© Revista de Matemática: Teoría y Aplicaciones 2020 **27**(1) : 93–121
cimpa – ucr   issn: 1409-2433 (Print), 2215-3373 (Online)
doi: `https://doi.org/10.15517/rmta.v27i1.39951`

FEASIBLE AND ETHICAL ALLOCATION OF INTERVENTION RESOURCES FOR INFECTIOUS DISEASES USING LINEAR PROGRAMMING

ASIGNACIÓN FACTIBLE ÉTICA DE RECURSOS DE INTERVENCIÓN PARA ENFERMEDADES INFECCIOSAS MEDIANTE PROGRAMACIÓN LINEAL

David J. Gerberry[*]   Sally Blower[†]

Received: 19/Jul/2019; Revised: 17/Sep/2019;
Accepted: 31/Oct/2019

---

[*] Xavier University, Department of Mathematics, Cincinnati, Ohio, United States. E-Mail: david.gerberry@xavier.edu

[†] University of California, David Geffen School of Medicine, Los Angeles, California, United States. E-Mail: sblower@mednet.ucla.edu

Abstract

In this work, we demonstrate that the consideration of a fixed epidemic and the use of linear programming can be an effective tool for designing rollout strategies for infectious disease interventions. Specifically, we argue that the approach can be more flexible, more amenable to detailed allocation plans and more in line with the way that public policy decisions are made than standard optimal control allocations. We also show how feasibility and ethical constraints can be incorporated into resource allocations.

As an application, we consider the initial rollout of Treatment as Prevention (TasP) resources for HIV (human immunodeficiency virus) in South Africa that began within the last decade. Going back to TasP's initial rollout allows us to demonstrate the strengths of this approach.

Keywords: mathematical model; infectious disease; resource allocation; linear programming; HIV; treatment as prevention; South Africa.

Resumen

En este trabajo, demostramos que la consideración de una epidemia fija y el uso de la programación lineal puede ser una herramienta efectiva para diseñar estrategias de lanzamiento para intervenciones de enfermedades infecciosas. Específicamente, argumentamos que el enfoque puede ser más flexible, más susceptible a planes de asignación detallados y más en línea con la forma en que se toman las decisiones de política pública que las asignaciones de control óptimo estándar. También, mostramos cómo la viabilidad y las restricciones éticas pueden incorporarse en las asignaciones de recursos.

Como aplicación, consideramos la implementación inicial de los recursos de Tratamiento como Prevención (TasP) para el VIH (virus de inmunodeficiencia humana) en Sudáfrica que comenzó en la última década. Volver al lanzamiento inicial de TasP nos permite demostrar las fortalezas de este enfoque.

Palabras clave: modelo matemático; enfermedad infecciosa; asignación de recursos; programación lineal; VIH; tratamiento como prevención; Sudáfrica.

Mathematics Subject Classification: 92B05, 92D30, 90C05.

1 Introduction

For many public health policymakers, mathematical modeling is never more important than at times when new interventions for infectious diseases are discovered. Modelers are relied up to answer important questions like, “Extrapolating the $x\%$ efficacy of the clinical trial, what would the population-level impact be?” and “Given y million dollars, how should resources be allocated to prevent the most infections?”

A major contribution of mathematical modeling in addressing such questions lies in the ability to predict the future consequences of events. For example, preventing 100 infections today results in fewer infectious people later which means even more infections will be prevented in the future. Without mathematical modeling, quantifying such feedback (positive feedback in this example) is nearly impossible. Including these feedbacks into the problem of optimally allocating intervention resources relies on the theory of optimal control. In an optimal control model, a dynamic model (e.g. a disease transmission model) is expanded to include a time-dependent control (e.g. in our case, time-dependent resource allocation). Using such a system allows for the control (i.e. resource allocation in our context) to evolve over time as the intervention affects the dynamics of infection.

From a purely mathematical perspective, optimal control theory is the perfect tool to address such resource allocation questions. The dynamic model captures the essence of disease transmission and the solution gives the optimal control function that prevents the absolute maximum number infections over the given time window. Unfortunately, public policy decisions are typically more nuanced than their mathematical formulations. Specifically, feasibility and ethical concerns may render a mathematically optimal resource allocation impractical at best and untenable at worst for a variety of reasons.

In this work, we detail an approach for planning the initial phase of disease interventions that utilizes linear programming to determine optimal resource allocation based on the consideration of a fixed epidemic. We use the phrase “fixed epidemic” to reflect the fact that the allocation strategy does not incorporate future epidemiological dynamics, but rather assumes a time window short enough that dynamic effects would be minimal. This assumption does not mean the disease is necessarily at an endemic equilibrium. Such an approach is not for all diseases and interventions. However, when appropriate the framework provides detailed, flexible and practical allocations that are in line with the decision-making process. In particular, we highlight a few characteristics of diseases and interventions for which this new approach is particularly well-suited.

- *Diseases with slow dynamics.* The major limitation of this allocation framework is the assumption of a fixed epidemic. This simplification is appropriate when considering a time window of the intervention that is small relative to the speed of disease dynamics, (e.g. if one was considering the first year rollout of a tuberculosis (TB) intervention given that epidemiological dynamics of TB occur on the order of decades). Other examples of diseases with slow dynamics include HIV and hepatitis. For diseases such as influenza and measles with much faster dynamics, the fixed epidemic simplification is unlikely to be appropriate.

- *Availability of data.* A significant strength of the approach proposed in this work is its ability to create detailed allocation plans. This virtue can only be fully leveraged if data estimates are available to inform the detailed linear programming. If a disease is spread homogeneously through a population (i.e. without variation in groups, ages, etc.) a simple model and optimal control formulation may be viable. When transmission and infection risk are highly structured (with reliable data estimates), the flexibility of the linear programming framework becomes crucial.

- *Feasibility, ethical and practicality concerns.* While mathematical optimization can find allocation plans that prevent the maximum number of infections over a given time period, it does not always produce a plan that is in line with sound public health policy. For example, optimal control investigations of infectious disease often result in "bang-bang" controls where an entire resource budget is spent as quickly as possible in order to get the maximum feedback effects over the time period. However, running out of resources is rarely a tenable policy plan and can be especially dangerous when drug resistance is a concern.

  Allocations that simply prevent the most infections can be impractical for other reasons. For instance, preventing the most infections often dictates prioritizing certain population groups perhaps by gender, age, sexual identity, geographic region, race, etc., in ways that may not be politically feasible or even ethical. In such situations, the flexibility of the linear programming framework allows for feasibility constraints to be easily added and evaluated for effectiveness.

While the use of mathematical optimization techniques to allocate limited HIV resources is not new in the field of health economics [11, 8, 13, 12, 4], our approach is unique in emphasizing the capacity of the fixed epidemic assumption (when appropriate) to allow for flexible and robust allocation analyses.

In this paper, we will use the rollout of Treatment as Prevention (TasP) resources for HIV in South Africa as a case study of our resource allocation technique. Doing so will highlight aspects of epidemics that are suitable to this approach and the robustness of the approach to answering important allocation questions.

In Section 2, we describe the Treatment as Prevention intervention for HIV. In Section 3, we use the Actuarial Society of South Africa (ASSA) HIV/AIDS model to examine South Africa's HIV epidemic at the time of TasP rollout. In Section 4, we use ASSA data to implement a linear program for allocating TasP resources. In Section 5, we examine the optimal allocation plans produced by the linear program, implement multiple feasibility and ethical constraints and compare the effectiveness of multiple TasP allocation strategies. In Section 6, we return to a general discussion of the resource allocation paradigm with the concrete example of the TasP case study.

2 Treatment as prevention for HIV

Treatment as Prevention (TasP) is the strategy of expanding antiretroviral treatment to early-stage HIV-infected individuals in order to prevent future transmission. The idea behind the approach is simple. When an HIV-infected individual is put on antiretroviral treatment (ART) their viral load decreases significantly, frequently below detectable limits [5]. This reduction in viral load not only improves the health and prognosis of the patient but also reduces their level of infectiousness and hence reduces the likelihood of them infecting others [5].

While not the first to address the idea of TasP, a mathematical modeling study proved to be a major impetus towards the global community's awareness of the strategy. Appearing in Granich et al. [9] used a relatively simple compartmental model to show that the HIV epidemic could be ended within decades if all individuals were tested and infected individuals were soon after put on ART.

Soon after the modeling work, clinical trial results provided more evidence for optimism regarding TasP. In 2010, the iPrEx clinical trial [10] showed a 92% reduction in transmission among men who have sex with men for those with a detectable level of drug-level (i.e. those who took ART as directed). In 2011, the HPTN 052 clinical trial [5] showed that ART reduced HIV transmission by 96% in stable serodiscordant couples (i.e. one partner HIV$+$ and one not).

With such a huge protective effect, the public health community and government officials moved toward designing plans to rollout TasP. Of course, doing so presented significant challenges including cost and feasibility.

As expensive HIV medications would be taken daily and indefinitely, it is clear that resource constraints would be a significant consideration. Moreover, adherence issues and the large-scale infusion of antiretroviral drugs could lead to drug resistance issues that render HIV much more complicated and expensive to treat.

In what follows, we place ourselves in the position of a health official in South Africa circa 2012-15 designing a TasP rollout plan. We focus primarily on the first challenge above, specifically:

> What is the best way to allocate a fixed quantity of TasP resources?

3 Actuarial Society of South Africa HIV/AIDS model

One of the strengths of the linear programming approach to allocating resources is its robustness to detail. To illustrate this, we build a TasP allocation plan around the extremely intricate Actuarial Society of South Africa (ASSA 2008) HIV/AIDS model.

The Actuarial Society of South Africa HIV/AIDS model is a demographic model of HIV in South Africa. Rather than being formulated in the language of differential equations that mathematical biologists are accustomed, the model is implemented in a spreadsheet form that is available for download at [1] with a user guide at [6]. The model is calibrated to HIV prevalence data at antenatal clinics and produces a detailed estimate of the HIV epidemic in South Africa. It is worth noting that what modeling papers often refer to as "data" (e.g. country-specific HIV prevalence estimate in Joint United Nations Programme on HIV/AIDS (UNAIDS) or World Health Organization (WHO) factsheets) are in fact only estimates of prevalence and are outputs of these types of actuarial models. For South Africa in particular, published national prevalence estimates came from the ASSA model. Though referred to as "ASSA 2008", the model was last updated and released in 2011 and was consequently state of the art as of 2012 when TasP rollout began.

The ASSA model stratifies the population of South Africa according to age, gender, race, sexual risk behavior, HIV stage of infection and province residence. The ASSA model does not include homosexual transmission of HIV. To more clearly illustrate our results, we consider South Africa as a whole and ignore province residence. We also ignore race as we believe that allocating resources based on race would be both infeasible and unethical.

3.1 Assumptions of ASSA model

The ASSA model divides the population into the following four sexual risk groups:

- **NOT:** Individuals who are not at risk of HIV infection. This includes individuals who are not sexually active, in a stable committed partnership with an uninfected partner, or do not engage in unprotected sex.
- **RSK:** Individuals with a relatively low level of sexual activity, but who are still at risk from HIV in that they have, on average, one new partner per year and do engage in unprotected sex.
- **STD:** Individuals whose level of sexual activity and infrequency of using protection is such that they are regularly infected with sexually transmitted diseases (STDs).
- **PRO:** Commercial sex workers and individuals whose level of sexual activity and infrequency of using protection is similar to that of commercial sex workers (e.g. clients of commercial sex workers).

Individuals below the age of 14 and above the age of 60 are assumed to not be sexually active. The initial distribution of the adult population into the four risk groups is shown in Table 2.

Individuals infected with HIV are stratified by the now-antiquated WHO Clinical Staging System for HIV, which include: **Stage 1:** Asymptomatic, **Stage 2:** Mild symptoms, **Stage 3:** Advanced symptoms, **Stage 4:** Severe symptoms, **Stage 5:** Receiving anti-retroviral treatment and **Stage 6:** Discontinued anti-retroviral treatment. Today, HIV progression is instead described by the number of CD4 immune cells per microliter of blood (i.e. CD4 count). A rough equivalence between the two is that Stage 1 is CD4 > 500, Stage 2 is $350 <$ CD4 < 499, Stage 3 is $200 <$ CD4 < 349 and Stage 4 is CD4 < 200 [16].

In modeling sexual transmitted diseases of humans, capturing reliable sexual behavior data is always a challenge. Even when data can be found, incorporating the nuance of such behavior into a compartment ordinary differential equation (ODE) or partial differential equation (PDE) model can be infeasible. Given its spreadsheet nature and detailed breakdown of the population, the ASSA model is able to make precise estimates of sexual mixing and activity levels that are not possible in many other structures.

In general, the underlying philosophy of the ASSA model is to make assumptions regarding the sexual behavior of females (e.g. initiation of sexual activity, sexual active levels, etc.) and the mixing patterns between females and males. The sexual behavior of males then follows as a consequence of these assumptions.

The sexual behavior component of the ASSA model is extensive and detailed, including, but not limited to, sexual activities levels, condom usage, partnership formation, transmission rates and increased susceptibility of young females to infection. Moreover, these quantities are largely specific to risk group, age, gender and disease stage. For clarity of presentation, we refer the reader to the ASSA model itself [1] the user guide at [6] and include graphics and tables in the Appendix to illustrate the major assumptions regarding sexual behavior.

3.2 State of the epidemic before TasP allocation

As described previously, the Actuarial Society of South Africa HIV/AIDS model provides a detailed approximation of the state of the HIV epidemic in South Africa. Before proceeding to the primary issue of resource allocation, it is valuable to assess our starting point. For this, we consider the epidemic at the start of 2015, a time when the expansion of treatment (i.e. TasP) began to ramp up in earnest.

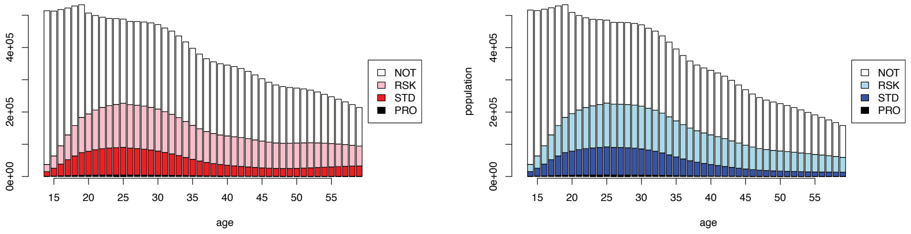
As the ASSA model was last updated in 2011, we first verify that it provides a reasonable estimate of HIV in South Africa in 2015. In Table 1, we see it produces estimates of overall HIV prevalence and the number of people living with HIV (LWH) that are comparable to estimates from other sources.

Table 1: Comparison of HIV estimates for HIV prevalence and the number of people living with HIV (LWH) from ASSA model and other sources.

Statistic	ASSA	Comparison	Source	Reference
HIV prevalence (15-49)	16.7%	19.0%	UNAIDS	[15]
HIV prevalence (15-49)	16.7%	18.8%	HSRC survey	[14]
Women LWH (15 & up)	3,285,754	3,500,000	UNAIDS	[15]
Adults LWH (15 & up)	5,425,066	5,900,000	UNAIDS	[15]

Moving to the detailed results of the ASSA model, we present the structure of the population of South Africa in terms of age and sexual risk group in Figure 1.

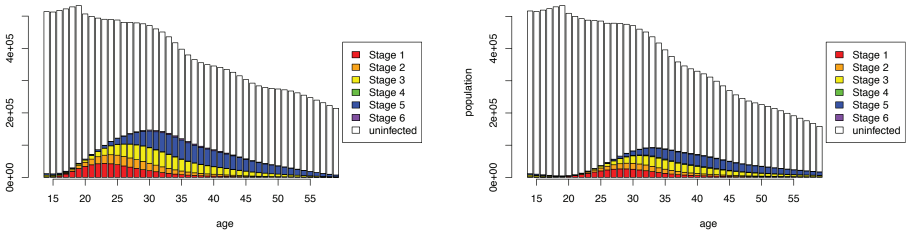
The starting point of our resource allocation in 2015 is a result of the initial distribution assumptions of the population into sexual risk groups in Table 2 and the model dynamics of infection, mortality, etc. The structure of the HIV epidemic in 2015 is shown in Figure 2.



(a) Female **(b)** Male

Figure 1: Risk group structure of the 2015 population of South Africa from the ASSA model. Across all ages, this results in 11042261, 4357205, 2172812 and 103128 females in the NOT, RSK, STD and PRO risk groups, respectively. For males, there are 10434387, 4161919, 2086901 and 103505 individuals in the NOT, RSK, STD and PRO risk groups, respectively.

Again, we emphasize that the WHO Clinical Staging System is no longer in use having been replaced by CD4 count levels for disease staging. Nevertheless, Figure 2 provides valuable insight into HIV in South Africa. Notably, we see that many more females are infected with HIV than males, a situation that is true of all HIV epidemics where transmission is mainly through heterosexual intercourse. In addition, we see that females tend to be infected with HIV at younger ages than males. For comparison, we see that the prevalence of HIV in females aged 20-25 is drastically larger than that of males aged 20-25.



(a) Female **(b)** Male

Figure 2: Population of South Africa as of 2015, structured by age and HIV stage from the ASSA model.

4 Linear program model for resource allocation

In this section, we describe the process of building a linear program model for allocating TasP resources in South Africa. Upon doing so, we can find optimal allocations based on a variety of resource constraints.

4.1 Reversing transmission calculations of ASSA model

ASSA 2008 is a demographic model of the epidemiology of HIV in South Africa. The transmission process is modeled by calculating the probability of a *susceptible* individual in each population subgroup getting infected in a year. Therefore, the ASSA model does not explicitly model an individual's infectious potential. Since TasP is given to HIV-infected individuals, we must calculate the transmission potential of an *infected* individual in each population subgroup in order to determine optimal resource allocations.

To illustrate this difference, consider the following example. ASSA calculations give us the probability that a 22 year-old female in the STD risk group gets infected in a given year. What our allocation program requires instead is the expected number of new infections that would be caused in a year by an infected 22 year-old female in the STD risk group. Going into more detail, the ASSA model calculates

$$
\begin{aligned}
\text{ASSA:}\quad & \Pr[\text{22 year-old female in STD risk group gets infected}] && (1)\\
&= 1 - \Pr[\text{she doesn't get infected}] && (2)\\
&= 1 - \prod_{\text{partnerships}} \Pr[\text{not infected in a particular partnership}] && (3)\\
&= 1 - \prod_{\text{partnerships}} (1 - \Pr[\text{infection per sex act}])^{\text{\# acts per partnership}}. && (4)
\end{aligned}
$$

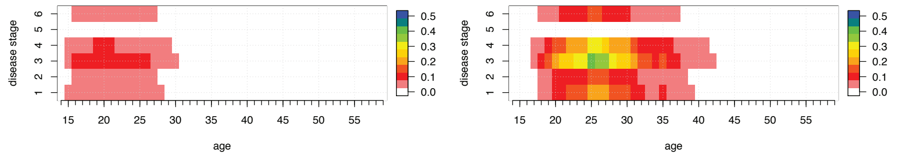
To reverse the transmission calculations to get the infectious potential of individuals, we note that

$$
\begin{aligned}
& \mathbb{E}[\text{new infections caused by 22 year-old female in STD risk group}] && (5)\\
&= \sum_{\text{partnerships}} \Pr[\text{she infects a particular partner}] && (6)\\
&= \sum_{\text{partnerships}} [1 - \Pr[\text{she doesn't infect a particular partner}]] && (7)\\
&= \sum_{\text{partnerships}} \left[1 - (1 - \Pr[\text{infection per sex act}])^{\text{\# acts per partnership}}\right], && (8)
\end{aligned}
$$

where the per act transmission probability, number of sexual contacts per partnership and number of partnerships comes from the ASSA model. More specifically, these values are actually averages over all combinations of partner ages, risk group, infection status, condom usage, etc., weighted according to assumed sexual mixing patterns (for more details, see theASSA model itself [1] and user guide [6]). Our reversed calculation is not exact in the sense that it overcounts new infections in the situation where a susceptible individual could have been infected by multiple individuals in a given year. As such situations are exceeding rare, this approximation causes no significant differences. Simulating the dynamics of the ASSA model and those of the "reversed" calculations resulted in new infection totals that were within $\pm 1\%$.

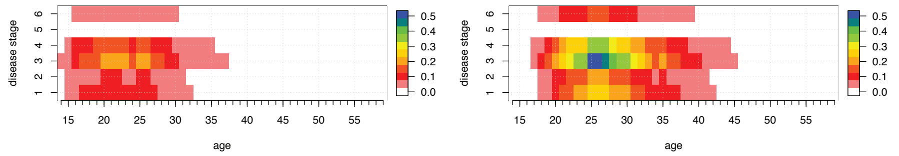
4.2 Infectious potential of individuals

Using the procedure described above, we calculate the transmission potential of an individual given their gender, sexual risk group, age and HIV stage. We denote this expected number of new infections caused in a year as $I_{i,j,k,l}$, where i, j, k and l denote the individual's gender, sexual risk group, age and HIV stage, respectively. The results of these calculations are presented in Figure 3.



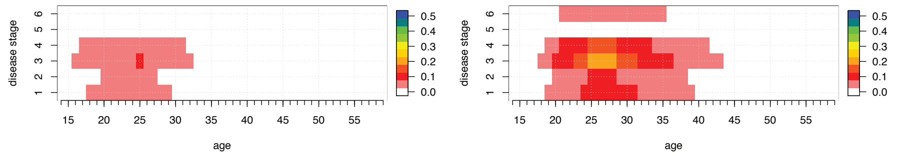
(a) Females in PRO risk group

(b) Males in PRO risk group



(c) Females in STD risk group

(d) Males in STD risk group



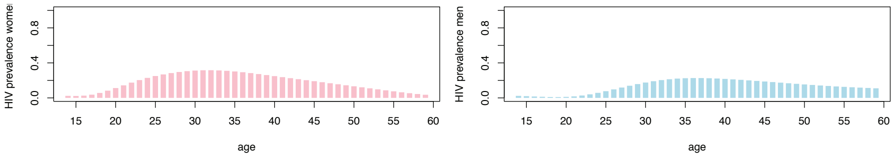
(e) Females in RSK risk group

(f) Males in RSK risk group

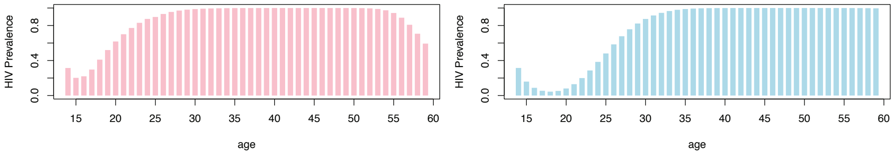
Figure 3: Per capita expected number of new infections per year caused by individuals in each population subgroup stratified by age and disease stage.

As the results in Figure 3 drive our resource allocations, observations are warranted. We note that HIV Stage 5 shows minimal infectious potential (i.e. white horizontal bands in each graph). As Stage 5 is composed of individuals receiving antiretroviral treatment, we see that the ASSA model already reflects that treatment drastically reduces one's infectiousness.

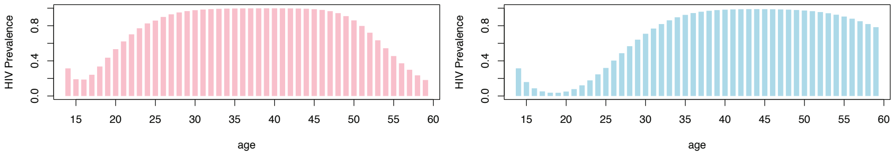
Comparing the left and right columns, we see that males have a significantly higher infectious potential. This results from the well-established fact that STDs in general are more transmissible from males to females as opposed to from females to male due to the insertive/receptive nature of heterosexual intercourse. Consequently it is true that HIV prevalence is higher in females than males in epidemics driven by heterosexual transmission (see Figures 4a and 4b). ASSA estimates HIV prevalences of 18% and 13.7% in female and males, respectively, in South Africa.



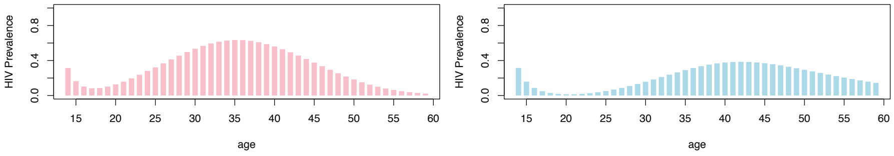
(a) Females in general population **(b)** Males in general population



(c) Females in PRO risk group **(d)** Males in PRO risk group



(e) Females in STD risk group **(f)** Males in STD risk group



(g) Females in RSK risk group **(h)** Males in RSK risk group

Figure 4: HIV prevalence in each population subgroup by age.

In Figure 3, we see that the ASSA model suggests that members of the STD sexual risk group have a higher infectious potential than those in the PRO risk group. This unexpected situation results from the fact that members of the PRO risk group interact primarily with members in the PRO risk group (75% of contacts, see Table 3) and that ASSA projects very high prevalence rates in the PRO groups (see Figures 4c and 4d). Thus, an infected member of the PRO risk group is primarily having contacts with others that are already infected. The accuracy of these extremely high prevalences in high-risk sexual groups (often referred to as "core groups" in the literature) in the actual population is debatable. However, obtaining accurate information regarding sexual behavior at the population level and stratifying that behavior into discrete sexual risk groups is a major limitation of many modeling efforts of HIV.

4.3 Resource allocation using linear programming

The values of the expected number of new infections caused in a year $I_{i,j,k,l}$ (illustrated in Figure 3) form the basis for our resource allocation. Quite simply, the optimal allocation of a fixed quantity of TasP resources is to prioritize the population subgroups with the highest infectious potential (i.e. $I_{i,j,k,l}$). While not necessary to understand the allocation plan, linear programming allows us to formulate the allocation succinctly and, more importantly, provides a robust framework for incorporating additional constraints.

To formalize our allocation problem, we assume a resource constraint R. To avoid the intricacies of pricing pharmaceuticals (price negotiations, discounts of scale, etc.), we will assume that R represents the number of new individuals that resources will allow to be put on treatment as part of the Treatment as Prevention (TasP) strategy rather than specific dollar amounts. Individuals in Stage 5 are already on treatment and are therefore assumed to remain so without requiring intervention resources (i.e. their continued treatment does not come from the intervention constraint, R). Since it is infeasible to get 100% of any population subgroup on treatment, we assume a maximum coverage of 90% in each subgroup. While it will not change the allocation plan, we assume that treatment reduces transmission by 95% to get estimates of the number of infections that could be prevented. Our linear program simply maximizes the number of infections prevented subject to the given resource constraints:

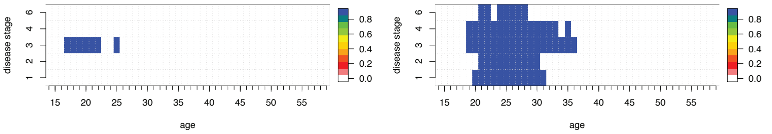
$$\begin{aligned} \max \quad & \sum_{i,j,k,l} (.95)\, P_{i,j,k,l}\, I_{i,j,k,l}\, X_{i,j,k,l}, \\ \text{s.t.} \quad & \sum_{i,j,k,l} P_{i,j,k,l}\, X_{i,j,k,l} \leq R, \\ & 0 \leq X_{i,j,k,l} \leq 0.90. \end{aligned} \tag{9}$$

Here, $P_{i,j,k,l}$ denotes the number of individuals in each subgroup (from ASSA), $I_{i,j,k,l}$ is the expected number of new infections caused per year per member of a subgroup (calculated from ASSA in Section 4.1) and $X_{i,j,k,l}$ is the TasP coverage in each subgroup (the output of the linear program, i.e. optimal allocation). Indices i, j, k, l are used to identify population subgroups where $i \in \{\text{male, female}\}, j \in \{\text{PRO, STD, RSK}\}, k \in \{15, ..., 49\}$, and $l \in \{\text{Stage 1,..., Stage 6}\}$.

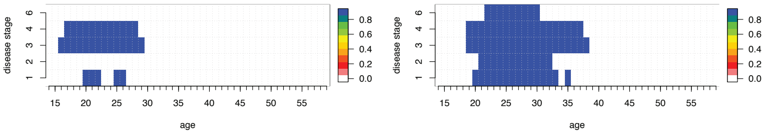
5 Results

5.1 Optimal allocation

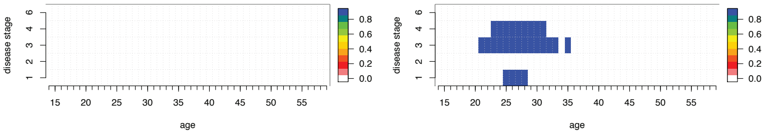
Once formalized in linear program (LP) (9), the optimal allocation of TasP resources is found by solving the linear program. Again, here we define optimal as maximizing the number of infections prevented given the fixed resources and define the allocation itself as the TasP coverage as specified in each group (i.e. the solution $X_{i,j,k,l}$ of LP (9)). The optimal allocations when resources are available to provide TasP to 25% and 10% of eligible individuals (i.e. infected and not already receiving ART) are illustrated in Figures 5 and 6, respectively.



(a) Females in PRO risk group **(b)** Males in PRO risk group

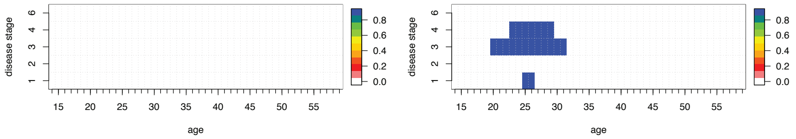


(c) Females in STD risk group **(d)** Males in STD risk group

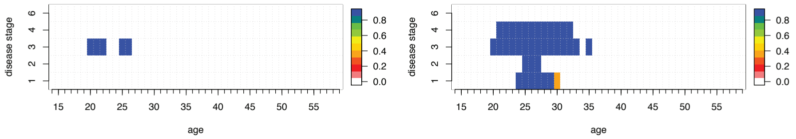


(e) Females in RSK risk group **(f)** Males in RSK risk group

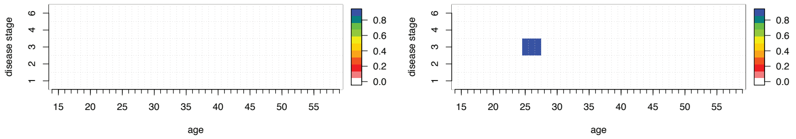
Figure 5: Allocation of TasP in each population subgroup stratified by age and disease stage resulting from solving LP (9) when resources are sufficient to cover 25% of eligible individuals.



(a) Females in PRO risk group **(b)** Males in PRO risk group

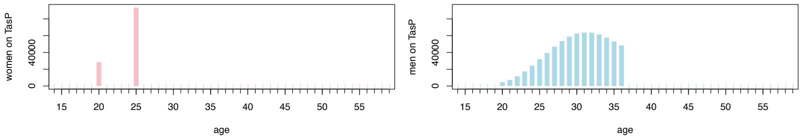


(c) Females in STD risk group **(d)** Males in STD risk group

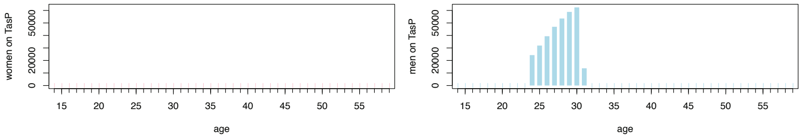


(e) Females in RSK risk group **(f)** Males in RSK risk group

Figure 6: Allocation of TasP in each population subgroup stratified by age and disease stage resulting from solving LP (9) when resources are sufficient to cover 10% of eligible individuals.



(a) Females on TasP when resources cover 25% of eligible individuals. **(b)** Males on TasP when resources cover 25% of eligible individuals.



(c) Females on TasP when resources cover 10% of eligible individuals. **(d)** Males on TasP when resources cover 10% of eligible individuals.

Figure 7: Age-stratified numbers of females and males put on TasP by allocation resulting from solving LP (9) for scenarios when resources are sufficient to cover 25% and 10% of eligible individuals.

As mentioned previously, the linear programming allocation essentially prioritizes population subgroups by their infectious potential. The full allocation (i.e. 90% maximum coverage) is then given to subgroups until resources run out. In both resource scenarios, Figures 5 and 6 show that more male subgroups are allocated TasP than female subgroups. Looking at the number of individuals on TasP (rather than TasP coverage) in Figure 7, we see the vast majority of TasP resources go to males. In fact, we see that *all* TasP resources go to men when resources are only available for 10% of eligible individuals. From an optimization perspective, this makes sense because male-to-female HIV transmission is nearly twice as likely as female-to-male transmission [7, 3]. Such an allocation plan of prioritizing men for TasP is consistent with modeling work of [2] which recommended TasP for males in the general population but not for women.

The optimal allocation when resources allow for 10% of eligible individuals to be put on TasP would prevent approximately 80,000 infections in the first year of rollout. If TasP were available for 25% of eligible individuals, the number of infections prevented rises to roughly 145,000.

5.2 Imposing additional constraints

When using more technical frameworks such as optimal control theory, it can be difficult for modeling works to go deeper than the initial optimal allocations of the previous section. In this section, we demonstrate the robustness of our simple linear programming approach to incorporating a variety of additional constraints.

5.2.1 Ethical constraints

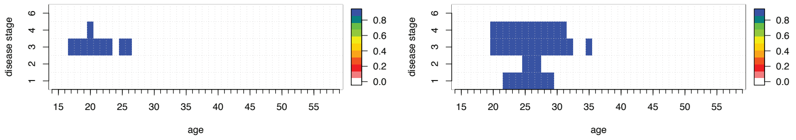
As mentioned in Section 5.1, our initial optimization resulted in a large imbalance of resources allocated to men as opposed to women. This imbalance makes sense when trying to maximize the number of infections prevented. However, there are clear ethical concerns in having such a gender imbalance.

If TasP resources go primarily to men (or entirely in the 10% resource scenario), it follows that new HIV infections prevented (IP) due to TasP would be primarily (or entirely) among women. Is it fair for the HIV prevention to go mostly to women? On the other hand, early initiation of antiretroviral treatment (as TasP is) does much more than reduce transmission potential. It is also associated with significant increases in life expectancy and improved quality of life in the infected individual. Is it fair that these positive effects of TasP would go primarily (or entirely) to men?

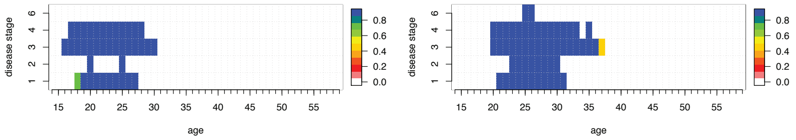
It is our view that answering these questions of health policy ethics are beyond the expertise of most mathematical modelers. Instead, modelers can offer information regarding the cost-effectiveness tradeoffs of imposing such ethical constraints into our resource allocation plan. Fortunately, our linear programming approach allows us to simply add an additional constraint requiring the number of males and females receiving TasP to be equal. The resulting linear program is now

$$\begin{aligned}
\max \quad & \sum_{i,j,k,l} (.95)\, P_{i,j,k,l}\, I_{i,j,k,l}\, X_{i,j,k,l}, \\
\text{s.t.} \quad & \sum_{i,j,k,l} P_{i,j,k,l}\, X_{i,j,k,l} \le R, \\
& \sum_{j,k,l} X_{\text{male},j,k,l}\, P_{\text{male},j,k,l} = \sum_{j,k,l} X_{\text{female},j,k,l}\, P_{\text{female},j,k,l}, \\
& 0 \le X_{i,j,k,l} \le 0.90.
\end{aligned} \tag{10}$$

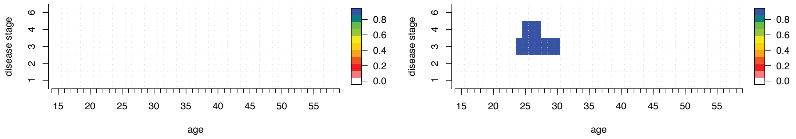
The optimal allocation, including the gender equity constraint (i.e. solution to LP 10) when resources are available to provide TasP to 25% and 10% of the infected individuals not already receiving ART are illustrated in Figures 8 and 9, respectively. The numbers of individuals on TasP by gender and age are presented in Figure 10.



(a) Females in PRO risk group **(b)** Males in PRO risk group

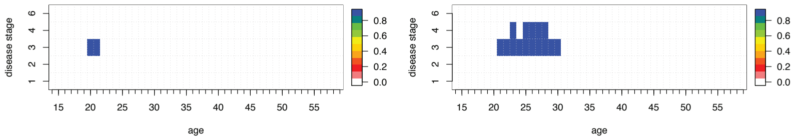


(c) Females in STD risk group **(d)** Males in STD risk group

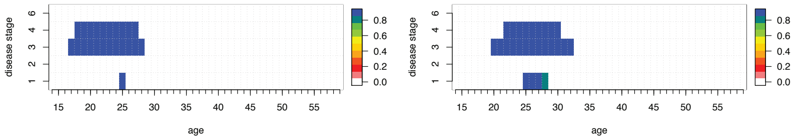


(e) Females in RSK risk group **(f)** Males in RSK risk group

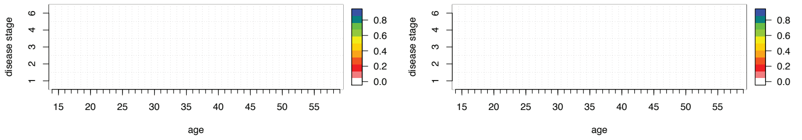
Figure 8: Allocation of TasP in each population subgroup stratified by age and disease stage, including gender equity constraint, resulting from solving LP (10) when resources are sufficient to cover 25% of eligible individuals.



(a) Females in PRO risk group **(b)** Males in PRO risk group

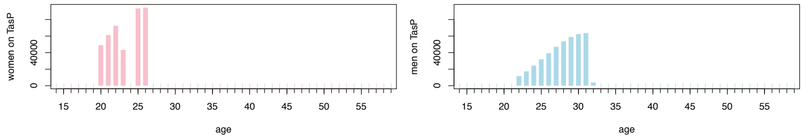


(c) Females in STD risk group **(d)** Males in STD risk group

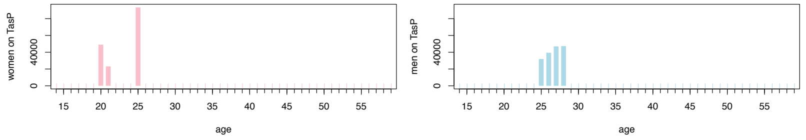


(e) Females in RSK risk group **(f)** Males in RSK risk group

Figure 9: Allocation of TasP in each population subgroup stratified by age and disease stage, including gender equity constraint, resulting from solving LP (10) when resources are sufficient to cover 10% of eligible individuals.



(a) Females on TasP when resources cover 25% of eligible individuals **(b)** Males on TasP when resources cover 25% of eligible individuals



(c) Females on TasP when resources cover 10% of eligible individuals **(d)** Males on TasP when resources cover 10% of eligible individuals

Figure 10: Age-stratified numbers of females and males put on TasP by allocation with gender equity constraint resulting from solving LP (10) for scenarios when resources are sufficient to cover 25% and 10% of eligible individuals.

In Figures 8 and 9, we see that even with gender equity enforced, more male population subgroups are allocated TasP than female subgroups. In Figure 10, we see that this is due to the larger number of HIV-infected females in each subgroup. Interestingly, we see a significant difference in the ages of female and males that would be prioritized for TasP in Figure 10. Specifically, we see that females in the age range of 20-25 would get TasP initially as opposed to a 25-30 age range for males.

5.2.2 Feasibility constraints

One of the larger debates in the mathematical modeling of HIV interventions is the relative importance of "core groups" and the practicality of relying on them in designing HIV interventions. On one hand, it is quite true that there is tremendous variation in HIV risk at the population level, even in settings with widespread HIV epidemics. In any model, the incorporation of core groups (i.e. those at elevated levels of risk such as the PRO and STD groups in the ASSA model) will estimate any intervention to be much more cost-effective than an analogous model with a more homogeneous risk of infection would (e.g. no core groups or less distinct groups).

While variation in HIV risk certainly exists, finding reliable data on sexual behavior is notoriously challenging. Even if such data were available and could be appropriately stratified in model risk groups, a practical issues remains of how public health officials in a setting would go about identifying individuals in different risk groups. Taking the optimistic view that the challenge could be overcome, it is clear that doing so would be require a significant amount of resources in itself, possibly enough to outweigh the cost-effectiveness gained by targeting such core groups.

Fortunately, our linear programming framework allows us to easily incorporate and evaluate feasibility constraints. To consider the case where we assume that we can not distinguish between sexual risk groups (and therefore not use risk groups in our allocation), we include additional constraints requiring that TasP allocations be identical across risk groups in the following linear program:

$$\begin{aligned} \max \quad & \sum_{i,j,k,l} (.95)\, P_{i,j,k,l}\, I_{i,j,k,l}\, X_{i,j,k,l}, \\ \text{s.t.} \quad & \sum_{i,j,k,l} P_{i,j,k,l}\, X_{i,j,k,l} \leq R, \\ & \sum_{j,k,l} X_{\text{male},j,k,l}\, P_{\text{male},j,k,l}, = \sum_{j,k,l} X_{\text{female},j,k,l}\, P_{\text{female},j,k,l}, \\ & \quad X_{i,PRO,j,k} = X_{i,STD,j,k} = X_{i,RSK,j,k}, \\ & \quad 0 \leq X_{i,j,k,l} \leq 0.90. \end{aligned} \tag{11}$$

While our discussion here focuses on the feasibility of using risk groups in our allocation plan, there are numerous feasibility concerns that could arise and be investigated with this framework. For instance, in Figure 10 we see in the 25% scenario that optimization allocates TasP to females aged 20-23 and 25-26, but not for those that 24 which may not be feasible in practice. Consequently, public policy officials may prefer that an allocation plan be based on age groups (i.e.g 15-19, 20-25, etc.) which could be implemented into our linear program. Perhaps, it may be viewed as too confusing for age eligibility to be different between females and males. If so, constraints could be added to the linear program to have allocations be independent of gender. In general, we can address any feasibility issue of resource allocation (provided they pertain to characteristics of the ASSA model: gender, age, HIV stage of infection, risk group) via constraints on our linear program.

5.3 Sensitivity analysis

In the previous section, we demonstrated the versatility of the resource allocation framework in addressing ethical and feasibility concerns. In this section, we use this versatility to compare a variety of allocation strategies across the spectrum of resource availability. In doing so, we shift the research question from "How should a fixed quantity of resources be allocated?" to "Given the possibility of a new intervention, what resources are necessary to ethically and feasibly meet the objectives of the society?"

To do so, we wish to evaluate how sensitive the outcomes (i.e. number of infections prevented) are to the amount of resources available and the allocation strategy employed. We let the resource constraint vary from being able to cover 3% to 90% of individuals eligible for TasP and consider allocations that are based on all combinations of age, disease stage and risk group. For each, we compare allocations that include the gender equity constraints and ones that do not. The results are presented in Figure 11 which shows the total number of infections prevented for each allocation strategy at each resource level.

In Figure 11, we see the number of infections prevented in the first year of TasP varying from around 25,000 to over 200,000 depending on the resources available. As the dotted curves represent infections prevented using gender equitable allocation plans, we see that the cost of gender equity (i.e. number of fewer infections prevented compared to plan without gender equity) is minimal with more detailed allocation plans (e.g. allocating by all factors, age and disease stage, sexual risk group and age) and more significant in plans based on stage or risk group only. In all allocation plans, this difference is also heavily dependent on the amount of available resources.

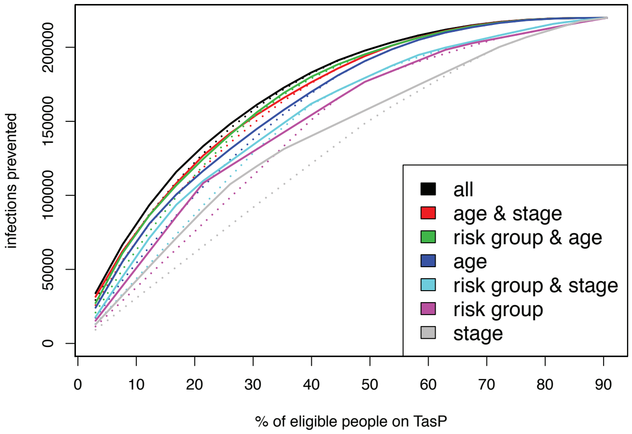


Figure 11: Total number of HIV infections prevented for each resource allocation strategy as a function of resource level. Solid curves illustrate infections prevented when allocations do not include gender equity. Infections prevented by allocations with gender equity depicted by dotted curves.

To better contrast the different allocation strategies, we visualize the same results of Figure 11 using a different metric: cost-effectiveness. To do so, we calculate the number of HIV infections prevented per 100 people on TasP. The results are presented in Figure 12.

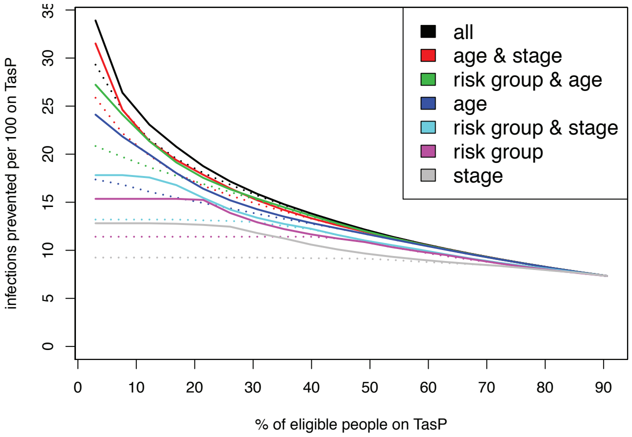


Figure 12: Cost-effectiveness, measured in HIV infections prevented per 100 people on TasP, for each resource allocation strategy as a function of resource level. Solid curves illustrate results when allocations do not include gender equity. Results from allocations with gender equity depicted by dotted curves.

As one would expect, the allocation plan based on all population subgroups that does not require gender equity is always the most cost-effective plan. Unexpectedly, we see that a plan that does not include risk groups (i.e. based on age and disease stage alone) is almost just as cost effective. Given the difficulty of identifying sexual risk groups discussed in Section 5.2.2, this welcome finding suggests that doing so may actually be unnecessary.

In Figure 12, we notice that age is a particularly influential element of allocation plans. The four most cost-effective plans all include age and the three least effective plans all do not. Moreover, the plan based on the single factor of age outperforms the plan based on both risk group and stage.

Figure 12 also illustrates the relationship between the amount of resources available and the importance of properly allocating those resources. When prevention resources are scarce, the choice of allocation plan has a large impact on cost-effectiveness. This impact decreases if resources are less constrained.

In terms of the complicated issue of gender equity in allocating TasP resources discussed in Section 5.2.1, our modeling also provides encouraging results. We find that imposing gender equity (depicted by dotted curves in Figure 12) affects cost effective significantly only when allocation plans are less detailed (i.e. those based on risk group and disease stage, risk group, or disease stage) and resource availability is low. In more detailed allocation plans, the cost-effectiveness reductions from imposing gender equity largely disappear if resources are available for 30% or more of eligible individuals.

6 Discussion

In this work, we have described a new framework of allocating intervention resources for infectious diseases based on linear programming. We have demonstrated how this approach could produce detailed, ethical and equitable allocation plans for the rollout of Treatment as Prevention resources for HIV in South Africa.

Ideally, we would all like to believe that health policy would be informed by a thorough investigation that takes future consequences into deep consideration. However, the reality is that funding decisions are often made by elected officials which means that short-term effects may be more heavily weighted than longer-term consequences. While hopefully not usually the case, the incentive is for leaders to pursue endeavors that show immediate results rather than longer-term results. From a less cynical view, it is still true that long-term funding is rarely guaranteed for a disease intervention even under thoughtful leadership.

In this reality, it is important for an intervention to produce measurable positive results in the short term in order to secure long-term funding commitment.

As mentioned previously, resource allocations resulting from optimal control theory often dictate that all resources are used as quickly as possible to get the full feedback effect of reduced transmission throughout the time period. For HIV, as considered in this case study, such a strategy could be disastrous. The cessation of antiretroviral therapy for HIV-infected individuals could dramatically increase levels of drug-resistant HIV. Such a scenario would leave an HIV epidemic that is more deadly and more difficult and expensive to manage. For such a reason, initiating an individual on antiretroviral therapy should be viewed by policy makers as a commitment to keep them on ART indefinitely.

From a mathematical perspective, the problem of allocating a certain fixed amount of resources to prevent the most infections is a well-defined and solvable mathematical problem, but such a question is rarely asked. Instead, a related, but more realistic question is a better goal for modeling: "How can we make a convincing argument that X resources should be allocated to a certain intervention?" In addition, we have seen that practicality and ethical concerns can be addressed with this robust allocation strategy. For example, rather than going down the ethical rabbit hole of deciding whether imposing gender equity is the proper course of action, even though it will reduce overall HIV reductions, modeling can shift the discussion to how do we successfully argue for the resources for 40% coverage so that gender equity is a non-issue.

References

[1] Actuarial Society of South Africa, *AIDS models*. Available at `https://www.actuarialsociety.org.za/downloads/committee-activities/aids-models/`

[2] S.J. Anderson, P. Cherutich, N. Kilonzo, I. Cremin, D. Fecht, D. Kimanga, M. Harper, . . . , T.B. Hallett, *Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: a modelling study*, The Lancet **384**(2014), no. 9939, 249–256. doi: `10.1016/S0140-6736(14)61053-9`

[3] M.C. Boily, R.F. Baggaley, L. Wang, B. Masse, R.G. White, R.J. Hayes, M. Alary, *Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies*, Lancet Infect. Dis. **9**(2009), no. 2, 118–129. doi: `10.1016/S1473-3099(09)70021-0`

[4] M.L. Brandeau, G.S. Zaric, A. Richter, *Resource allocation for control of infectious diseases in multiple independent populations: beyond cost-effectiveness analysis*, Journal of Health Economics **22**(2003), no. 4, 575–598. doi: `10.1016/S0167-6296(03)00043-2`

[5] M.S. Cohen, Y.Q. Chen, M. McCauley, T. Gamble, M.C. Hosseinipour, N. Kumarasamy, J.G. Hakim, . . . , T.R. Fleming, *Prevention of HIV-1 infection with early antiretroviral therapy*, N. Engl. J. Med. **365**(2011), no. 6, 493–505. doi: `10.1056/NEJMoa1105243`

[6] R. Dorrington, L.F. Johnson, D. Budlender, *ASSA2008 AIDS and demographic models*. Centre for Actuarial Research, University of Cape Town, 2010.

[7] A. Duriux-Smith, J.T. Goodman, European Study Group on Heterosexual Transmission of HIV, *Comparison of female to male and male to female transmission of HIV in 563 stable couples*, British Medical Journal **304**(1992), no. 6830, 809–813. doi: `10.1136/bmj.304.6830.809`

[8] S.R. Earnshaw, K. Hicks, A. Richter, A. Honeycutt, *A linear programming model for allocating HIV prevention funds with state agencies: a pilot study*, Health Care Manag. Sci. **10**(2007), no. 3, 239–252. doi: `10.1007/s10729-007-9017-8`

[9] R.M. Granich, C.F. Gilks, C. Dye, K.M. De Cock, B.G. Williams, *Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model*, The Lancet **373**(2009), no. 9657, 48–57. doi: `10.1016/S0140-6736(08)61697-9`

[10] R.M. Grant, J.R. Lama, P.L. Anderson, V. McMahan, A.Y. Liu, L. Vargas, P. Goicochea, . . . , D. V. Glidden, *Preexposure chemoprophylaxis for HIV prevention in men who have sex with men*, N. Engl. J. Med. **363**(2010), no. 27, 2587–2599. doi: `10.1056/NEJMoa1011205`

[11] J.L. Juusola, M.L. Brandeau, *HIV treatment and prevention: a simple model to determine optimal investment*, Medical Decision Making **36**(2016), no. 3, 391–409. doi: `10.1177/0272989X15598528`

[12] E.H. Kaplan, M.H. Merson, *Allocating HIV-prevention resources: balancing efficiency and equity*, Am J Public Health **92**(2002), no. 12, 1905–1907. doi: `10.2105/ajph.92.12.1905`

[13] A. Lasry, G.S. Zaric, M.W. Carter, *Multi-level resource allocation for HIV prevention: a model for developing countries*, European Journal of Operational Research **180**(2007), no. 2, 786–799. doi: `10.1016/j.ejor.2006.02.043`

[14] O. Shisana, T. Rehle, L.C. Simbayi, K. Zuma, S. Jooste, N. Zungu, D. Labadarios, D. Onoya, *South African national HIV prevalence, incidence and behaviour survey, 2012*, Human Sciences Research Council (HSRC) Press, Cape Town, South Africa, 2014. Available from: `http://www.hsrc.ac.za/en/research-outputs/view/6871`

[15] UNAIDS, *UNAIDS country data: South Africa*, 2015. `https://www.unaids.org/en/regionscountries/countries/southafrica`

[16] World Health Organization, *Interim WHO clinical staging of HIV/AIDS and HIV/AIDS case definitions for surveillance: African region*, World Health Organization, Switzerland, 2005. `https://www.who.int/hiv/pub/guidelines/casedefinitions/en/`

Appendix

Table 2: ASSA model initial distribution of population into sexual risk groups.

Risk Group	% of Male Pop.	% of Female Pop.
PRO	1%	1%
STD	18%	18%
RSK	28%	28%
NOT	53%	53%

Table 3: ASSA model assumptions of sexual mixing between sex risk groups. Female selection of partners is assumed. Male behavior is derived from the assumptions for females and the assumed sexual mixing patterns between males and females.

Female	Source of male partner			
risk group	PRO	STD	RSK	NOT
PRO	0.75	0.25	0	0
STD	0.2	0.75	0.05	0
RSK	0	0.4	0.6	0
NOT	0	0	0	1

Male	Source of female partner			
risk group	PRO	STD	RSK	NOT
PRO	0.81	0.19	0	0
STD	0.27	0.69	0.05	0
RSK	0.00	0.39	0.61	0
NOT	0	0	0	1

Table 4: ASSA model assumptions of new partners per year. Again, female behavior is assumed and male behavior is derived from from the distribution for females and the age mix of the partners of the women.

Risk group	New partners per year year for females	New partners per year for males
PRO	250	231
STD	12.00	13.09
RSK	1.00	0.99

Table 5: ASSA model assumptions of the average number of contacts per sexual partnership.

Female	Contacts per male partner		
risk group	PRO	STD	RSK
PRO	1	1	0
STD	3	13	45
RSK	0	50	95

Table 6: Baseline per act transmission probabilities for sexual contacts between risk groups and factors of transmission efficiency based on disease stage. Differences of baseline transmission rates among risk groups reflect prevalences of other sexually transmitted diseases among the risk groups. Transmission efficiency factors reflect viral load levels at different stages of infection.

Female to male transmission

	PRO	STD	RSK
PRO	0.005	0.005	
STD	0.005	0.005	0.003
RSK		0.003	0.001

Male to female transmission

	PRO	STD	RSK
PRO	0.007	0.007	
STD	0.007	0.007	0.0045
RSK		0.0045	0.002

Transmission efficiency

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
0.5	0.4	1.5	2.9	0.134	2.9

Table 7: Factor for increased susceptibility to HIV for females by age.

Age of female	Increased susceptibility factor
14	3.725207494
15	3.239310864
16	2.816792056
17	2.449384396
18	2.129899475
19	1.8520865
20	1.61051
21	1.4641
22	1.331
23	1.21
24	1.1
25	1
⋮	⋮

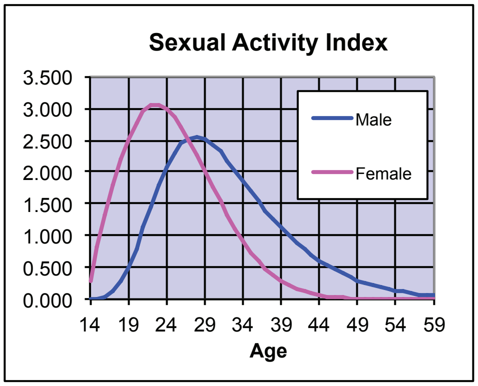


Figure 13: Age-based “sexual activity index”; a multiple for the baseline number of new sexual partners per year (from Table 4) based on age.

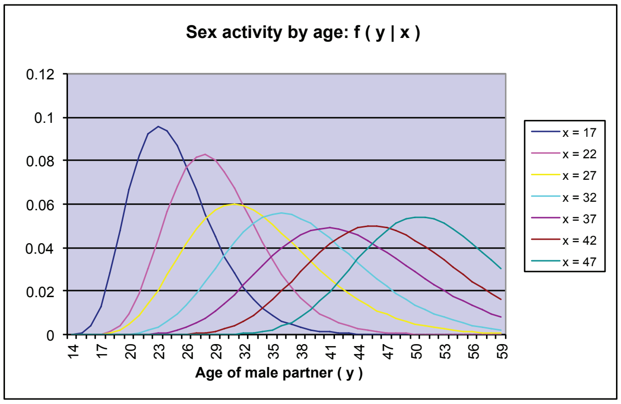


Figure 14: Age distributions of the sexual partners for females of ages $x = 17, 22, 27, ..., 47$. Again, female behavior is assumed and male behavior is derived from the distribution for females.

Table 8: ASSA model multiples for disease-stage adjustments to sexual behavior and condom usage.

Amount of sexual contacts		
	Males	Females
Stage 1	0.986	0.964
Stage 2	0.976	0.946
Stage 3	0.632	0.616
Stage 4	0.242	0.236
Stage 5	0.552	0.552
Stage 6	0.173	0.170

Condom non-usage		
	Males	Females
Stage 1	0.974	0.932
Stage 2	0.955	0.897
Stage 3	0.947	0.900
Stage 4	0.941	0.898
Stage 5	0.470	0.470
Stage 6	0.470	0.470

Table 9: Condom usage by age and sexual risk group.

Age	PRO	STD	RSK
14	0.947	0.817	0.690
15	0.947	0.816	0.689
16	0.946	0.815	0.688
17	0.946	0.815	0.688
18	0.946	0.815	0.688
19	0.946	0.816	0.689
20	0.927	0.760	0.613
21	0.927	0.760	0.613
22	0.927	0.760	0.612
23	0.926	0.759	0.612
24	0.926	0.758	0.611
25	0.864	0.614	0.443
26	0.864	0.614	0.443
27	0.864	0.613	0.442
28	0.864	0.613	0.442
29	0.863	0.612	0.441
30	0.845	0.577	0.405
31	0.845	0.576	0.405
32	0.844	0.575	0.404
33	0.844	0.574	0.403
34	0.843	0.573	0.402
35	0.705	0.374	0.230
36	0.704	0.373	0.229
37	0.703	0.372	0.229
38	0.703	0.371	0.228
39	0.702	0.370	0.227
40	0.701	0.369	0.226
41	0.700	0.369	0.226
42	0.701	0.369	0.227
43	0.702	0.371	0.227
44	0.703	0.372	0.229
45	0.705	0.374	0.230
46	0.705	0.374	0.230
47	0.706	0.375	0.231
48	0.705	0.374	0.230
49	0.705	0.374	0.230
50	0.705	0.374	0.230
51	0.705	0.374	0.230
52	0.705	0.374	0.230
53	0.705	0.374	0.230
54	0.706	0.375	0.230
55	0.706	0.375	0.231
56	0.706	0.375	0.231
57	0.705	0.374	0.230
58	0.704	0.373	0.229
59	0.703	0.372	0.228