subset of unpaired agents is selected to enter a mating pool. If there's an odd number of agents in the mating pool, a randomly chosen one is removed.

Match Agents in the mating pool are matched with each other.

An exception to this is when the model is running in *stabilisation mode*. This is sometimes done at the beginning of a simulation to stabilise the number of daily breakups and pairings. During this phase neither the *age nor* infect events are executed. [4]

[3] The code is available under the GNU General Public License Version 3. It is written in C++ and available on Github at: `https://github.com/nathangeffen/faststi`.

[4] It is often hard in microsimulations to start the model with roughly the same number of breakups and pairings in each time step. The initial number of relationships in the model may be unknown or may need to be determined empirically. So, for example, a simulation may be initialised to have no pairs, or too few pairs, to start with. During the stabilisation period the number of pairings increases on each time step, until, hopefully, the simulation reaches a stage where the pairings and breakups roughly match on each time step.

Algorithm 1 Structure of a discrete microsimulation from [10]

```
for each time step do
    for each event e do
        for each agent a do
            if e has to be applied to a then
                Apply e to a
            end if
        end for
    end for
end for
```

Each agent has a sex, sexual orientation and age which can be used to identify the corresponding characteristics of its preferred partner. An agent's sex and age determines its daily risk of entering the mating pool if it is single, or breaking up if it is in a relationship.

Population characteristics and agent behavior

All singles agents were initialized to between 12 and 50 years old, proportionate to the German population. The vast majority of agents in partnerships were initialised to between 12 and 50. [5]

The behavior of the agents includes probabilities for breakups, entering the mating pool for long-term relationships and casual sex encounters. Casual sex encounters last one time step, which corresponds to one day. Only agents not currently in a relationship can enter a casual relationship, a limitation of our model. To model heterogeneous versus homogeneous behavior, two different strategies, called RANDOM and DATA, were implemented.

- The RANDOM strategy contains no heterogeneous behavior as it randomly selects a set of partnerships to break up and a set of agents to enter the mating pool. On average the same number of agents break up that enter the mating pool each day.
- The DATA strategy estimates daily, group-level (i.e., age- and sex-specific) probabilities for breakups, partnership formation and casual

[5] The initialization routine sometimes creates agents outside this age range to be partners of agents in the 12 to 50 age range.

sex contacts from *The German Family Panel* (PAIRFAM) [22, 14]. This longitudinal dataset provides information about the complete relationship history of the study participants before the beginning of the survey as well as during the survey period. For the latter, information about the frequency of sexual intercourse is provided for the three months preceeding the interview date of each wave of the survey. To determine the probabilities of breakups and entering a new long-term relationship, the beginnings and ends of partnerships were extracted for each study participant for all the relationships and these data were summarised for all study participants by each age year. The probabilities were then calculated by dividing the number of breakups by the number of relationships at each age and the number of new relationships by the number of single persons at each age. To estimate the probability of casual sex contacts the number of sexual intercourses in the three months before the survey were converted to daily probabilities for people who indicated that they had not been in a relationship during that period. When calculating the risk of a breakup, the sum of the agents' breakup probabilities are averaged.

As men-who-have-sex-with-men (MSM) or women-who-have-sex-with-women (WSW) are not represented well in the data set, the probabilities have been estimated irrespective of sexual orientation. The estimates of the age-specific model parameters can be found in Appendix 2. Agents older than the highest age — 50 years — for which data is available, are treated the same as 50-year-olds.

By default the DATA strategy models group-level heterogeneity: differences in sexual behaviour by sex and age. The implementation is as follows: Consider how an agent enters the mating pool (i.e. the Select event). A uniform random number between 0 and 1 is generated. If the agent's probability, p, of entering the mating pool, calculated based on sex and age, is less than this number, the agent enters the mating pool. Analogous mechanisms are used for determining whether the agent will have a casual relationship, and whether a relationship breaks up (in this case the mean of the probabilities of the two agents is used).

To additionally model unobservable, individual-level heterogeneity, p is multiplied by a factor with normal distribution of mean 1 and standard deviation 0.3. This factor is set individually for each agent at the beginning of the simulation and does not change over time. There are actually three such

factors: one for entering the mating pool in search of a casual sex relationship, one for entering the mating pool for a non-casual sex relationship, and one for breaking up, thereby generating a wide variety of individual behaviour. However, except for one set of experiments, this individual-level heterogeneity is deactivated.

While some of the parameters chosen to do these simulations are arbitrary (such as the standard deviation of 0.3 in the above paragraph), our aim is proof of concept. Modelling specific diseases will require choosing appropriate parameters, ideally informed by data.

Matching procedures and algorithms

By default the simulation matches all agents in the mating pool. [6] The matching algorithms have been described and their performances analysed in [10]. They depend on the existence of a distance function that measures the suitability of two agents for matching based on sex, sexual orientation and age according to the distribution of relationships in the population. The smaller the distance the more suitable are the agents for matching. All but one of the algorithms attempt to minimise the sum of the distances of all matches.

Using distance to measure the suitability of a match has advantages: (1) the algorithms can be kept generic with domain specific details confined to the distance function, (2) category mismatches (such as agents with different sexual orientations) can be dealt with by the distance function returning very large values rendering such matches unlikely or even impossible if the penalization is higher than a maximal distance threshold for matching, and (3) it provides a measurement for comparing how closely algorithms estimate the underlying distribution. Algorithm 2 provides the distance function we used.

Two algorithms serve as upper and lower boundaries of the quality of matches. Random-pair matching (RPM), in which agents in the mating pool are paired randomly, sets the lower limit on quality. The average distance between paired agents in the mating pool that the other algorithms generate should be much smaller than that of RPM.

On the other end of the scale, the Blossom algorithm, first described by [8], finds the minimum sum of the distances of pairs of vertices in a graph.

[6] If the number of agents in the market is odd, one is randomly removed.

Algorithm 2 Distance function used in simulations

Parameters:
a, b**: agents between whom to measure distance**
Return value: real number that determines how likely a partnership is between a **and** b **according to the distribution of partnerships in the population being studied.**
0 is a good match, while 50 or more is a poor one.
A modification in some simulations we ran was to remove the previous partnership penalty.

function DISTANCE(a, b)
    $ageProb \leftarrow$ lookup probability of matching a with b based on ages.
    $agePenalty \leftarrow (1 - ageProb) * 50$
    if mismatch on sex based on sexual orientation **then**
        $orientationPenalty \leftarrow 50$
    end if
    if a and b have been partners previously **then**
        $prevPenalty \leftarrow 50$
    end if
    Return $agePenalty + orientationPenalty + PrevPenalty$
end function

We use it by first generating a fully connected undirected graph in which each vertex represents an agent in the mating pool and each edge represents the distance between two agents. The Blossom algorithm finds the theoretically closest set of pairs to the distribution of relationships in the population being studied within the constraints of (1) our methodology that uses a distance function to assess the suitability of a relationship, and (2) the quality of the data on those relationships.

The problem however with Blossom is that it is impractically slow when there are a large number of agents in the mating pool, even when using a highly optimised recent implementation called Blossom V [18]. The time to create the graph increases quadratically with the number of agents in the mating pool, and the time for the Blossom V algorithm increases approximately cubicly with the number of agents in the mating pool. Generally, algorithms in a simulation whose execution time increases more than linearithmically ($n \log n$) with the number of agents are impractical if modellers wish to do sensitivity analysis, calibrate parameters, or build stochastic error (or confidence) intervals (see Appendix 1). Furthermore it is not a stochastic algorithm which is often a desired feature of pair-matching.

Between the high and low precision of Blossom and RPM respectively are algorithms that approximate the minimum sum of distances:

- Random-K Pair-Matching (RKPM) is similar to RPM. For each agent a in the mating pool that still needs to be matched, it examines up to k adjacent neighbouring agents in the mating pool — where k is a user-defined constant positive integer that is usually much smaller than the number of agents in the mating pool — and matches a with the agent with the lowest distance to it. RPM is essentially RKPM with $k = 1$.

- Brute-force pair matching (BFPM) is similar to RKPM except that k is set to a value equal to or greater than the maximum number of agents in the mating pool. This means that for each agent a that still needs to be matched, it will be partnered with the remaining unmatched agent that has the shortest distance to it.

- Cluster Shuffle Pair-Matching (CSPM) relies on the existence of a cluster function as described by [10]. It sorts the agents by the value returned by the cluster function. The sorted agents are divided into a

user-specified number of clusters. Each cluster is then shuffled to introduce stochasticism. Next, as with RKPM, for each unmatched agent a, it examines the k adjacent neighbouring agents in the mating pool, choosing the one with the lowest distance to a. Since it is more complex than the above two algorithms, pseudocode for the CSPM algorithm is presented in Algorithm 3.

[10] discuss the selection of values of k in the CSPM and RKPM algorithms, and the number of clusters in the CSPM algorithm. In this work we varied k between 30 and 300, and we varied the number of clusters between 10 and 100, depending on the number of agents in the simulation's population.

In summary, a simulation therefore has a breakup and mating pool strategy (DATA vs RANDOM) coupled with a pair-matching algorithm (Blossom vs BFPM vs CSPM vs RKPM vs RPM). [7] The DATA strategy coupled with Blossom, BFPM, CSPM or RKPM accounts for the heterogeneity of agents, while the RANDOM strategy coupled with RPM treats all agents homogeneously. DATA coupled with RPM, and RANDOM coupled with Blossom, BFPM, CSPM or RKPM represent compromises between treating agents heterogeneously and homogeneously.

Accounting for group-level heterogeneity makes matches more representative of the population being studied, and prevents the model from overestimating the spread of the STI. However, all the algorithms generate at least some poor matches, i.e. matching agents across vastly different age groups or with differing sexual orientations. These "mismatches", provided they are not too frequent, actually assist the simulation by ensuring the STIs eventually cross into different subgroups. It is possible to set the model to ignore poor matches, but this comes with severe disadvantages: (a) the number of matches, especially with the RANDOM strategy or RPM algorithm, would be too few, and (b) we would in essence have several entirely independent STIs in subgroups such as as MSM and men who have sex with women (MSW) having no effect on each other, which is not realistic and defeats the purpose of using a microsimulation as opposed to an equation-based model with multiple compartments.

[7] We use "strategy" to describe the *Select and* Breakup events, and "algorithm" for the *Match event so as to keep RANDOM and DATA differentiated from Blossom, BFPM, CSPM, RKPM and RPM in the reader's mind.*

Algorithm 3 Cluster shuffle pair-matching (CSPM)

Parameters:
Agents, **an array of agents, with subscripts** $0..n-1$, **where** n **is the number of agents. For simplicity assume** n **is even.**
c, **the number of clusters to divide the agents into. For simplicity assume** c **divides into** n**.**
k, **the number of adjacent agents to consider when finding a suitable partner.**

```
1: function clusterShuffleMatch(Agents, c, k)
2:     for each agent, a, in Agents do
3:         a.weight ← cluster(a)
4:     end for
5:     sort Agents by weight
6:     clusterSize ← n/c
7:     i ← 0
8:     for each cluster do
9:         first ← i * clusterSize
10:        last ← first + clusterSize
11:        shuffle Agents[first...last − 1]        ▷ to introduce stochasticism
12:        i ← i + 1
13:    end for
14:    for each unmatched agent a in Agents do
15:        best ← ∞
16:        for each unmatched agent b in one of up to k positions
   in the array after a do
17:            d ← distance(a, b)
18:            if d < best then
19:                best ← d
20:                bestPartner ← b
21:            end if
22:        end for
23:        Make a and bestPartner partners
24:    end for
25: end function
```

Simulation set-up

General set-up

Our aim is to show the qualitative effects of design decisions on prevalence. Hence, the results presented here should be seen as illustrative; their precise impact will be a function of the particular disease, setting and time being modelled.

In these experiments, the group of MSM between the ages of 15 and 20 are initiated to an infected state in the initial population. The generic STI spreads out from this group across the population. This results in about 1 in 1,000 agents being infected in the initial population. (We also re-ran several simulations which, instead, distributed the initial infections across various demographic groups. There was no qualitative difference in the results.)

Experiment set-ups

To determine the effect of the different pair-matching algorithms on prevalence under different risk of transmission scenarios (research question I), we ran the algorithms in simulations of 20,000 and one million agents for ten years with a time step of one day. It was only feasible to execute the Blossom algorithm with 20,000 agents as it is simply too slow on a population of 1 million agents. For this experiment we chose different transmission probability scenarios (low, medium and high risk of infection as given in Table 1). Since the model is stochastic, we generally repeated each simulation 30 times to build mean final prevalence and stochastic error intervals.

A second experiment was set up to further analyse the effect of the different algorithms in combination with various different population sizes, ranging from from 10,000 to 1 million agents, answering research question II.

To analyse the effect of agent heterogeneity a two-step experiment was conducted (research question III). First, the effect of group-level heterogeneity was analysed by comparing the RANDOM and DATA strategy for agent behavior. To stabilise the number of daily breakups and mating pool entrants in the DATA strategy, the stabilisation period was set to 60 days (i.e. no ageing or infections occur in this period). To avoid bias in our comparison of strategies, the stabilisation period was also run with the RANDOM strategy.

	Risk Scenario					
	Low		Medium		High	
	Male	Female	Male	Female	Male	Female
Male	0.002	0.001	0.02	0.01	0.2	0.1
Female	0.002	0.001	0.02	0.01	0.2	0.1

Table 1: Daily risk of infection for sero-negative agent in sero-discordant relationship

Relevance to various STIs

Table 2 summarises estimated transmission probabilities for various STIs. These data suggest our findings are particularly relevant to models of HPV and gonorrhea where transmission risk is very high. However, even the transmission risks of chlamydia, HIV and syphilis may be high enough to render microsimulations of these infections sensitive to the number of agents.

An alternative explanation is that the effect of declining prevalence with increasing number of agents is a consequence of the distance-based methodology to do pair-matching that accounts for heterogeneous sexual behaviour. This would mean what we are seeing is merely an artefact of a computer algorithm, and that this is not how STIs work in the real-world. While we cannot rule out this possibility, at present we do not see why this should be the case.

Results

Effect of pair matching algorithms and transmission probability

Table 3 shows the results of the first experiment. In the low-risk scenario, there was no significant difference in prevalence by algorithm, irrespective of whether a small (20,000) or large (1 million) number of agents was used. But as the risk increased RKPM, CSPM, BFPM and Blossom showed a trend towards lower prevalence compared to RPM, and this trend was significant for simulations with 1 million agents. Moreover, all the algorithms, except RPM, calculated lower prevalence with a higher number of agents. The higher the risk of transmission the more sensitive the final prevalence was to

STI	Unit	Tranmission probability	Comment	Source
HIV	act	0.014 [95%CI 0.002;0.025]	URAI	[2]
	partner	0.404 [95%CI 0.060;0.749]	URAI	
	partner	0.217 [95%CI 0.160;0.429]	UIAI	
Syphilis	act	0.014	UAI	[12]
	partner	0.627		[1]
HPV	act	0.400 (range 0.050–1.000)	Simulated	[6]
	partner	0.270 [95%CI 0.210;0.350]	MtoF	[5]
	partner	0.310 [95%CI 0.240;0.400]	FtoM	[5]
Gonorrhea	day	0.150/0.600 (steady/casual)	MtoF	[20]
	day	0.063/0.250 (steady/casual)	FtoM	
Chlamydia	day	0.039/0.154 (steady/casual)	MtoF	[20]
	day	0.305/0.122 (steady/casual)	FtoM	

Table 2: Probabilities of infection for different STIs. (URAI = Unprotected, receptive anal intercourse; UIAI = Unprotected, insertive anal intercourse; MtoF = male to female transmission; FtoM = female to male transmission)

the number of agents in the simulation. Also, the stochastic error intervals were narrower for 1 million versus 20,000 agents for all algorithms.

Except for RPM, all the pair-matching algorithms in the high-risk scenario — and to a lesser extent with the medium-risk scenario — resulted in lower prevalence in the population with 1 million agents. By contrast, RPM generated the same prevalence, irrespective of the population size.

Blossom is too slow to run with 1 million agents, but with 100,000 agents the mean final prevalence over 12 runs was 42%, compared to 47.7% with 20,000 agents (see Table 4).

Effect of population size

To follow up the different results for the different population sizes, we ran further simulations in the high-transmission risk scenario for 10,000, 50,000 and 100,000 agents for all algorithms, and 300,000 and 600,000 agents for all algorithms except Blossom. As figure 1 shows, the simulations reveal a pattern of lower prevalence after 10 years for a higher number of agents and for the more complex algorithms.

N	Algorithm	Infection risk scenario		
		Low	Medium	High
20,000	RPM	0.3% [0.2;0.4]	1.1% [0.7;1.8]	50.8% [47.9;52.0]
	RKPM	0.3% [0.2;0.4]	1.1% [0.8;1.5]	48.9% [46.5;51.6]
	BFPM	0.3% [0.2;0.4]	1.0% [0.6;1.4]	49.3% [47.7;51.4]
	CSPM	0.4% [0.2;0.5]	1.0% [0.5;1.5]	48.2% [45.3;49.8]
	BLOSSOM	0.3% [0.2;0.4]	1.0% [0.8;1.4]	47.7% [46.5;48.9]
1,000,000	RPM	0.3% [0.3;0.3]	1.1% [1.0;1.1]	51.1% [50.9;51.3]
	RKPM	0.3% [0.3;0.3]	0.8% [0.8;0.8]	46.2% [46.0;46.5]
	BFPM	0.3% [0.2;0.3]	0.5% [0.4;0.5]	43.8% [43.4;44.4]
	CSPM	0.3% [0.3;0.3]	0.8% [0.7;0.8]	37.7% [36.3;39.2]
	Blossom	NA	NA	NA

Table 3: Prevalence after 10 years of low, medium and high infection risk scenarios for pair-matching algorithms, sorted by prevalence of high risk scenario. Each entry in the Low, Medium and High columns is the mean and 95% stochastic error interval of 30 runs.

Population	CSPM	Blossom
10,000	47.9%	48.7%
50,000	45.5%	44.3%
100,000	43.9%	42.0%

Table 4: Comparison of prevalence for CSPM against Blossom for three different population sizes. Each entry in the CSPM and Blossom columns is the mean of 12 runs.

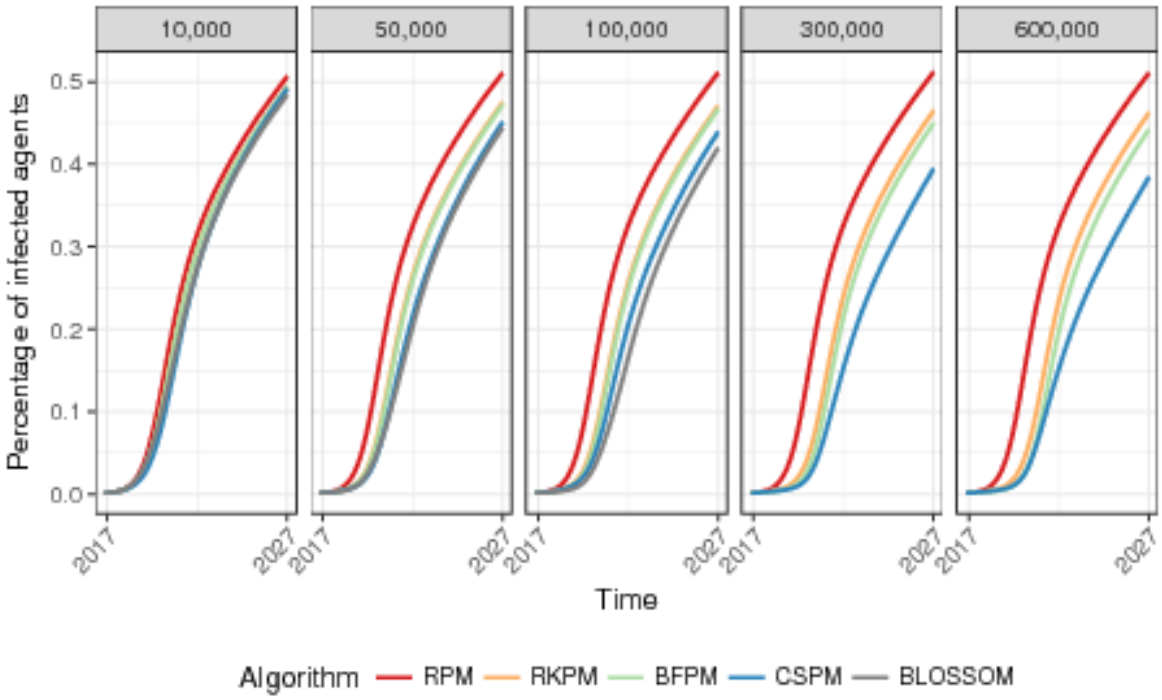

Figure 1: **Disease prevalence by population**
Mean disease prevalence after 10 years of 30 simulation runs for different population sizes of 10,000, to 600,000 agents.

We also examined the behavior of the CSPM algorithm for different population sizes (10,000, 20,000, 50,000, 100,000, 500,000 and 1 million) to explore if the effect of decreasing prevalence with increasing number of agents tapers off. Figure 2 depicts this visually, showing that reduced prevalence is much greater moving from 100,000 to 500,000 agents, than from 500,000 to 1 million agents.

Running the medium-risk scenario for long enough demonstrated that the effect of declining prevalence as the number of agents increases that we saw with the high risk scenario was still present, but took longer to be as noticeable. Average final prevalence for 10,000 agents for a 100 years using CSPM was 81.7% [95%CI: 65.4;90.2 over 200 runs] for 10,000 agents versus 74.1% [95%CI 73.2;74.9 over 16 runs] for 1 million agents.

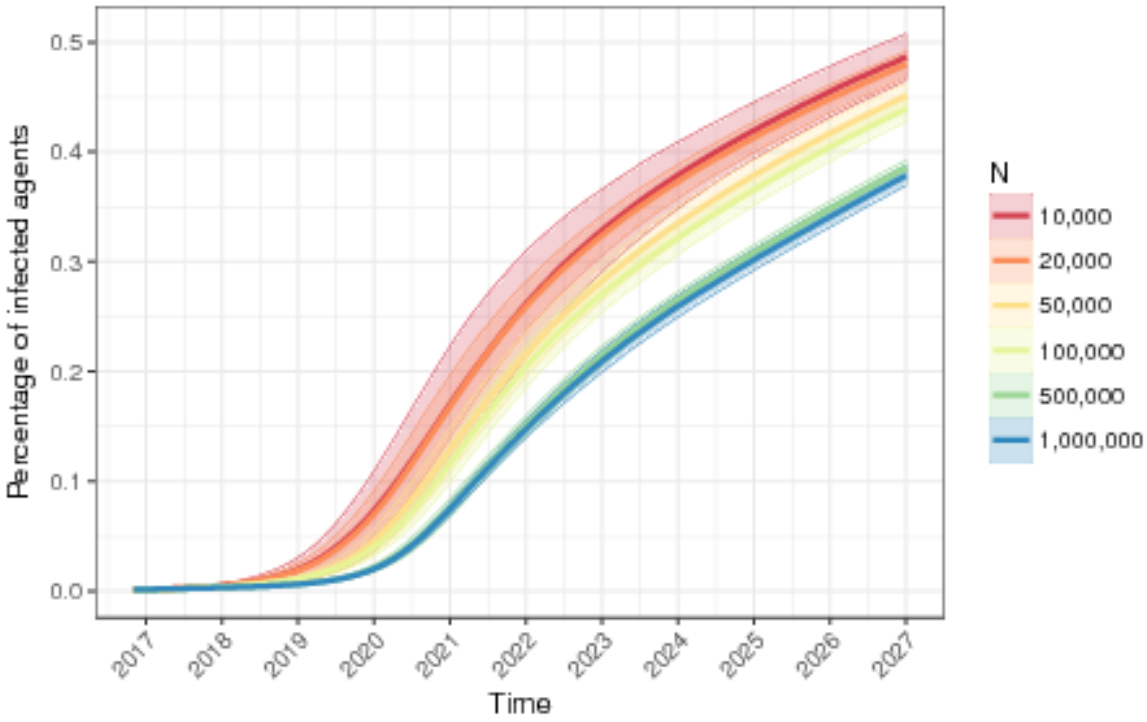

Figure 2: **DATA/CSPM simulations for different population sizes.** DATA/CSPM simulations run for ten years (3,650 days) on 20,000, 300,000, 500,000 and 1 million agents. The lower prevalence with higher number of agents appears to be a consequence of the longer time that the STI takes to begin growing rapidly in its early stage.

Heterogeneity of agents entering the mating pool and breaking up

Effect of group-level heterogeneity

We used a similar methodology to compare the DATA and RANDOM strategies (using only CSPM and RPM as the pair-matching algorithms). The daily number of breakups and agents entering the mating pool in the RANDOM strategy was set to closely match (less than 0.2% difference) the daily average for the DATA strategy. The results are presented in Table 5.

The DATA strategy coupled with CSPM resulted in the lowest estimates of prevalence in all risk scenarios, followed by RANDOM coupled with CSPM.

The approach that treats agents entirely homogeneously, RANDOM coupled with RPM, estimated the highest prevalence.

A further finding, one that's expected, is that in simulations using the DATA strategy and non-random pair-matching algorithms sub-groups of agents at higher risk (e.g. MSM or younger adults) have a relatively higher prevalence, and lower risk agents, such as older adults have relatively lower prevalence than random approaches. For example if we set the initial infection rates uniformly to 1% across the population, the final prevalence in a 10-year simulation using DATA strategy coupled with CSPM shows much higher prevalence at younger, higher risk age groups, and lower prevalence at older age groups. With a RANDOM strategy coupled with RPM, the distribution across age groups is more uniform.

Strategy	Algorithm	# Agents	Infection risk scenario		
			Low	Medium	High
DATA	CSPM	1,000,000	0.3% [0.3;0.3]	0.8% [0.7;0.8]	37.7% [36.3;39.2]
		20,000	0.4% [0.2;0.5]	1.0% [0.5;1.5]	48.2% [45.3;49.8]
RANDOM	CSPM	1,000,000	0.9% [0.8;0.9]	39.0% [37.8;40.2]	81.7% [78.3;88.1]
		20,000	3.2% [2.1;4.0]	100% [99.9;100]	100% [100;100]
	RPM	1,000,000	3.5% [2.9;3.9]	100% [100;100]	100% [100;100]
		20,000	3.4% [1.3;4.8]	100% [100;100]	100% [100;100]

Table 5: Prevalence after 10 years of low, medium and high infection risk scenarios for breakup and mating pool strategies, sorted by prevalence of high risk scenario. Each entry in the Low, Mean and High columns is the mean and 95% stochastic error interval of 30 runs.

Effect of individual-level heterogeneity

To test individual-level heterogeneity we ran 30 simulations each with DATA coupled with CSPM for 10,000, 50,000, 100,000, and 500,000 agents over 10 years with the high risk scenario with individual-level heterogeneity switched on and then switched off. The results are presented in Table 6. There were no significant differences in mean final prevalence. Stochastic error intervals were also roughly the same width.

Increasing heterogeneity from group- to individual-level in this way and to this extent does not appear to affect *overall* prevalence. However, we do not rule out the possibility that a more in-depth analysis may revise this finding.

Agents	Group	Individual
10,000	48.5% [45.8;51.0]	47.8% [42.2;51.9]
50,000	45.9% [44.5;47.4]	46.3% [44.4;48.1]
100,000	43.9% [41.6;45.5]	44.4% [43.1;45.8]
500,000	38.6% [37.5;39.7]	40.2% [38.5;41.9]

Table 6: Mean prevalence and 95% stochastic error interval over 30 runs comparing group- (age, sex and sexual orientation) versus individual-level heterogeneity (age, sex and sexual orientation modified by factors set for each agent).

Effects of other design decisions

We explored several other issues that may effect incidence and prevalence estimates.

Changing the number of partnerships

In the default data set we used, the average number of partners per agent per year was 2.9. 98% of these are casual interactions. To see the effect of reducing partnerships across the model, for the DATA strategy coupled with CSPM, we compared the effect on prevalence of reducing the casual partnerships by 50%, 75% and 90%.

Table 7 presents the results of these simulations, showing that even after massively reducing the partnerships the prevalence still declines with an increase in the number of agents for the medium- and high-risk scenarios. However, the effect becomes less pronounced as the number of partnerships declines.

# Agents	% of default	Infection risk scenario		
		Low	Medium	High
20,000	50	0.4% [0.2;0.5]	0.7% [0.4;1.1]	29% [24.1;32.8]
	25	0.3% [0.2;0.4]	0.6% [0.4;0.8]	6.1% [3.8;8.1]
	10	0.3% [0.2;0.4]	0.5% [0.3;0.7]	1.1% [0.6;1.6]
1,000,000	50	0.3% [0.3;0.3]	0.5% [0.5;0.6]	16.7% [16.1;17.7]
	25	0.3% [0.3;0.3]	0.5% [0.4;0.5]	2.8% [2.6;3]
	10	0.3% [0.3;0.3]	0.4% [0.4;0.4]	0.8% [0.8;0.9]

Table 7: Results of simulations with reduced number of casual partnerships. Each entry in the Low, Medium and High is the mean and 95% stochastic error interval over 30 runs.

Discouraging previous partnerships

A dilemma we had was how to deal with a potential new partnership between agents who had previously been partners. We implemented two variations of the distance function, one that keeps track of all partnerships and penalises potential pairings between agents who have previously been in a relationship, and one that does not keep track of partnerships at all. We found no material difference in results using these two methods.

There are practical implementation consequences of this finding. Our largest simulations (40 million agents, see Appendix 1) use a large amount of memory. The number of simulations we could run in parallel was limited by the available memory on our machine. The largest data structure by far in our simulations, despite extensive optimisation, is the one that keeps track of previous partnerships. With the finding that penalising previous partnerships made no difference to the results, this data structure could be disabled on large simulations, allowing more simulations to be run in parallel.

Discussion

Our experiments found the following:

- Random mating of agents in an STI microsimulation leads to higher

infection incidence. This is unsurprising and a consequence of not accounting for heterogeneous sexual behaviour.

With less randomness in pair-matching, the STI is more confined to subgroups that have riskier sexual behaviour profiles. While when mating and breaking up randomly (i.e. ignoring heterogeneity), the infection spreads uniformly through the population.

- Stochastic error intervals narrow as the number of agents in the simulation population increases.

- The incidence and prevalence that microsimulations using sophisticated pair-matching algorithms estimate are sensitive to the number of agents in the model. As the risk of infection for the sero-negative partner in a sero-discordant partnership increases, or as the frequency of partnership formation and breakups in the population increases, the more sensitive to the number of agents the model becomes.

- CSPM is a pair-matching algorithm that (1) appears to produce results comparable to the Blossom algorithm (which optimally approximates the distribution of partnerships in a mating pool), (2) is practical to use in microsimulations with very large numbers of agents, and (3) is practical to use when many thousands of simulations need to be run in a reasonable amount of time.

Effect of population size

The finding that incidence declines as population increases when heterogeneity is accounted for is surprising.

This does not appear to be explained by the increasing quality of matches as the mating pool increased. The average distance between the agents in partnerships in the algorithms is approximately 14, 20.5, 26.5, 29.5 and 61 for Blossom, BFPM, CSPM, RKPM and RPM respectively (higher scores mean worse matches). These values do not change much as the number of agents increases. In fact, CSPM has a slightly lower average score for 20,000 agents than 1 million agents. If the larger mating pools resulted in better quality matches, we would expect the average distance to decrease. Nor is it explained by the number of partnerships, which are proportional to the population size.

Figure 2 depicts what is occurring. During the early stages of the epidemic incidence is lower for a longer period of time, resulting in lower prevalence at any given time in the ten year period of the simulation. When we started simulations in an already mature epidemic (e.g. 10% prevalence), the effect of prevalence being lower with higher numbers of agents disappeared.

A possible explanation for what is happening is that as the number of agents decreases, an infection occurring in a relatively low-risk subgroup of agents (e.g. WSW aged between 45 and 50 years old) has a disproportionately greater effect on the number of infections that will subsequently occur in that subgroup. For example, if there are only 10 WSW aged 45 to 50 in a simulation and one of them, via a poor match or stochastic variation, becomes infected, then 10% of the subgroup is immediately infected. However if in a much bigger population there are 100 agents in this subgroup, then only 1% of this low-risk subgroup is infected. On consequent time-steps, the risk of prevalence increasing substantially in this subgroup is much higher for the smaller population. This is particularly the case for Blossom by virtue of the fact that it generally makes better matches than the other algorithms. For CSPM the effect may be due to it clustering agents that are more likely to be paired; even poor matches will be in neighbouring or nearby clusters, and partnering within clusters is accentuated as the population grows.

A similar explanation suggested to the authors is that the greater assortativity — the propensity with which agents are more likely to be matched with agents similar to themselves — associated with the non-random pair-matching algorithms is affected by population size. Perhaps this creates networks of agents who by change are more isolated from the other agents. Consequently, while infections may spread rapidly in some of these sub-networks, they don't easily cross into the main network of agents. Alternately, in some of these sub-networks, infections may never seed. Both of these possibilities may result in lower prevalence.

These explanations are currently fuzzy and insufficiently compelling as they stand. Further research is needed to understand this phenomenon properly and offer an explanation with confidence.

The heterogeneity-dilemma

Our results raise a dilemma for STI microsimulation modellers. The advantage of microsimulations over equation-based models is that the former can practically account for greater heterogeneity. However, a pair-matching algo-

rithm such as Blossom that best accounts for this heterogeneity is extremely slow. It can be used effectively only with smaller population sizes or choosing small subgroups of a population (e.g.,[3]). However, when the risk of infection is high — e.g. for HPV and gonorrhea —, the time horizon is long or the turnover of partnerships is high, modelling with an agent population that is much lower than the real-world population of interest may considerably underestimate incidence and prevalence. Although in practice this would be corrected by calibrating the model to real-world data points, the calibration process may in turn result in parameter values (i.e., for infection rates) being set far off their real-world values, so that the further into the future the model projects the greater will be the error in its estimates.

Modellers may wish to consider using an algorithm such as CSPM that usually offers a good trade-off between speed and approximation of the distribution of relationships in the population being studied. Ideally a simulation should have a similar number of agents as the population being studied. This is often impractical though, and even where it is practical, the poor quality of data on the distribution of partnerships based on sex, age, sexual orientation and even the role of geographical location, is a much bigger problem.

Of course, if the risk of infection per serodiscordant partnership is low then using a large population for the microsimulation may be unnecessary. The same is true if the simulation begins when an epidemic that is already mature.

Limitations

Our analysis has several limitations:

- The DATA strategy is based on sex survey data, with the well-documented problems that this presents [23].
- Our modelling of casual relationships is unsophisticated, and possibly overstates casual sex as one-night stands and understates short-term relationships involving a few sexual encounters. However, our results appear to be robust when accounting for this by greatly reducing the number of casual partnerships.
- We did not model varying the risk of transmission over the course of an infection, for example, the higher transmission risk of HIV during

primary infection. This would be particularly interesting to examine in research that extends our work.

- We did not model concurrent relationships, although they might play a significant role in the spread of STIs [21]. This too would be an interesting way to extend our research.

- As noted in the introduction, we did not model death, healing, condom use, circumcision and other factors, as including these would confuse the analysis and make it difficult to isolate the role of the mating pool and breakup strategy, and pair-matching algorithm.

- We have not done systematic subgroup analysis, e.g. the effect of the different algorithms on a particular 5-year age group or sexual orientation.

In preparing this paper we ran tens of thousands of simulations with the number of agents ranging from 10,000 to as high as 40 million using affordable consumer hardware. The feasibility of this is likely of interest to other modellers; Appendix 1 contains further notes on our implementation.

Further research needs to be done refining the cluster function of CSPM, as well as identifying ideal values of k (for RKPM as well) and the number of clusters. A deeper analysis of poor matches, and how frequently to block or allow them is also needed. We also recommend examining the effect of using the CSPM algorithm to model HPV or gonorrhoea in real-world populations, using different numbers of agents in the model.

Conclusion

Microsimulations have become a popular method for the analysis of the spread of STIs and for the evaluation of interventions to alleviate them. While the effects of structural assumptions about pair formation and infectivity are well known for the classical method of differential equations, the analysis of design decisions of microsimulations are less well understood.

Our findings contribute to closing this gap by providing insights into the effect of different matching algorithms for various infection rates. Additionally, we found that there exist fast pair-matching algorithms that provide a practical way for microsimulation modellers to account for heterogeneity in

sexual behaviour and without limiting the population to a small subgroup. Our findings may also inform reviewers of STI microsimulations about the extent to which the pair matching methodology can influence the results of a model.

Supporting information

S1 Appendix. This appendix aims to provide a few pointers to implementing fast microsimulations that can accommodate large numbers of agents. It is mainly aimed at microsimulation modellers who do not have a formal computer science background.

Our model is capable of doing large microsimulations fast. For example, running single-threaded, it can match nearly 600,000 agents per second in a mating pool using the CSPM algorithm on a laptop with an i5 processor running at 2.5GHz. With the agent population set to 40 million, it takes approximately 5.2 hours to run five 10-year simulations in parallel, updated daily, on an Intel Xeon with 20 cores at 2.3Ghz, with 32GB RAM (the RAM was the limiting factor; we could not use more than 5 or 6 threads at a time without running out of memory). Once birth, death, healing, and migration events, as well as a concurrency feature are added, this will slow down, but not substantially, because we expect the time taken for all of these events to increase linearly with the number of agents. On the same machine, using multithreading, we also ran 20,000 simulations of 10,000 agents (10 years, time step equal to a day) in a little more than 6 hours. This is a mean of just over 22 seconds per simulation. Running multiple threads in parallel slows down the execution of a single run because of resource contention. A single-threaded run takes about two seconds.

In a discrete microsimulation of the form described by Algorithm 1, it is the events that consume the most time. The time taken for most events to complete increases linearly with the number of agents. These include ageing, death, healing, select, and breakup events. However, pair-matching requires interaction between agents. The time to run BFPM increases quadratically with the size of the mating pool. The time taken to run the Blossom algorithm is worse: approximately cubic with the number of agents. It takes over 2 hours to match 5,000 agents in a mating pool on an i5 processor running at 2.5GHz (using our default data that would be a simulation containing approximately 1.4 million agents). So simulating 10 years with a 60 day sta-

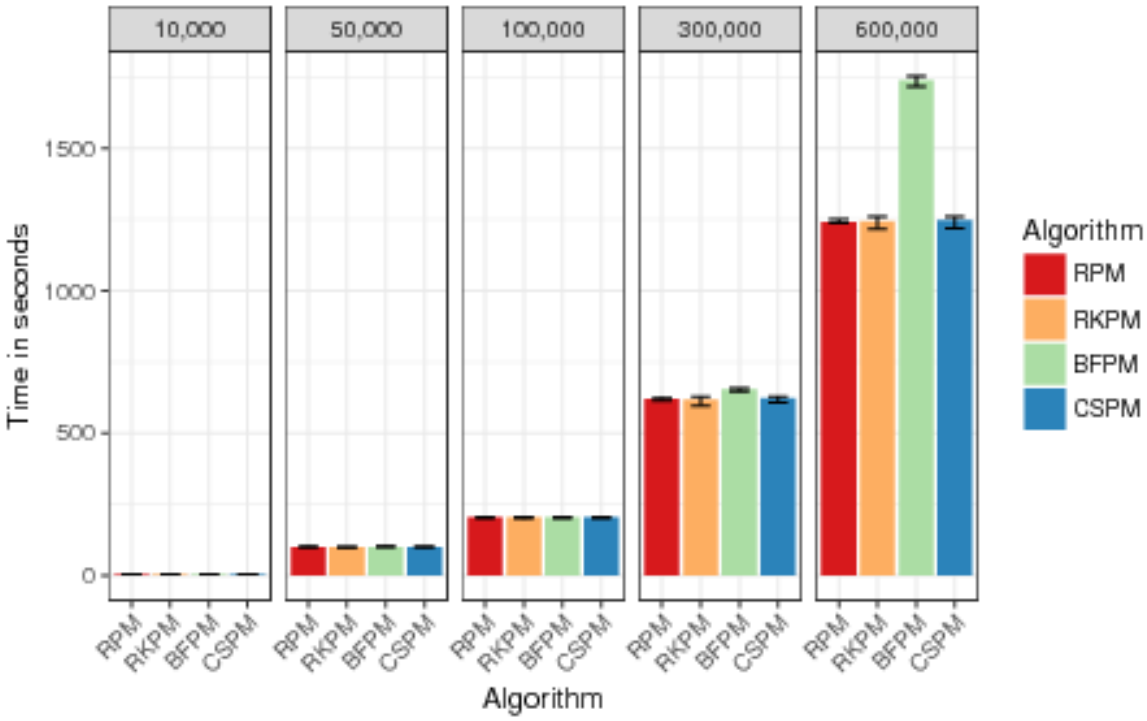

Figure 3: **Algorithm execution times**
Average execution time in seconds for each algorithm (except Blossom) when 30 simulations are run in parallel. Note that there are other time-consuming events in a simulation other than pair-matching, and that times are approximate.

bilisation period would take about 7,420 hours, or over 300 days. Perhaps waiting this long for the outcome of a physics or chemistry experiment with precise inputs and outputs may be worth it, but waiting more than a day or two for results of a single simulation in the world of STI modelling where nearly every input is a rough approximation seems pointless to us.

Table 8 and Figure 3 present execution times of simulations using the different pair-matching algorithms.

RPM's time increases linearly, but it does very poor matching. The time for CSPM and RKPM increases loglinearly ($n \log n$, where n is the size of the mating pool) with the number of agents in the mating pool. For a large simulation of 40 million agents to be done in a practical amount of time, it

is vital to use an algorithm whose time efficiency is loglinear or better.

Also it is vital to put agents who need to be paired on a particularly time step into a mating pool. This pool is almost always a small fraction of the total number of agents. Determining which agents enter the mating pool requires an algorithm with linear efficiency. Pairing the agents intelligently requires an algorithm with at least loglinear efficiency (as far as we can tell). Skipping the mating pool creation, and subsuming it into the pair-matching algorithm is false efficiency, because instead of the pair-matching algorithm executing in time that increases loglinearly with the number of agents in the mating pool, it will execute with time that increases loglinearly with the total number of agents in the simulation, a much larger number.

In computer science, we describe algorithms whose efficiency increases linearly, loglinearly, quadratically and cubicly with a dataset with n elements as $O(n)$, $O(n \log n)$, $O(n^2)$ and $O(n^3)$ respectively. We call this Big O notation and the preceding list is ordered from fastest to slowest efficiency [8]. When optimising code, changing a time-consuming algorithm to one whose Big O efficiency is better is usually more productive than tinkering with the implementation details of a fundamentally inefficient algorithm. For example, no matter how much we optimise our implementation of the BFPM algorithm it will not be as fast as an unoptimised competent implementation of the CSPM algorithm. This is a necessarily simplified discussion of Big O. For further details see one of many textbooks on the subject, eg. [17].

From our knowledge of the field and discussions with other modellers, it appears some leading microsimulation models are coded in interpreted languages such as R, Visual Basic or Python. While microsimulations coded in these languages have many uses, there is unfortunately little prospect that these languages can develop microsimulations that can manage the kind of loads described in this paper. For heavy-duty simulation it is vital to use a programming language that is either compiled directly into native machine code (e.g. C, C++, Rust, Fortran, Common Lisp, Swift, Go) or the byte code of a virtual machine (e.g. Java, Clojure, Scala).

Our microsimulation, which we call FastSTI, is open source and, we hope, well commented and easy to read. Modellers are encouraged to download it, adapt it for their needs, report bugs to us, and send us queries.

[8] This is a very simplified discussion and intentionally omits the point that Big O notation are sets to which algorithms belong, or that a less efficient set includes all the algorithms in more efficient sets. These details are simply unnecessary here.

Agents	Blossom	BFPM	CSPM	RKPM	RPM
10,000	23	2	2	2	1
100,000	2,960	81	43	42	12
1,000,000	N/A	8,810	633	601	218

Table 8: Time in seconds for a single execution on an Intel Xeon i7 2.3GHZ without any other significant load on the machine. Note that there are other time-consuming events in a simulation other than pair-matching, and that times are approximate.

S2 Appendix. As stated in the Materials and Methods section, the input parameters have been derived from the PAIRFAM study. The data includes eight waves of observations from 2008 to 2016 for three age cohorts born 1971-73, 1981-83 and 1991-93. The data is freely available at `http://www.pairfam.de/` after becoming a registered user.

While the probabilities for breakups and entering the mating pool were calulated directly from the data, the probability of casual sex was smoothed using a cubic-spline regression with age as the only explanatory variable and using the predicted values for the complete age-range of 12 to 50.

Acknowledgments

This paper uses data from the German Family Panel pairfam, coordinated by Josef Bruederl, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz Neyer, and Sabine Walper. pairfam is funded as long-term project by the German Research Foundation (DFG).

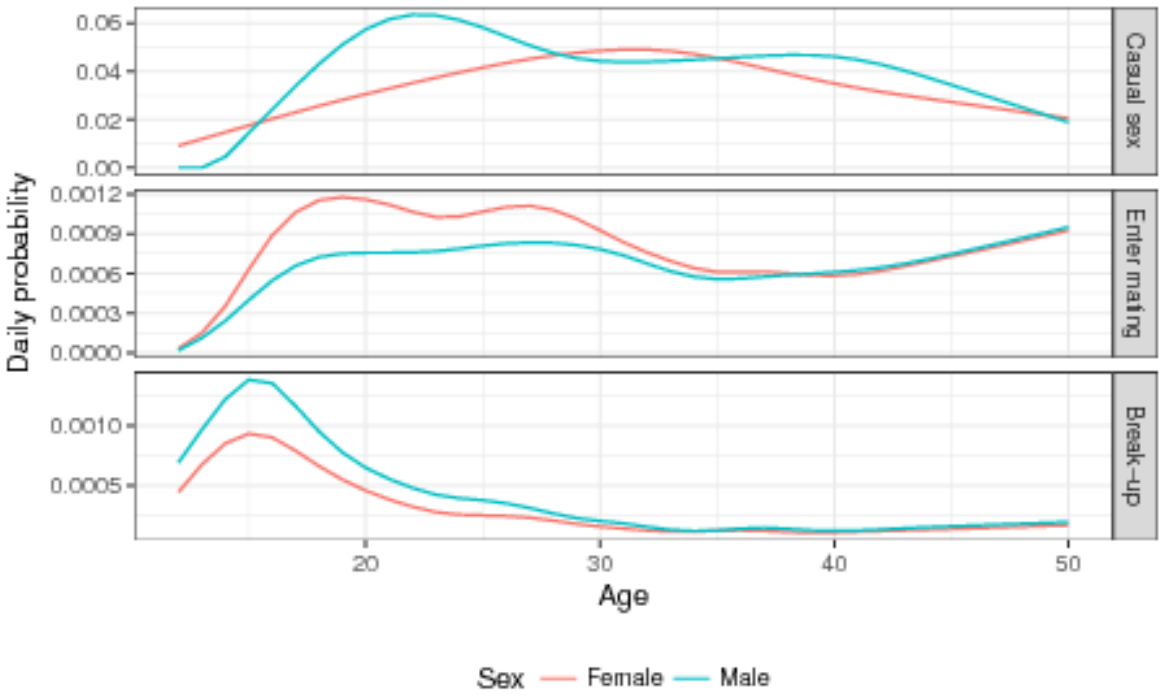

Figure 4: **Mating pool probabilities**
Daily probabilities (by age and sex) of entering the mating pool for agents who are single, or breaking up for agents in relationships.

Age	Casual Sex		Long-term relationships			
	Enter mating pool		Enter mating pool		Breakup	
	Female	Male	Male	Female	Male	Female
12	0.00914	0.00000	0.00002	0.00003	0.00069	0.00044
13	0.01194	0.00000	0.00011	0.00015	0.00097	0.00067
14	0.01474	0.00453	0.00024	0.00035	0.00122	0.00085
15	0.01753	0.01432	0.00040	0.00063	0.00139	0.00093
16	0.02030	0.02418	0.00055	0.00089	0.00136	0.00090
17	0.02302	0.03392	0.00066	0.00106	0.00117	0.00079
18	0.02564	0.04307	0.00072	0.00115	0.00095	0.00066
19	0.02814	0.05110	0.00075	0.00118	0.00078	0.00055
20	0.03056	0.05744	0.00076	0.00116	0.00065	0.00045
21	0.03291	0.06168	0.00076	0.00112	0.00055	0.00038
22	0.03519	0.06356	0.00076	0.00106	0.00048	0.00032
23	0.03741	0.06327	0.00077	0.00102	0.00042	0.00027
24	0.03956	0.06130	0.00079	0.00103	0.00039	0.00025
25	0.04160	0.05818	0.00081	0.00107	0.00037	0.00025
26	0.04349	0.05447	0.00083	0.00110	0.00035	0.00024
27	0.04517	0.05077	0.00083	0.00111	0.00031	0.00023
28	0.04658	0.04765	0.00083	0.00108	0.00026	0.00020
29	0.04773	0.04543	0.00081	0.00101	0.00022	0.00017
30	0.04859	0.04426	0.00078	0.00093	0.00020	0.00015
31	0.04907	0.04389	0.00074	0.00084	0.00018	0.00014
32	0.04905	0.04400	0.00068	0.00076	0.00015	0.00012
33	0.04846	0.04436	0.00062	0.00069	0.00013	0.00011
34	0.04725	0.04484	0.00057	0.00064	0.00012	0.00011
35	0.04555	0.04538	0.00056	0.00061	0.00012	0.00012
36	0.04352	0.04595	0.00056	0.00061	0.00014	0.00012
37	0.04132	0.04646	0.00057	0.00061	0.00014	0.00012
38	0.03910	0.04681	0.00059	0.00060	0.00013	0.00011
39	0.03696	0.04677	0.00060	0.00059	0.00012	0.00010
40	0.03496	0.04616	0.00061	0.00058	0.00012	0.00010
41	0.03315	0.04487	0.00062	0.00059	0.00012	0.00011
42	0.03153	0.04289	0.00064	0.00062	0.00013	0.00012
43	0.03005	0.04036	0.00067	0.00065	0.00014	0.00012
44	0.02865	0.03745	0.00071	0.00069	0.00015	0.00013
45	0.02728	0.03439	0.00075	0.00073	0.00016	0.00013
46	0.02593	0.03130	0.00078	0.00077	0.00016	0.00014
47	0.02458	0.02820	0.00082	0.00081	0.00017	0.00015
48	0.02323	0.02511	0.00087	0.00084	0.00018	0.00015
49	0.02188	0.02201	0.00091	0.00088	0.00018	0.00016
50	0.02053	0.01891	0.00095	0.00093	0.00019	0.00016

Table 9: Daily probabilities of entering the mating pool for agents who are single, or breaking up for agents in relationships.

References

[1] L. J. Alexander and A. G. Schoch. Prevention of Syphilis. *Archives of Dermatology and Syphilology*, 59(1):1–10, 1949.

[2] R. F. Baggaley, R. G. White, and M. C. Boily. HIV transmission risk through anal intercourse: Systematic review, meta-analysis and implications for HIV prevention. *International Journal of Epidemiology*, 39(4):1048–1063, 2010.

[3] E. Beck, M. Birkett, B. Armbruster, and B. Mustanski. A data-driven simulation of hiv spread among young men who have sex with men: the role of age and race mixing, and stis. *Journal of acquired immune deficiency syndromes*, 70(2):186–194, 2015.

[4] R. Brookmeyer, D. Boren, S. D. Baral, L. G. Bekker, N. Phaswana-Mafuya, C. Beyrer, and P. S. Sullivan. Combination HIV prevention among MSM in South Africa: Results from agent-based modeling. *PLoS ONE*, 9(11), 2014.

[5] A. N. Burchell, F. Coutlee, P. P. Tellier, J. Hanley, and E. L. Franco. Genital transmission of human papillomavirus in recently formed heterosexual couples. *Journal of Infectious Diseases*, 204(11):1723–1729, 2011.

[6] A. N. Burchell, H. Richardson, S. M. Mahmud, H. Trottier, P. P. Tellier, J. Hanley, F. Coutlée, and E. L. Franco. Modeling the sexual transmissibility of human papillomavirus infection using stochastic computer simulation and empirical data from a cohort study of young women in Montreal, Canada. *American Journal of Epidemiology*, 163(6):534–543, 2006.

[7] J. W. Eaton, L. F. Johnson, J. A. Salomon, T. Bärnighausen, E. Bendavid, A. Bershteyn, D. E. Bloom, V. Cambiano, C. Fraser, J. A. C. Hontelez, S. Humair, D. J. Klein, E. F. Long, A. N. Phillips, C. Pretorius, J. Stover, E. A. Wenger, B. G. Williams, and T. B. Hallett. HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. *PLoS medicine*, 9(7):e1001245, 2012.

[8] J. Edmonds. Maximum matching and a polyhedron with $0, 1$ vertices. *J. of Res. the Nat. Bureau of Standards*, 69 B:125–130, 1965.

[9] E. Escobar, R. Durgham, O. Dammann, and T. J. Stopka. Agent-based computational model of the prevalence of gonococcal infections after the implementation of HIV pre-exposure prophylaxis guidelines. *Online journal of public health informatics*, 7(3):e224, 2015.

[10] N. Geffen and S. Scholz. Efficient and effective pair-matching algorithms for agent-based models. *Journal of Artificial Societies and Social Simulation*, 20(4):8, 2017.

[11] J. D. Goldhaber-Fiebert, N. K. Stout, J. Ortendahl, K. M. Kuntz, S. J. Goldie, and J. A. Salomon. Modeling human papillomavirus and cervical cancer in the united states for analyses of screening and vaccination. *Population Health Metrics*, 5(1):11, Oct 2007.

[12] R. T. Gray, A. Hoare, P. D. McCann, J. Bradley, I. Down, B. Donovan, G. Prestage, and D. P. Wilson. Will changes in gay men's sexual behavior reduce syphilis rates? *Sexually transmitted diseases*, 38(12):1151–1158, dec 2011.

[13] J. A. C. Hontelez, M. N. Lurie, T. Bärnighausen, R. Bakker, R. Baltussen, F. Tanser, T. B. Hallett, M.-L. Newell, and S. J. de Vlas. Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study. *PLoS medicine*, 10(10):e1001534, Oct. 2013.

[14] J. Huinink, J. Brüderl, B. Nauck, S. Walper, L. Castiglioni, and M. Feldhaus. Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung - Journal of Family Research*, 23:77–101, 2011.

[15] J. A. Jacquez, C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry. Modeling and analyzing hiv transmission: the effect of contact patterns. *Mathematical Biosciences*, 92(2):119–199, 1988.

[16] L. F. Johnson and N. Geffen. A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. *Sexually Transmitted Diseases*, 43(3):139–146, Mar. 2016.

[17] D. E. Knuth. *The Art of Computer Programming, Volume 1: (3rd Ed.) Fundamental Algorithms*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1997.

[18] V. Kolmogorov. Blossom V: a new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation*, 1(1):43–67, June 2009.

[19] M. Kretzschmar and K. Dietz. The effect of pair formation and variable infectivity on the spread of an infection without recovery. *Mathematical Biosciences*, 148(1):83–113, 1998.

[20] M. Kretzschmar, Y. T. Van Duynhoven, and A. J. Severijnen. Modeling prevention strategies for gonorrhea and Chlamydia using stochastic network simulations. *American Journal of Epidemiology*, 144(3):306–317, 1996.

[21] M. Morris and M. Kretzschmar. A microsimulation study of the effect of concurrent partnerships on the spread of HIV in Uganda. *Mathematical Population Studies*, 8(2):109–133, 2000.

[22] B. Nauck, J. Brüderl, J. Huinink, and S. Walper. The German Family Panel (pairfam), 2014.

[23] K. E. E. Schroder, M. P. Carey, and P. A. Vanable. Methodological Challenges in Research on Sexual Risk Behavior: II. Accuracy of Self-Reports. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 26(2):104–123, Oct. 2003.

6.2 Does STI modelling increase our knowledge?

The article in this chapter has shown that with different design decisions a model can estimate markedly different outcomes. This finding presents an opportunity to consider in more depth the question of what knowledge, if any, models actually provide.

Here are examples of models giving inconsistent results, either with other models or with direct measurements of the natural population being modelled:

- In the last two articles of this work we showed how using different algorithms for matching agents results in different outcomes. And in the final article, we found that incidence drops as the number of agents in the microsimulation increases, meaning that the same model can give different results simply by changing the number of agents . Of course models will be calibrated so that incidence or prevalence matches estimates of the natural population being studied, but that does not inspire confidence in projections beyond the calibrated data.

- Eaton et al. (2015) assessed ten models of the South African HIV epidemic against survey data. All the models estimated lower prevalence for 2012 than a survey estimate, with eight estimating below the survey's 95% confidence interval. Eight models estimated that prevalence would stay the same or decline between 2008 and 2012 whereas it increased across two surveys. The models' estimates also differed sig-

nificantly from each other on some outputs, though they did match survey data in some respects, including predicting approximately the same number of people on ART. Also, it is possible that the survey data was inconsistent and perhaps faultier than some of the models.

- As explained in the introduction to this dissertation, Johnson et al. (2016) compared a microsimulation and a deterministic compartmental model on six STIs, with the latter model always estimating greater prevalence.

Do these contradictory results mean that modelling is presently at best an immature field that fails to increase our knowledge of epidemics? The question may be asked of non-STI disease models too. Butler (2014) discussed the fact that Ebola models overestimated the the number of cases in Liberia during the 2014 epidemic. The article was subtitled "Rate of infection in Liberia seems to plateau, raising questions over the usefulness of models in an outbreak", sparking a strongly worded response by Rivers (2014) who wrote: "Your assertion that models of the Ebola epidemic have failed to project its course misrepresents their aims ... They helped to inspire and inform the strong international response that may at last be slowing the epidemic". The letter concluded "Epidemics are affected by countless variables, so uncertainty is a given. Models synthesize available information. Without them, there is little to guide decision-makers during an outbreak. Their importance goes beyond providing forecasts."

Perhaps this is the best we can hope for with STI models too. After all, even for the most sophisticated microsimulations, the assumptions of how

sexual behaviour occurs are highly idealised. Sexual partnership formation and dissolution in the natural world, as well as the risk of infection per partnership, is extremely complicated and impossible to capture precisely in models.

Yet models have had a crucial role in HIV-related debates as the first three articles of this dissertation show. Models cannot be expected to be *precise*, but they can be expected to give broadly *accurate* answers to the questions they try to answer [2]. So the first models of HIV in South Africa in the early 1990s, such as Doyle et al. (1991), accurately showed that a large epidemic was underway and that it would become massive, but they could not predict the size of the epidemic with much precision. In the early 2000s, the ASSA models and subsequent work were accurate in that they showed that ART would save many lives and that implementing it would be affordable (Geffen et al., 2003). But they could not precisely predict how many lives would be saved, how much life-expectancy would increase, or exactly how much it would cost. In the late 2000s, the Granich model showed — hopefully accurately; it's too early to tell — that a policy of universal ART would prevent many new infections in South Africa, but its projections will almost certainly be imprecise (Granich et al., 2009).

Meticulously constructed models have made plausible estimates of prevalence and life-expectancy. They did not precisely match subsequent survey data, but they were often close enough for practical purposes, e.g. increasing

[2]Accuracy in this context means providing an answer that is true, even if the range of the answer is wide, Precision in this context means providing an exact answer. e.g. "Prevalence is between 8% and 10%." may be accurate, while "Prevalence is 8.23%" is precise, but almost certainly inaccurate. This differentiation between precision and accuracy was once related to the author by Edwin Cameron.

public awareness of the seriousness of HIV (Doyle, 1993), influencing policy debates and court cases (Dorrington et al., 2001; Dorrington, 2002), and helping to assess the current state of the epidemic (Johnson, 2014).

Furthermore, as discussed in Chapter 1 and 4.2, adding complexity to models, such as by using microsimulations rather than deterministic equation-based ones, has potential pitfalls. In the attempt to create more realism with complex microsimulations (or even complex deterministic equation-based models) modellers sometimes produces results that are difficult to explain or even counter-intuitive, such as those in this chapter's paper. Simple models may, in some cases, produce results are easier to explain and are more consistent with theory, even if they sacrifice realism.

There is much research to be done to improve STI modelling. We have shown how design and methodological difference between models produce different results. We have barely considered how in addition to this, models can reach dramatically different — and often wrong — conclusions because of methodological differences in, or problems with, data collection, such as sex surveys, prevalence surveys and death registration. Nevertheless, despite considerable concerns, models have their place in clarifying our knowledge of STIs and, in particular, the South African HIV epidemic.

Chapter 7

Conclusions

In this work, we have considered models of STIs from their role in policy making about the South African HIV epidemic through to technical considerations about how to match individuals in sexual relationships.

The first publication of this dissertation provides an overview of HIV modelling. The second described modelling controversies with respect to policy-making. The third examined the debate on when to start ART, which was informed by modelling. Moreover, after interest receded in this debate for some years, the work of Granich et al. (2009) reignited interest.

The next two publications presented technical work. The first described and analysed algorithms that pair agents in microsimulations of STI epidemics, while the final article compared these algorithms in an actual microsimulation.

The relevance of mathematical or computer models of diseases can only be understood if the needs of society emanating from the disease are understood. The first three articles provide the background to this, culminating

in a detailed analysis of what was one of the most difficult questions of the HIV epidemic: when to start treatment. Other vital questions have also been asked of models: What are the costs and benefits of providing antiretrovirals? How will providing circumcision in public health facilities change HIV incidence? How effective must a microbicide or vaccine be to have an effect on incidence? How many orphans has the South African HIV epidemic produced? What role do sexual behaviours such as concurrent relationships or wide age-gap relationships have, if any, in the spread of HIV? Why have some countries had large HIV epidemics, while other comparatively small ones? How long will it take for a policy of universal treatment to virtually eliminate the epidemic? And many more.

How then do the final two technical articles help answer such questions?

First, the fundamental way in which models of STIs differ from other diseases is in partnership formation and dissolution. If a model assumes that people randomly match into partnerships it overestimates incidence, unless the model only accounts for a subset of individuals with similar sexual behaviour. The challenge of pair-matching is to account for heterogeneous sexual behaviour and, consequently, heterogeneous risk. Deterministic compartmental models account for heterogeneity by dividing the population into compartments, but this has practical limits; the number of compartments needs to be manageable and it is difficult to model the borders between compartments in which there is crossover pairing. Heterogeneity is easier to account for in agent-based models, and the pair-matching algorithm is key to this. The presentation and analysis of the pair-matching algorithms, we hope, adds to the understanding of this subject, and will make it easier for

modellers to implement microsimulations that better account for heterogeneity of sexual behaviour.

Second, we have found that as the agent population increases, incidence decreases, and that this is particularly apparent if the risk of infection is high. This finding is troubling and interesting. It occurs with all the pair-matching algorithms except random-pair matching, which means it occurs when heterogeneity is taken into account. We are unsure if it is an artefact of our methodology which is based on assigning a distance between agents to compare their suitability for matching.

If this phenomenon is not an algorithmic artefact — and there is no compelling reason to believe it is — then it appears to follow that setting the agent population to approximately the same size as the population being studied will yield more accurate estimates of incidence and prevalence. This is a hard task when it comes to large populations, even with highly optimised programming on high-end hardware. Of course, models are usually calibrated so that the prevalence they estimate matches multiple known data points. However, this in turn results in adjusting model parameters, e.g. infection rates, such that they may no longer match what is actually happening in the natural world STI. Nevertheless, the data, such as sex surveys, used to model STIs is quite possibly a greater source of error than this phenomenon.

Several research questions follow from our research:

- What is the cause of incidence decreasing as the agent population rises with the pair-matching algorithms presented here? The final article hypothesises an answer, but we have neither proven it, nor are we

confident of it. It is also unclear how to answer it. The number of agents we deal with, from 10,000 to 40 million, is so large that ways of visualising or manually analysing the phenomenon are not obvious.

- Is it possible to test whether in natural populations, incidence decreases for high-transmission rate STIs as population increases? There are so many variables involved in STI transmission that it is not obvious how to do this.

- If population size does affect STI incidence, can studying smaller subpopulations within a larger population produce more relevant results than studying the entire population. For example, in studying HIV incidence in the South African population, under what circumstances does it make sense for a model to consider only a subset of the population (e.g. men who have sex with men)?

- How do Blossom V, BFPM and CSPM affect the outcomes of microsimulations of actual, as opposed to fictitious, STIs? In particular, we are interested in testing our algorithms in microsimulations of HIV, gonorrhoea and HPV. The latter two are of interest because of the high risk of infection in serodiscordant partnerships.

- Can microsimulations that do pair-matching, perhaps using the algorithms described in this dissertation, be used to identify groups of people who should be prioritised for ART or other STI treatments?

In addition to these questions, there is a need to examine, analyse and compare other pair-matching methodologies that are not based on a distance

function with our algorithms. This will consolidate the theoretical understanding of pair-matching.

There are fascinating research questions that need to be considered to better understand and improve mathematical models of STIs. Answering them will improve our understanding of natural world STIs, and potentially improve our policy responses.

Bibliography

Abbasi, K., Bartlett, J., and Bangsberg, D. (2012). Investigating the Impact of Treatment on New HIV Infections. *PLoS Collections*.

Actuarial Society of South Africa (2011). ASSA2008 model. Accessed: 3/11/2012.

Asimov, I. (1991). *Foundation*. Bantam Spectra Books, New York, revised edition edition.

Bacaër, N., Pretorius, C., and Auvert, B. (2010). An Age-Structured Model for the Potential Impact of Generalized Access to Antiretrovirals on the South African HIV Epidemic. *Bulletin of Mathematical Biology*, 72(8):2180–2198.

Birnbaum, J. K., Murray, C. J., and Lozano, R. (2011). Exposing misclassified HIV/AIDS deaths in South Africa. *Bulletin of the World Health Organization*, 89(4):278–285.

Boulle, A., Kenyon, C., and Abdullah, F. (2003). A review of antiretroviral costing models in South Africa. In *Economics of AIDS and Access to*

HIV/AIDS Care in Developing Countries: Issues and Challenges, pages 293–309. ANRS, Paris.

Butler, D. (2014). Models overestimate Ebola cases. *Nature News*, 515(7525):18.

Cambiano, V., Bertagnolio, S., Jordan, M. R., Lundgren, J. D., and Phillips, A. (2013). Transmission of Drug Resistant HIV and Its Potential Impact on Mortality and Treatment Outcomes in Resource-Limited Settings. *The Journal of Infectious Diseases*, 207(suppl_2):S57–S62.

Cambiano, V., Bertagnolio, S., Jordan, M. R., Pillay, D., Perriëns, J. H., Venter, F., Lundgren, J., and Phillips, A. (2014). Predicted levels of HIV drug resistance: potential impact of expanding diagnosis, retention, and eligibility criteria for antiretroviral therapy initiation. *AIDS (London, England)*, 28 Suppl 1:S15–23.

Collins, S. and Geffen, N. (2014). Community views: balancing the public health benefits of earlier antiretroviral treatment with the implications for individual patients - perspectives from the community. *Current opinion in HIV and AIDS*, 9(1):4–10.

Cook, W. and Rohe, A. (1999). Computing Minimum-Weight Perfect Matchings. *INFORMS J. on Computing*, 11(2):138–148.

Cori, A., Ayles, H., Beyers, N., Schaap, A., Floyd, S., Sabapathy, K., Eaton, J. W., Hauck, K., Smith, P., Griffith, S., Moore, A., Donnell, D., Vermund, S. H., Fidler, S., Hayes, R., Fraser, C., and HPTN 071 PopART

Study Team (2014). HPTN 071 (PopART): a cluster-randomized trial of the population impact of an HIV combination prevention intervention including universal testing and treatment: mathematical model. *PloS One*, 9(1):e84511.

Dorrington, R. E. (1998). Assa600: An aids model of the third kind? *Transactions of the Actuarial Society of South Africa*, 13(1):99–153.

Dorrington, R. E. (2002). Expert affidavit in TAC Competition Commission complaint.

Dorrington, R. E., Bourne, D., Bradshaw, D., Laubscher, R., and Timaeus, I. M. (2001). The Impact of HIV/AIDS on Adult Mortality in South Africa. Technical report, Medical Research Council.

Doyle, P. (1993). The Demographic Impact of AIDS on the South African Population. In Cross, S. and Whiteside, A., editors, *Facing up to AIDS*, pages 87–112. Palgrave Macmillan UK. DOI: 10.1007/978-1-349-22597-2_5.

Doyle, P., Broomberg, J., Steinberg, M., Masobe, P., and Behr, G. (1991). AIDS in South Africa: the demographic and economic implications. Technical Report 23, Centre for Health Policy, University of the Witwatersrand.

Eaton, J. W., Bacaër, N., Bershteyn, A., Cambiano, V., Cori, A., Dorrington, R. E., Fraser, C., Gopalappa, C., Hontelez, J. A. C., Johnson, L. F., Klein, D. J., Phillips, A. N., Pretorius, C., Stover, J., Rehle, T. M., and Hallett, T. B. (2015). 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. *The Lancet Global Health*, 3(10):e598–e608.

Eaton, J. W. and Hallett, T. B. (2014). Why the proportion of transmission during early-stage HIV infection does not predict the long-term impact of treatment on HIV incidence. *Proceedings of the National Academy of Sciences*, 111(45):16202–16207.

Eaton, J. W., Johnson, L. F., Salomon, J. A., Bärnighausen, T., Bendavid, E., Bershteyn, A., Bloom, D. E., Cambiano, V., Fraser, C., Hontelez, J. A. C., Humair, S., Klein, D. J., Long, E. F., Phillips, A. N., Pretorius, C., Stover, J., Wenger, E. A., Williams, B. G., and Hallett, T. B. (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.

Edmonds, J. (1965). Maximum matching and a polyhedron with o,1-vertices. *Journal of Research of the Na tional Bureau of Standards-B. Mathematics and Mathematical Physics*, 69B(1 and 2):125–130.

Geffen, N. (2001). Cost and cost-effectiveness of mother-to-child transmission prevention of hiv. Briefing paper.

Geffen, N. (2013). Squinting into the future. *NSP Review*, 1(5):22–25.

Geffen, N. (2015). Anything to Stay Alive: The Challenges of a Campaign for an Experimental Drug. *Developing World Bioethics*.

Geffen, N. (2017). The role of activists in access to HIV and tuberculosis

treatment and prevention. *Current opinion in HIV and AIDS*, 12(4):398–402.

Geffen, N., Aagaard, P., Corbelli, G. M., Meulbroek, M., Peavy, D., Rappoport, C., Schwarze, S., Collins, S., and International Network for Strategic Initiatives in Global HIV Trials (INSIGHT) Community Advisory Board (2015). Community perspective on the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial. *HIV medicine*, 16 Suppl 1:10–13.

Geffen, N. and Low, M. (2017). When to start antiretroviral treatment? a history and analysis of a scientific controversy. *Southern African Journal of HIV Medicine*, 18(1):8.

Geffen, N., Raubenheimer, C., and Nattrass, N. (2003). The cost of hiv prevention and treatment interventions in south africa. Technical Report 28, Centre for Social Science Research.

Geffen, N. and Scholz, S. (2017). Efficient and effective pair-matching algorithms for agent-based models. *Journal of Artificial Societies and Social Simulation*, 20(4):8.

Geffen, N. and Welte, A. (2018). Modelling the human immunodeficiency virus (hiv) epidemic: A review of the substance and role of models in south africa. *Southern African Journal of HIV Medicine*, 19(1):15.

Gottlieb, M., Schanker, H., Fan, P., Saxon, A., and Weisman, J. (1981). Pneumocystis pneumonia — los angeles. *MMWR Weekly*, 30(21):1–3.

Granich, R. M., Gilks, C. F., Dye, C., De Cock, K. M., and Williams, B. G. (2009). Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *The Lancet*, 373(9657):48–57.

Heard, D., Dent, G., Schifeling, T., and Banks, D. (2015). Agent-Based Models and Microsimulation. *Annual Review of Statistics and Its Application*, 2(1):259–272.

Hontelez, J. A. C., Lurie, M. N., Bärnighausen, T., Bakker, R., Baltussen, R., Tanser, F., Hallett, T. B., Newell, M.-L., and de Vlas, S. J. (2013). Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study. *PLoS medicine*, 10(10):e1001534.

JAS (2017). ABM and Microsimulations. Web page.

Johnson, L. (2004). An introduction to the mathematics of HIV/AIDS modelling.

Johnson, L. (2014). THEMBISA version 1.0: A model for evaluating the impact of HIV/AIDS in South Africa.

Johnson, L., Dorrington, R., Bradshaw, D., Pillay-Van Wyk, V., and Rehle, T. (2009). Sexual behaviour patterns in South Africa and their association with the spread of HIV: insights from a mathematical model. *Demographic Research*, 21(11):289–340.

Johnson, L. F. (2012). Access to antiretroviral treatment in South Africa, 2004 - 2011. *Southern African Journal of HIV Medicine*, 13(1):22–27.

Johnson, L. F., Davies, M.-A., Moultrie, H., Sherman, G. G., Bland, R. M., Rehle, T. M., Dorrington, R. E., and Newell, M.-L. (2012). The effect of early initiation of antiretroviral treatment in infants on pediatric AIDS mortality in South Africa: a model-based analysis. *The Pediatric Infectious Disease Journal*, 31(5):474–480.

Johnson, L. F., Dorrington, R. E., and Moolla, H. (2016). Modelling the impact of HIV in South Africa's provinces.

Johnson, L. F. and Geffen, N. (2016). A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. *Sexually Transmitted Diseases*, 43(3):139–146.

Klein, D. (2012). Relationship formation and flow control algorithms for generating age-structured networks in HIV modeling. In *2012 IEEE 51st Annual Conference on Decision and Control (CDC)*, pages 1041–1046.

Klein, D. J., Bershteyn, A., and Eckhoff, P. A. (2014). Dropout and re-enrollment: implications for epidemiological projections of treatment programs. *AIDS (London, England)*, 28 Suppl 1:S47–59.

Kolmogorov, V. (2009). Blossom V: a new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation*, 1(1):43–67.

Nattrass, N. (2001). Affidavit in Treatment Action Campaign and Others versus Minister of Health and Others.

Nattrass, N. and Geffen, N. (2005). The impact of reduced drug prices on

the cost-effectiveness of HAART in South Africa. *African Journal of AIDS Research*, 4(1):65–67.

Network, H. P. T. (2016). Population effects of antiretroviral therapy to reduce hiv transmission (popart) (popart).

Orcutt, G. H. (1957). A New Type of Socio-Economic System. *The Review of Economics and Statistics*, 39(2):116–123.

Phillips, A. N., Pillay, D., Garnett, G., Bennett, D., Vitoria, M., Cambiano, V., and Lundgren, J. (2011). Effect on transmission of HIV-1 resistance of timing of implementation of viral load monitoring to determine switches from first to second-line antiretroviral regimens in resource-limited settings. *AIDS (London, England)*, 25(6):843–850. PMID: 21192233.

Rivers, C. (2014). Ebola: models do more than forecast. *Nature*, 515(7528):492–492.

Simbayi, L., Shisana, O., Rehle, T., Onoya, D., Jooste, S., Zunga, N., Labadarios, D., and Zuma, K. (2014). South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. Technical report, HSRC.

Statistics South Africa (2002). Causes of death in south africa 1997-2001: Advance release of recorded causes of death. P0309.2, Statistics South Africa, Pretoria.

UNAIDS (2016). Aids by the numbers.

Williams, B. (2014). Elimination of HIV in South Africa through expanded

access to antiretroviral therapy: Cautions, caveats and the importance of parsimony. *arXiv:1403.7104 [q-bio]*. arXiv: 1403.7104.