Chapter 4

Modelling HIV Transmission in Sexual Partnerships: Equations and Assumptions

A key equation in transmission models defines the force of infection λ: the instantaneous rate at which susceptible individuals acquire infection. In the simplest compartmental transmission models — i.e., without any population stratification, repeated contacts, etc. — this rate is defined as:[1]

$$\lambda = C\beta\frac{I}{N} \tag{4.1}$$

where: C is the average contact rate per-person; β is the average probability of transmission per contact; and I/N is the current prevalence of infection.

If the population is stratified into multiple groups i, the infection is stratified into multiple infectious stages h, and contacts are stratified into multiple types p, then Eq. (4.1) can be generalized to:

$$\lambda_i = \sum_{pi'h'} C_{pii'}\beta_{ph'}\frac{I_{i'h'}}{N_{i'}} \tag{4.2}$$

where: $C_{pii'}$ is the average rate of type-p contacts per-person among group i with group i', $\beta_{ph'}$ is the average probability of transmission per type-p contact given infection stage h', and $I_{i'h'}/N_{i'}$ is the prevalence of infection stage h' among group i'. Note that Eq. (4.2) implicitly assumes that contact rate and mixing by infection status/stage is random.

The force of infection equation is further complicated by repeated contacts with the same individuals, such as in sexual partnerships (also household contacts, and other social relationships), where each contact reflects a single "sex act".[2] With repeated vs. random contacts, it is more likely that individuals who recently acquired or transmitted infection will continue to contact the same person, resulting in "wasted

[1] Eq. (4.1) also assumes frequency-dependent transmission vs. density-dependent transmission, which is almost always more appropriate for sexually-transmitted diseases [1].
[2] Sex involving vaginal *and* anal intercourse would be modelled as two sex acts.

contacts" (in terms of transmission),[3] and slower infection spread through the contact/partnership network.

This chapter explores different mathematical approaches to modelling HIV transmission within sexual partnerships in compartmental models, considering these "wasted contacts". In particular, I review prior approaches (§ 4.1), develop a new approach (§ 4.2), and compare the influence of each approach on selected model outputs (§ 4.3). A preliminary version of this approach was presented in [2].

4.1 Prior Approaches

The earliest HIV transmission models [3] were adapted from models of other sexually transmitted diseases, especially gonorrhea [4–6]. These early HIV transmission models did not explicitly model individual sex acts, but instead assumed an overall probability of transmission per partnership [7]. This assumption was initially justified via data suggesting that the probability of HIV transmission per partnership increased quickly and then saturated [8]. Such data were later explained by heterogeneity in infectiousness (e.g., due to infection stage, etc.) and/or susceptibility (e.g., due to genital ulcer disease, etc.) [9–11]. As this heterogeneity was quantified [9] and incorporated into HIV transmission models [12], the probability of transmission was increasingly parameterized per act vs. per partnership.

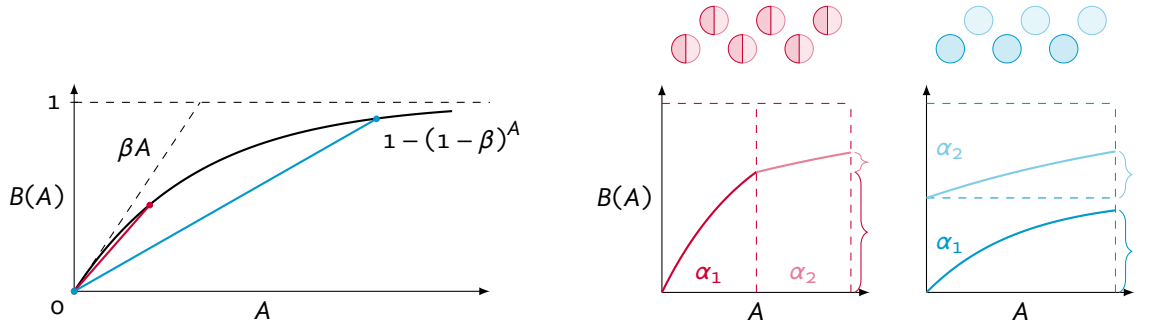
The shift to per-act vs. per-partnership parameterization highlighted a fundamental limitation of compartmental models (see § 4.1.5): compartmental models cannot model individual partnerships, because each "compartment" reflects a group of individuals whose characteristics are assumed to be homogeneous [13]. Therefore, the dynamics of sexual partnerships must be modelled using average rates of partnership change and average characteristics of those partnerships. As a result, partnerships are effectively modelled as instantaneous, such that the cumulative risk of transmission per partnership is applied at the moment of partnership change [14]. This cumulative risk can be defined in terms of the average total numbers of sex acts per partnership, but the timing of specific sex acts or other events within partnerships cannot be captured in compartmental models. Further implications of the "instantaneous partnership assumption" and alternate modelling frameworks which avoid this assumption are discussed in § 4.1.5, 4.1.6, and 4.2.

Thus, over the years, different force of infection equations have been designed for compartmental models which explicitly aggregate the risk of transmission across different numbers and types of sex acts, and likewise across different numbers and types of sexual partners/partnerships. The remainder of this section reviews these equations and their assumptions in detail.

4.1.1 Aggregating Sex Acts within a Partnership

To account for repeated sex acts with the same partner, the per-partnership probability of transmission B was conceptualized as follows [15]. Let A denote the total number of sex acts in the partnership, and β denote the probability of transmission per act. For now, β is assumed to be equal for all acts. With equal and independent β, the theoretical probability of n transmissions after A acts can be described by a

[3] Another conception of "wasted contacts" is "within-partnership competing risks".



(a) Probability of transmission per partnership B vs. number of sex acts A, comparing shorter (red) vs. longer (blue) partnerships

(b) Average accumulation of transmission probability for within-partnership heterogeneity (red) vs. between-partnership heterogeneity (blue)

Figure 4.1: Per-partnership robability of transmission vs. number of acts

B: probability of transmission per partnership; β: probability of transmission per act; A: total acts per partnership; α: fraction of total acts (within or between partnerships).

binomial distribution:

$$p(n) = \binom{A}{n} \beta^n (1-\beta)^{A-n} \tag{4.3}$$

Since transmission of HIV can actually only occur once, the per-partnership probability of transmission B is defined via the probability of "escaping" infection after all A acts:[4]

$$\begin{aligned} B &= 1 - p(n=0) \\ &= 1 - \binom{A}{0} \beta^0 (1-\beta)^A \\ &= 1 - (1-\beta)^A \end{aligned} \tag{4.4}$$

We can show that $B \leq \beta A$. Thus, a model with partnerships using Eq. (4.4) reduces transmission vs. a model without partnerships by adjusting for "wasted contacts". Figure 4.1a illustrates the shape of $B(A)$ (curve) and the corresponding effective probabilities of transmission per act B/A (tangent slopes) for a shorter (red) vs. longer (blue) partnership. Although $B(A)$ is monotonic, the effective probability of transmission per act decreases with the number of acts because, on average, more and more acts are "wasted" — occuring after transmission.

4.1.2 Heterogeneity in the Per-Act Probability of Transmission

As noted above, the per-act probability of transmission β is heterogeneous, varying with factors like: HIV infection stage, genital ulcer disease, condom use, etc. [11, 16]. The next step in developing a force of infection equation is to extend Eq. (4.4) to allow heterogeneity in β [15]. Let β_f denote the probability of transmission associated with a particular factor (or combination of factors) f; and let α_f denote the

[4] Eq. (4.4) can also be reasonably approximated via the Poisson distribution $B = 1 - e^{-\beta A}$ for small β.

proportion of acts A in an average partnership having transmission probability β_f (thus $\Sigma_f \alpha_f = 1$). There are two main approaches to aggregating β_f, reflecting different interpretations of α_f:[5] [6]

- **Within-Partnership Heterogeneity (WPH)**: modelled partnerships are identical, but comprise heterogeneous acts — α_f denotes a proportion of acts in each partnership (Figure 4.1b red).

$$B_{\text{WPH}} = 1 - \prod_f (1 - \beta_f)^{A\alpha_f} \tag{4.5}$$

- **Between-Partnership Heterogeneity (BPH)**: modelled partnerships are different, but each comprise identical acts — α_f denotes a proportion of partnerships (Figure 4.1b blue).

$$B_{\text{BPH}} = 1 - \sum_f \alpha_f (1 - \beta_f)^A \tag{4.6}$$

Figure 4.1b illustrates these approaches for a simple case with two factors. For WPH (red), each factor f marginally contributes to the probability of escaping infection in every partnership. For BPH (blue), the overall probability of escaping infection is modelled as a weighted average across partnerships, each affected by a single factor f. Both approaches guarantee $B < 1$, but we can show that $B_{\text{WPH}} \geq B_{\text{BPH}}$ by the (weighted) AM-GM inequality (see § A.3.1) [17]. Intuitively, this inequality arises because any large probability of transmission β_f has disproportionate influence in Eq. (4.5), even for a small proportion of acts affected α_f, whereas this influence is bounded by α_f in Eq. (4.6), as shown in Figure 4.1b.

The decision to use WPH vs. BPH for aggregating specific types of heterogeneity in β should be driven by the factor(s) in question. To this end, it is possible to combine Eq. (4.5) and Eq. (4.6) as follows to aggregate both types of factors simultaneously:

$$B_{wb} = 1 - \sum_g \gamma_g \prod_f (1 - \beta_{fg})^{A\alpha_{fg}} \tag{4.7}$$

where: f denotes WPH factor(s); g denotes BPH factor(s); and γ_g replaces α_f for BPH factors. Then, for example, if it is known or assumed that "50% condom use" reflects 50% condom use in 100% of partnerships, sex acts with condoms vs. without condoms should be aggregated as WPH, with $\alpha_f = 0.5$. By contrast, heterogeneity in individual-level factors like infection stage or treatment status should be aggregated as BPH,[7] with γ_g as the conditional prevalence of each stage/status g among infected partners. In fact, aggregating infection stage and treatment status is often deferred to the full incidence equation (see § 4.1.3) using an equivalent form, but where γ_g is replaced by the unconditional prevalence of stage/status g among *all* partners.

[5] In most compartmental models without repeated contacts (partnerships), this distinction is not possible or necessary, because all contacts (sex acts) between two compartments (risk groups) are assumed to be independent.

[6] In fact, further variations on Eqs. (4.5) and (4.6) have also been used; § **??** gives some examples.

[7] Individual-level factors should be aggregated as BPH because a given partner has exactly one current infection stage or treatment status; of course, this stage/status could evolve over the course the partnership, but this future trajectory is not explicitly modelled — which only serves to highlight the limitations of either approach to aggregating heterogeneity in β.

4.1.3 Aggregating Partnerships

Although we considered between-partnership heterogeneity in § 4.1.2, the modelled per-partnership probability of transmission B still corresponds to a single average partnership. Some population groups may have multiple partners per unit time (usually year), possibly including different types of partnerships, or less than one partnership per year, on average. Thus, the second step in constructing the incidence equation is to aggregate transmission risk across these various partnerships [15].

As in § 4.1.2, there are two main approaches to aggregating partnerships — indeed having similar equations to Eqs. (4.5) and (4.6). This time, the distinction may be more familiar:

- **Incidence Rate:** instantaneous rate of infection among susceptible individuals — transmission risks are addative; can have $\lambda_i > 1$.

$$\lambda_i^{\text{IR}} = \sum_{pi'h'} Q_{pii'} B_{pii'h'} \frac{I_{i'h'}}{N_{i'}} \tag{4.8}$$

- **Incidence Proportion:** cumulative proportion of susceptible individuals infected over a period Δ_t — transmission risks are competing; can only have $\lambda_i \leq 1$.

$$\lambda_i^{\text{IP}} = 1 - \prod_{pi'h'} \left(1 - B_{pii'h'} \frac{I_{i'h'}}{N_{i'}}\right)^{Q_{pii'}\Delta_t} \tag{4.9}$$

where: $Q_{pii'}$ is the rate of type-p partnership formation between groups i and i',[8] $B_{pii'h'}$ is the average per-partnership probability of transmission from group/infection stage $i'h'$ to group i' via partnership type p, and $I_{i'h'}/N_{i'}$ is the prevalence of infection stage h' among group i'. Similar to within- vs. between-partnership heterogeneity, we can show that $\lambda^{\text{IR}} \geq \lambda^{\text{IP}}$ (see § A.3.1).

By definition, the force of infection is a rate [18]. Yet in principle, incidence proportion could be more precise than incidence rate *over a given time period* Δ_t. Since most models are now solved computationally, the period Δ_t could be matched to the timestep of the numerical solver. However, the added precision may be insignificant, as Δ_t should be small.[9] Moreover, some applications of this approach have used a period of $\Delta_t = 1$ year in the equation, but then applied the resulting incidence proportion λ_i^{IP} as a rate over smaller timesteps. Such applications then erroneously reduce *current* transmission due to *future* "wasted contacts" across multiple partnerships. This reduction is then redundant because "wasted contacts" *within* partnerships are already captured via the per-partnership probability of transmission, while "wasted contacts" *between* partnerships are already captured via multiple opportunities for transmission over successive timesteps. In sum, unless the period Δ_t can be matched to the numerical solver timestep, incidence rate Eq. (4.8) is preferred over incidence proportion Eq. (4.9) as the force of infection equation.

[8] This fully stratified partnership formation rate $Q_{pii'}$ is often broken down into a per-person partnership formation rate Q_{pi} and a mixing matrix $\rho_{pii'}$, as in § **??**.

[9] Also, many popular numerical solvers do not pass the timestep Δ_t (only the current time t) to the derivative computing function, and may use adaptive timestep sizes for precision while solving — including: `scipy.integrate.odeint` in Python, `deSolve::lsoda` in R, and `ode45` in MATLAB.

4.1.4 Revisiting Partnership Duration

A final issue in constructing the force of infection equation relates to parameterization. In Eqs. (4.3)–(4.9), partnership durations δ are not explicitly modelled, but implied by the total numbers of sex acts per partnership A, and a presumed frequency of sex per partnership F, such that $A = F\delta$. By contrast, the partnership formation rate Q is often directly informed by survey questions like "*How many different people have you had sex with in the past 12 months?*" (see also § A.2.3). As such, the lowest possible value among sexually active individuals could naively be taken as $Q = 1$ (per year). Then, if $Q \geq 1$ is used in the model, the total sex acts per partnership can (and should) be reduced to reflect up to one year — i.e., $A \leq F$, or effectively $\delta \leq 1$ year. This change vs. $Q < 1, A > F$ can substantially reduce the proportions of sex acts which are modelled as "wasted", and thus increases transmission via what would be longer ($\delta > 1$ year) partnerships. On the other hand, using the true $Q < 1, A > F$ can effectively delay early transmission in longer partnerships, such that modelled HIV prevalence could lag behind true HIV prevalence. These dynamics are further explored in simulation experiments (§ 4.3.3).

Lastly, it is worth noting that partnership duration δ is further related to the average partnership formation rate Q and the average number of concurrent partners K by $Q = K/\delta$.[10] Thus, an alternate parameterization could specify the number of concurrent partners K and the frequency of sex with each partner F. The overall rate of sex would be the same: $QA = KF$, since $A = F\delta$. In some ways, this KF parameterization is more intuitive, and it will be useful later, in the new force of infection approach (§ 4.2).

4.1.5 Limitations of Prior Approaches

Evidently, prior approaches to modelling HIV transmission in sexual partnerships have several limitations (see also [13]). These limitations, and their implications for existing model-based evidence, can be summarized as follows.

Instantaneous Partnerships. Eqs. (4.8) and (4.9) both include the current HIV prevalence I/N directly in the force of infection. Thus, newly infected individuals are modelled to be immediately at risk of onward transmission, including via the exact same partnership by which they were infected. Similarly, individuals who recently transmitted to a given partner are also modelled to be at risk of transmitting (again) to the same partner, albeit with a small absolute rate reduction due to the smaller susceptible population. This modelling assumption acts to increase the modelled rate of transmission vs. "reality", especially for longer partnerships. As a result, the contribution of longer partnerships to overall transmission has likely been systematically overestimated by compartmental models, while the contribution of shorter partnerships has likely been underestimated. This point was directly illustrated in a comparison of compartmental vs. network-based models [19]. This bias has substantial implications for which types of interventions and populations have been prioritized using compartmental model-based evidence.

Aggregating Past/Future Sex Acts. The instantaneous partnerships assumption is in fact directly related to the "wasted contacts" issue, because the delay in onward transmission risk that is missing under instantaneous partnerships reflects the same post-transmission period within partnerships wherein contacts are "wasted". The prevailing solution to this issue is to define the per-partnership probability of

[10] Gaps between partnerships do not result in $Q < K/\delta$, because the average K would be reduced if some individuals are "between partnerships".

transmission B by aggregating competing risks from each sex act within a given partnership via Eq. (4.4) et al. However, as described in § 4.1.4, this approach introduces a trade off between capturing the true proportion of "wasted contacts" in longer partnerships (using the true partnership duration δ) vs. capturing the true magnitude of early transmission within partnerships (using $\delta \leq 1$). These two options would then underestimate or overestimate the contribution of longer partnerships to overall transmission, respectively, with broader implications as described above. Moreover, the sex acts aggregated within each partnership via Eq. (4.4) et al. are parameterized to reflect current conditions — i.e., HIV stage, treatment status, condom use, etc. — even though such conditions almost certainly evolve over the course of partnerships, especially longer partnerships. Again, we can see the direct connection to the instantaneous partnerships assumption. This issue parallels limitations of cross-sectional risk factor analyses (e.g., § **??**), where risk factors are modelled as static, but true risk accumulates via cumulative exposure to dynamic risk factors. The implications of aggregating these past/future sex acts are not immediately obvious, and likely depend on numerous factors and conditions.

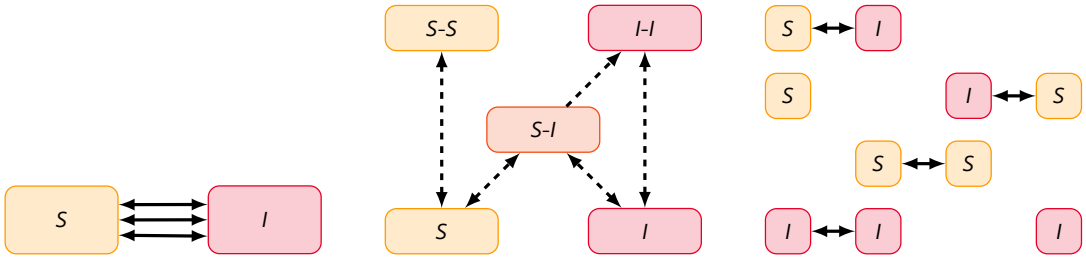
Incidence Proportion. Risk from multiple partnerships is sometimes aggregated as incidence proportion λ^{IP} via Eq. (4.9). As described in § 4.1.3, this approach is not inherently wrong, but the specification of time period $\Delta_t = 1$ often is. This Δ_t *should* be matched to the timestep of the numerical solver, but $\Delta_t = 1$ year is often used, and the resulting incidence applied as a rate over smaller timesteps, reducing transmission. Since λ^{IP} saturates at 1 — similar to $B(A)$ in Figure 4.1a — transmission to higher risk groups is disproportionately reduced. Thus, this error may *further* underestimate the importance of prevention among higher risk groups.

Within vs. Between Partnership Heterogeneity. The final limitation of prior approaches is the apparent lack of distinction between within- vs. between-partnership heterogeneity when computing the average per-partnership probability of transmission B. Both WPH and BPH — i.e., Eqs. (4.5) and (4.6) — and combinations thereof, have been used to model modified transmission risk in a proportion of sex acts due to HIV stage, treatment status, PrEP use, condom use, STI co-infection, circumcision, and more, but the choice of aggregation model is almost never explicitly justified. For some factors, there may be no "correct" choice, but modellers should be aware of the assumptions implied by their choice. The implications of model choice for transmission dynamics mainly derive from the fact that $B_{\text{WPH}} \geq B_{\text{BPH}}$, but even then the differences are often small (see § 4.3.1).

4.1.6 Alternate Modelling Frameworks

Recognizing the limitations of compartmental models in simulating infectious disease transmission via sexual partnerships, two main alternate modelling frameworks have been developed. These frameworks are illustrated in Figure 4.2. A more detailed system for classifying modelling frameworks is also given in Appendix 1 of [19].

Pair-Based Models. Pair-based models, also known as pair-formation models, were developed as early as 1988, with the explicit motivation to overcome limitations of classic compartmental models of STI transmission [14]. In pair-based models, the fundamental population stratification reflects different partnership configurations and health states [20], such as: susceptible and single, a susceptible/infected pair in a long-term partnership, etc. (Figure 4.2b). Such models can therefore track the numbers of partnerships where transmission is vs. is not possible, thereby avoiding the instantaneous partnership



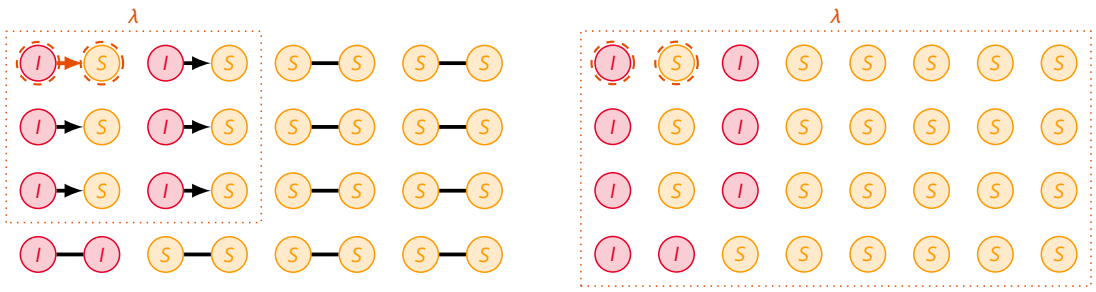
(a) Classic compartmental (b) Pair-based (c) Individual-based

Figure 4.2: Representations of health states and sexual partnerships under three different STI modelling frameworks

S: susceptible; *I*: infectious; solid arrows: partnerships; dashed arrows: state transitions.

assumption. Pair-based models have been applied to a variety of STIs [20]. However, the numbers of compartments required to reflect all possible partnership configurations *and* all possible health states among connected partners quickly become impractical [13, 20]. For example, a classic compartmental model with 2 risk groups and 2 health states would require $2 \times 2 = 4$ compartments; whereas even a first-order pair-based model (i.e., without "triples") would require 2×2 (singles) + $(2 \times 2)^2$ (pairs) $= 20$ compartments; a second-order pair-based model would require $4 + 4^2 + 4^3 = 84$ compartments. Thus, pair-based models are especially limited in their ability to model partnership concurrency — the role of which in HIV epidemiology remains controversial [21]. As such, pair-based models have seen little widespread adoption for STI or HIV transmission modelling [13].

Individual-Based Models. Individual-based models, also knowns as agent-based, network-based, or microsimulation models, explicitly simulate unique individuals (Figure 4.2c), representing a fundamental change in the model unit from groups of individuals — i.e., the "compartments" of compartmental models [13]. Individual-based models can therefore model unique partnerships, and track them over time. Such individuals and partnerships can then be parameterized in fundamentally different ways vs. compartmental models, including with continuous valued features like infection age and sex frequency, vs. predetermined categories like infection stages and partnership types [13, 22]. Parameters for each individual and partnership are thus sampled randomly and/or dynamically, allowing more complete and nuanced representations of risk heterogeneity and partnership dynamics. Such nuances can in fact be key determinants of epidemic dynamics and intervention impact [19, 23]. Evidently, many of the limitations of compartmental and pair-based models do not apply to individual-based models [13]. Yet these limitations are replaced with new challenges, especially related to implementing, parameterizing, and calibrating these powerful models [13, 22, 24]. For example, much effort has been dedicated to formalizing the statistical properties of dynamic networks via temporal exponential family random graph models (tERGM) [25] or latent order logistic models (LOLOG) [26], so that dynamic networks can be generated which are consistent with observed data. Although individual-based models have seen greater use than pair-based models, these challenges still prevent universal adoption over classic compartmental models [13].



(a) Pair-wise reality (b) Frequentist approximation

Figure 4.3: Comparison of pair-based reality and frequentist approximation for a population of 16 pairs with 25% infection prevalence, at the moment of one transmission event

S: susceptible; *I*: infectious; λ: force of infection; dashed circles: individuals involved in transmission event.

4.2 Proposed Approach

In this section, I propose a new approach to modelling HIV transmission within sexual partnerships in compartmental models. The approach overcomes the limitations of prior approaches described in § 4.1.5, without the need to change modelling frameworks.

4.2.1 Illustrative Toy Scenario

Consider the moment of one transmission event in a population of 16 monogamous partnerships, with 25% infection prevalence and random mixing by infection status (Figure 4.3a). Initially, infection prevalence is equal among partners of susceptible *S* and infectious *I* individuals: 6/24 and 2/8, respectively. Immediately after transmission, prevalence decreases to 5/23 among partners of *S* but increases to 4/9 among partners of *I*, decreasing the population-level transmission risk. Next, three events are possible:

(a) another transmission occurs among the remaining *S*-*I* partnerships, yielding 4/22 prevalence among partners of *S*, and 6/10 among partners of *I*; population-level transmission risk decreases further

(b) the partnership from the original transmission ends, and both individuals form new partnerships (assumed at random), yielding, on average, 9/32 prevalence among partners of both *S* and *I*; population-level transmission risk increases *above* the initial level (9/32 > 6/24)

(c) any other partnership ends, and both individuals form new partnerships (assumed at random); infection prevalence among *S* and *I*, and population-level transmission risk all remain unchanged, on average

Prior compartmental models have effectively assumed that event (b) always occurs before (a) — i.e., the "instantaneous partnership assumption". This assumption is reflected in Figure 4.3b, where the frequentist approximation does not explicitly model any individual partnerships. Evidently, this assumption is worse for longer partnerships.

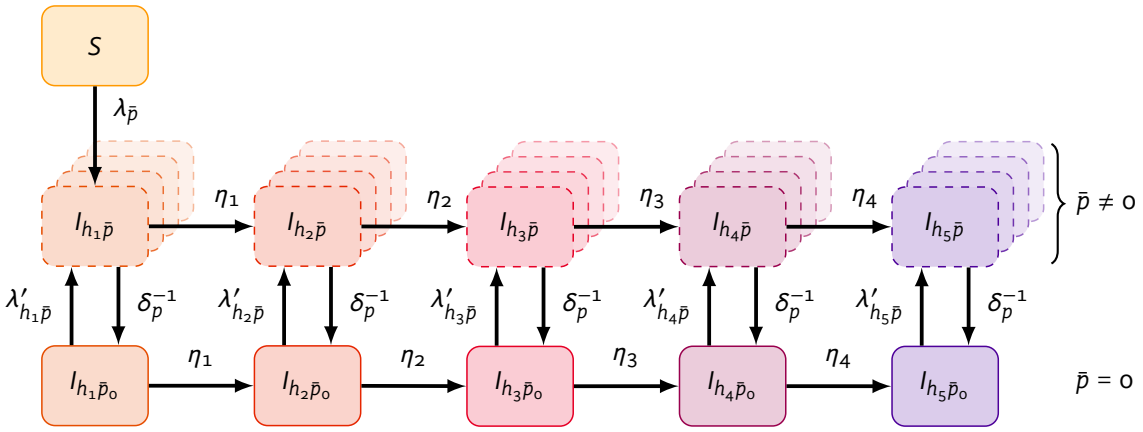


Figure 4.4: Modelled states and transitions related to HIV infection, and a new stratification $\bar{p}$ to track the proportions of individuals in partnerships where transmission already occurred

S: susceptible; I_h: infectious in stage h; p: partnership type; $\bar{p}$: new stratification, where $\bar{p} = 0$ reflects no recent transmission (all new partners), and $\bar{p} \neq 0$ reflects recent transmission via a type-p partnership; λ: force of infection per susceptible; λ': force of infection per infectious; η: rate of progression between infection stages; δ: partnership duration.

4.2.2 Conceptual Development

The toy scenario highlights how any partnership where transmission has occured should be "removed" from the force of infection. In a compartmental (non-pair-based) model, these partnerships can be tracked as proportions of individuals: namely, all individuals who recently acquired infection *and* all individuals who recently transmitted infection. Here, I use "recent" to mean *before individuals change partners*. If some individuals have multiple concurrent partnerships ($K > 1$), then these individuals should not be removed entirely, but their effective numbers of partners should be reduced by 1. If multiple types of partners are considered, then only the partnership type involved in the transmission should be reduced. This adjustment can then be applied until the individuals change partners — at an effective rate inversely related to partnership duration: δ^{-1}. However, during this period, these individuals should still be modelled to progress as usual through different stages of infection, activity group turnover, etc.

Using this conceptual basis, I propose a new stratification of the modelled infected population, denoted $\bar{p}$. The stratum $\bar{p} = 0$ corresponds to no recent transmission, or all "new" (potentially discordant) partners. Other strata $\bar{p} \neq 0$ correspond to recent transmission via (to or from) partnership type $\bar{p}$. Figure 4.4 illustrates the new stratification together with with the existing HIV infection stratification (Figure **??**). Following infection, all individuals enter a stratum $\bar{p} \neq 0$ corresponding to the partnership type p by which they were infected. Thus, the rate of entry to this stratum from S_i is defined by the incidence rate without aggregating across partnership types: λ_{ip}. Individuals may then transition from $\bar{p} \neq 0$ to $\bar{p} = 0$ upon forming a new partnership, at a rate δ_p^{-1}. Finally, individuals may re-enter any stratum $\bar{p} \neq 0$ if they transmit infection via partnership type p. I denote the corresponding rate as λ'_{ip}, representing the per-person rate of *transmission*, not *acquisition* as in λ_{ip}. This rate λ'_{ip} is not defined or needed in prior models (§ 4.1) but I develop the necessary equations below in § 4.2.3. The issue of transmission via multiple partnerships is discussed in § 4.2.4.

4.2.3 Equations[11]

Since partnership duration is now considered separately and explicitly, I do not define any adjusted per-partnership probability of transmission B. Rather, I define the force of infection to directly include the frequency of sex per partnership F and probability of transmission per sex act β. However, the mixing is now slightly more complicated, since the effective number of partnerships "offered" depends on infection status. In addition, these partnerships are now defined as numbers of concurrent partners K, rather than partnership formation rates Q.

Let $M_{pii'}$ be the total (population-level, not per-person) number of type-p partnerships between group i and group i'. As described § **??**, this "mixing matrix" $M_{pii'}$ can be defined in several ways, based on the total numbers of partnerships "offered" by each group: $M_{pi}, M_{pi'}$, plus some parameter(s) specifying mixing patterns (e.g., Φ). Working backwards, I start by defining M_{pi} (and likewise $M_{pi'}$) via the sum across health statuses — i.e., susceptible, and different stages of infection h:

$$M_{pi} = M_{S,pi} + \sum_h M_{I,pih} \tag{4.10}$$

I then define the total numbers of partnerships "offered" by susceptible individuals as:

$$M_{S,pi} = S_i K_{pi} \tag{4.11}$$

and likewise for individuals in infection stage h:

$$M_{I,pih} = I_{ih,\bar{p}=p}(K_{pi} - 1) + \sum\nolimits_{\bar{p}\neq p} I_{ih\bar{p}}\, K_{pi} \tag{4.12}$$

Eq. (4.12) is the key equation whereby the effective numbers of type-p partnerships "offered" by individuals in stratum $\bar{p}$ are reduced by 1. This reduction is then propagated through the mixing patterns when defining $M_{pii'}$. Thus, we are now ready to construct the overall force of infection equation as follows. I define the total (population-level, not per-person) rate of transmission from group i' and infection stage h' to group i via type-p partnerships as:

$$\Lambda_{pii'h'} = F_p \beta_{pii'h'} M_{pii'} \left(\frac{M_{S,pi}}{M_{pi}}\right)\left(\frac{M_{I,pi'h'}}{M_{pi'}}\right) \tag{4.13}$$

where the two fractions represent the proportions of all partnerships $M_{pii'}$ formed between susceptible individuals from group i ($M_{S,pi}$) and infectious individuals in group/infection stage $i'h'$ ($M_{I,pi'h'}$). The per-person transmission rates to group i, and from group $i'h'$ can then be defined as:

$$\lambda_{pi} = \sum_{i'h'} \Lambda_{pii'h'}\, S_i^{-1} \tag{4.14}$$

$$\lambda'_{pi'h'} = \sum_{i} \Lambda_{pii'h'}\, I_{i'h'}^{-1} \tag{4.15}$$

[11] "*Enough talk. Show me the $*" — LaTeX users.

For the purposes of solving the model, we can skip division by S_i and $I_{i'h'}$ in Eqs. (4.14) and (4.15), since λ'_{pi} and $\lambda'_{pi'h'}$ are immediately multiplied by S_i and $I_{i'h'}$, respectively, in the system of differential equations.

Finally, and to reiterate from above, infected individuals in stratum $I_{ih\bar{p}}$ are assumed to form new partnerships at a rate $\delta_{\bar{p}}^{-1}$, and thereby transition to stratum $I_{ih\bar{p}_0}$ ("all new partners"); and otherwise transition between infection stages, cascade of care, activity groups, etc. as usual, as illustrated in Figure 4.4.

4.2.4 Transmission via Multiple Partnerships

In the proposed approach (i.e., new stratification and equations), I do not explicitly model the proportions of infected individuals who recently acquired and/or transmitted infection via two different partnership types, or two partnerships of the same type. To do so, the required size of the new dimension $\bar{p}$ would be at least 2^P, not $P+1$, where P is the number of different types of partnerships modelled. For transmission via three different partnerships, the required size would be at least 3^P, etc. Indeed, this exponential relationship is related to the challenge of specifying all possible combinations of partnership states in pair-based models [20]. However, under frequentist assumptions, we can equivalently model two transmissions by one individual as one transmission each by two individuals. Thus, we can transfer two individuals from $I_{ih\bar{p}_0}$ to $I_{ih\bar{p}_1}$ and $I_{ih\bar{p}_2}$ (one each) under the $P+1$ stratification, instead of just one individual from $I_{ih\bar{p}_0}$ to $I_{ih\bar{p}_{12}}$ under any of the exponential x^P the stratifications.

In fact, $I_{ih\bar{p}_0}$ can be *negative* (but only for $\bar{p}=0$), because the dimension $\bar{p}$ is only relevant to Eq. (4.12); in all other contexts and equations, we use $I_{ih} = \sum_{\bar{p}} I_{ih\bar{p}}$, which must be positive as usual. Moreover, we can also have $I_{ih\bar{p}} > I_{ih}$, provided that:

$$I_{ih\bar{p}} \leq I_{ih} K_{pi} \tag{4.16}$$

reflecting the situation where 100% of I_{ih} have recently acquired and/or transmitted infection via at least one type-p partnership, or 50% via at least two partnerships, etc. This situation can therefore only arise in the context of multiple concurrent type-p partnerships: $K_{pi} > 1$. If $I_{ih\bar{p}} > I_{ih}$, then $I_{ih\bar{p}_0}$ *must* be negative, but it can be shown that Eq. (4.12) still yields the correct value of $M_{I,pih}$. With this perspective, the constraint in Eq. (4.16) may be intuitive: we cannot "remove" more than the total number of partnerships "offered". This constraint should also be easy to guarantee for small enough timesteps, because $M_{I,pih}$ approaches zero as $I_{ih\bar{p}}$ approaches $I_{ih}K_{pi}$ — i.e. all type-p partnerships become HIV+ concordant, and no more transmission can occur via these partnerships until partners change.

4.3 Experiments

In this section, I describe some simple experiments (methods and results) to highlight differences in the various equations and approaches to modelling HIV transmission in sexual partnerships.

4.3.1 Within- vs. Between-Partnership Heterogeneity

For computing an average per-partnership probability of transmission (B), § 4.1.2 clarified the interpretations of Eq. (4.5) vs. Eq. (4.6) as modelling within-partnership heterogeneity (WPH) vs. between-partnership heterogeneity (BPH), respectively. As shown in § A.3.1 (proof), $B_{\text{WPH}} \geq B_{\text{BPH}}$. Here I explore

under what conditions the ratio $B_{\text{WPH}} \,/\, B_{\text{BPH}}$ is maximized — i.e., when does choosing the correct approach matter most. For simplicity, I considered a single illustrative factor f affecting $\alpha_f \in [0, 1]$ proportion of sex acts ($1 - \alpha_f$ are unaffected), with relative probability of transmission $R_f \in [0.01,\ 10]$. I then computed B_{WPH} and B_{BPH} for $A \in [1, 1000]$ total sex acts, using a base per-act probability of transmission $\beta = 0.34\%$ as a representative value for HIV [11].

Figure 4.5 illustrates four 2-dimensional cross sections of $B(R, \alpha, A)$ under WPH vs. BPH, and the ratio $B_{\text{WPH}}/B_{\text{BPH}}$; the cross sections were at: $A = 32$, $\alpha = 0.5$, $R = 0.1$, and $R = 5$. Based on these results, the difference between approaches can be summarized as:

- negligible for $A < 10$, and small for $A < 100$
- increasing as R gets farther from 1 ($R \to 0$ or $R \to \infty$)
- maximized by specific values of (α, A) for a given R, including $\alpha > \frac{1}{2}$ for $R < 1$, and $\alpha < \frac{1}{2}$ for $R > 1$

The specific values of (α, A) which maximize the difference between approaches for a given R and β create a continuous curve (Figure 4.6), which slowly tends towards $\alpha \to 1, A \to \infty$ as $R \to 0$, and $\alpha \to 0, A \to 0$ as $R \to \infty$. The curve is sigmoidal for log-transformed A, and shifts left with increasing β. I did not derive an analytical expression, but it should be possible to do so. In the context of HIV, the difference between approaches would be larger for protective factors (e.g., condoms) affecting most of a large number of sex acts ($\alpha > 1000$); and likewise larger for risk-increasing factors (e.g., anal sex) affecting a minority of a moderate number of sex acts ($\alpha \approx 100$).

4.3.2 Partnership Durations

As described in § 4.1.3, multiple prior models have implicitly assumed a maximum partnership duration $\delta \leq 1$ year. As such, the adjustment for "wasted contacts" Eq. (4.4) would have less effect. This reduced effect can be quantified via the effective probability of transmission per sex act β' — i.e., tangent slopes in Figure 4.1a — defined as:

$$\beta' = \frac{B}{A} = \frac{1 - (1 - \beta)^A}{A} \tag{4.17}$$

Figure 4.7 illustrates the 1-year β'_1 vs. true-duration β'_δ, for different partnership durations $\delta \in [1, 30]$ and sex frequencies $F \in [1,\ 180]$ per year. Assuming $\delta \leq 1$ can considerably increase the modelled rate of transmission for partnerships with $F \geq 52$ (i.e., weekly) and a true duration $\delta \geq 5$ years, including up to *8-fold* difference with $F \approx 100$ and $\delta \approx 30$. Thus, prior models using $\delta \leq 1$ may have substantially overestimated the relative contribution of longer partnerships with frequent sex — including main/spousal partnerships — to overall transmission.

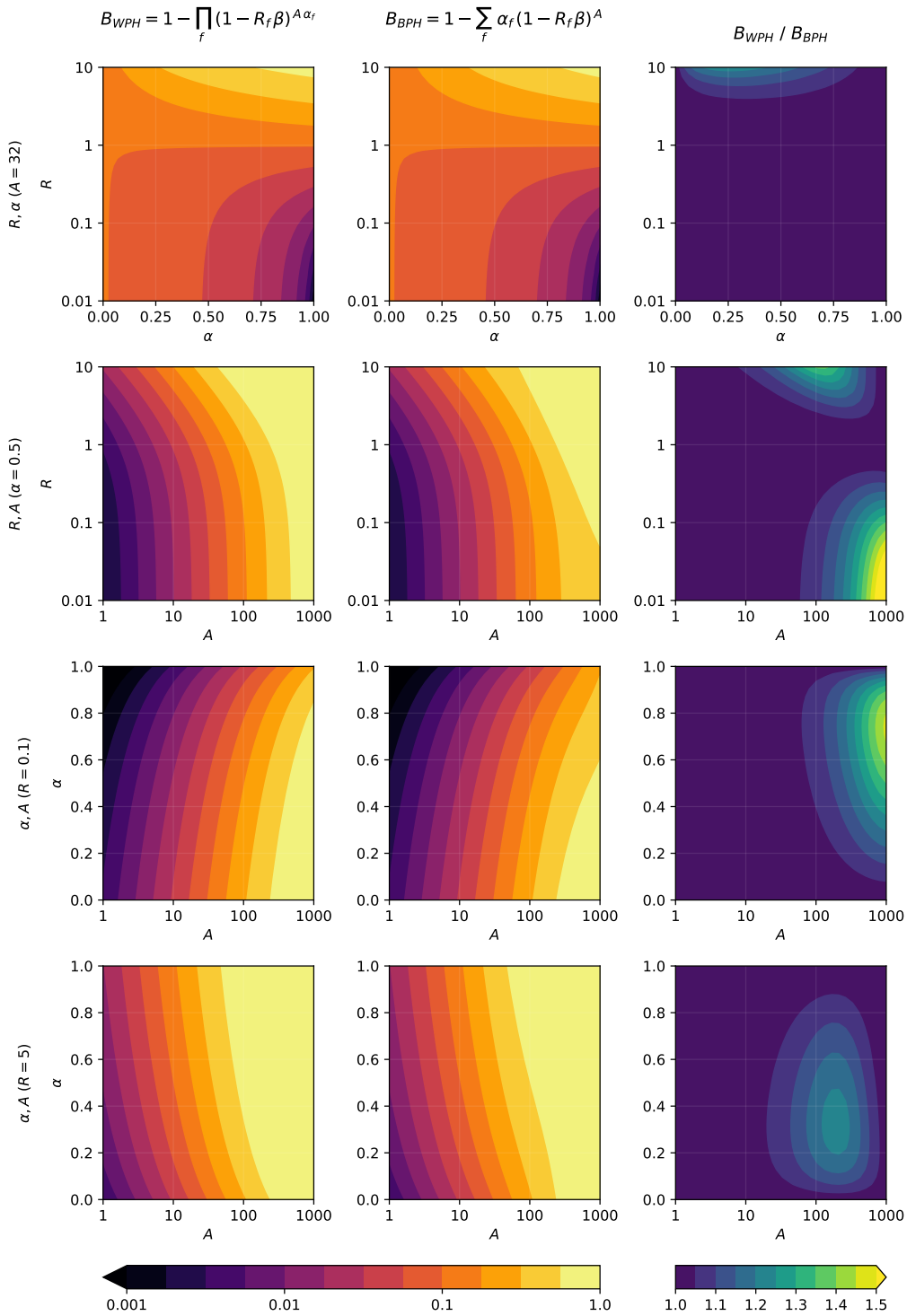


Figure 4.5: Average per-partnership probability of transmission B given heterogeneity in the per-act probability of transmission β within vs. between partnerships

B: probability of transmission per partnership (log scale colourmap); β = 0.34%: probability of transmission per sex act (fixed) [11]; A: total sex acts per partnership (log scale); α_f: proportion of sex acts affected by factor f (linear scale); R_f: relative β given factor f (log scale); WPH: within-partnership heterogeneity; BPH: between-partnership heterogeneity.

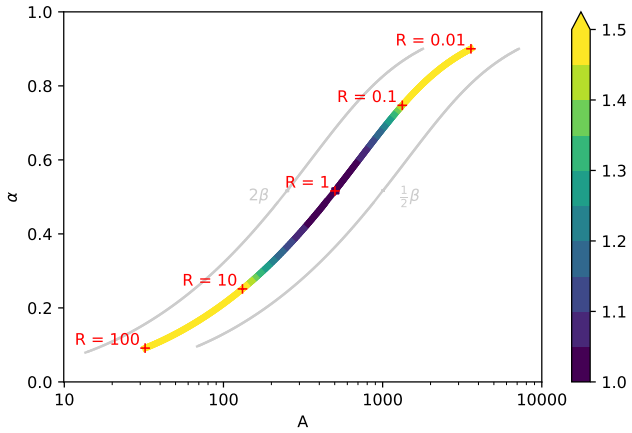


Figure 4.6: Parameter values (α, A) which maximize the difference between the average per-partnership probability of transmission given within- vs. between-partnership heterogeneity

B_{WPH} / B_{BPH}: line colour; β = 0.34%: probability of transmission per sex act (fixed) [11]; A: total sex acts per partnership (log scale); α_f: proportion of sex acts affected by factor f (linear scale); R_f: relative β given factor f (log scale); gray lines denote equivalent contours for 2β and $\frac{1}{2}\beta$.

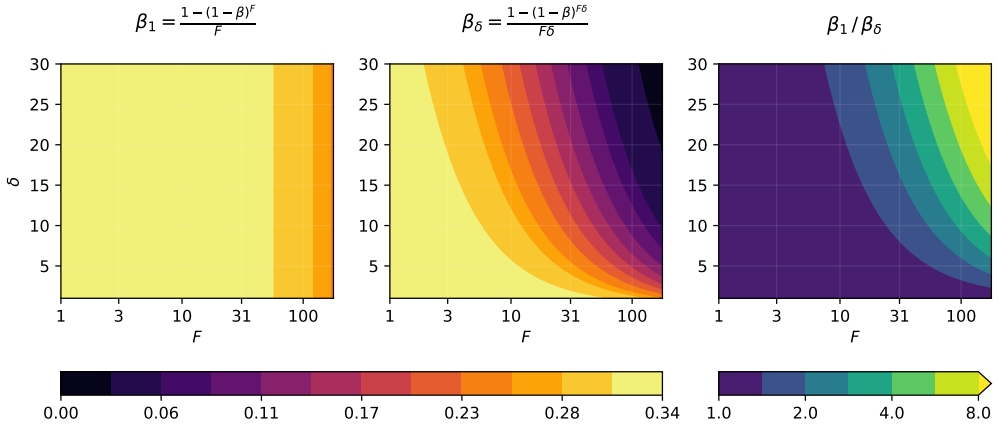


Figure 4.7: Effective probability of transmission per sex act over 1 year vs. total partnership duration

β = 0.34%: probability of transmission per sex act (fixed) [11]; F: frequency of sex per partnership (per year, log scale); δ: partnership duration (years, linear scale); β_1, β_δ: effective probability of transmission per sex act, for 1 year vs. total partnership duration, respectively.

4.3.3 Comparing Approaches in a Complete Model

Next, I sought to explore how different approaches to modelling HIV transmission via sexual partnerships — i.e., the force of infection — influence key outputs from a complete model. For this analysis, I focused on two aspects of prior approaches: whether incidence is aggregated across partnerships as a rate vs. proportion (§ 4.1.3), and whether or not partnership durations are effectively capped at 1 year (§ 4.1.4). Thus, I considered 4 prior approaches, plus the new proposed approach from § 4.2 (Table 4.1). I integrated each approach within the model from Chapter **??**.

Table 4.1: Compared approaches to modelling HIV transmission via sexual partnerships

ID	Name	Key Eqs.	Key Parameters
NP	New Proposed	(4.10)–(4.15)	K, F, δ
RD	Rate-Duration	(4.6), (4.8)	A, Q
RY	Rate-1-Year	(4.6), (4.8)	A_1, Q_1
PD	Proportion-Duration	(4.6), (4.9)	$A, Q,$
PY	Proportion-1-Year	(4.6), (4.9)	A_1, Q_1

K: number of concurrent partners; F: frequency of sex per partnership; δ: partnership duration; $A = F\delta$: total sex acts per partnership; $Q = K/\delta$: partnership formation rate; $A_1 = F\delta_1$, $Q_1 = K/\delta_1$, where $\delta_1 = \min(\delta, 1)$.

I then explored selected model outputs under each of the 5 approaches, with the aim of characterizing:

1. fundamental differences in transmission dynamics under each approach
2. differences in the model-estimated prevention priorities under each approach

For aim 1, I compared HIV incidence (overall and group-specific) using *equal* model parameters across approaches. For aim 2, I compared the transmission population attributable fraction (TPAF, details in § 4.3.3.2) of several transmission pathways, using *approach-specific* model parameters (recalibrated). A transmission pathway could reflect a given partnership type, or infections acquired among and/or transmitted from a given risk group, etc. Since applied models are usually calibrated to a given context, aim 2 thus provides a realistic comparison of how prevention priorities could differ when informed by models using each approach.

Equal vs. Approach-Specific Parameters. The new proposed (NP) approach uses the numbers of concurrent partnerships K, frequency of sex per partnership F, and partnership duration δ, while prior approaches use the total numbers of sex acts per-partnership A, and partnership formation rate Q. For *equal* model parameters, I used model fits (parameter sets) from NP (see § **??**), and converted $A = F\delta$ and $Q = K/\delta$ for all 4 prior approaches (RD, RY, PD, PY), with the additional adjustment $\delta_1 = \min(\delta, 1)$ for the 1-year approaches (RY, PY). For *approach-specific* parameters, I repeated the methods in § **??**, yielding 1000 unique model fits (parameter sets) for each prior approach.

4.3.3.1 Transmission Dynamics using Equal Parameters

Figure 4.8 illustrates HIV incidence under each approach among FSW, clients, and everybody else ("lower risk"). Specifically, Figure 4.8a illustrates incidence per person-year (NP repeated for comparison) and

Figure 4.8b illustrates relative differences vs. the NP approach.[12] I made the following observations — and *hypothesized* explanations, drawing on the complete network of modelled transmission under NP (Figure **??**):

- Incidence among lower risk was generally much higher under 1-year approaches (RY, PY) — underestimation of "wasted contacts" under these approaches disproportionately increases transmission via main/spousal partnerships, allowing more transmission to/from lower risk individuals, including a positive feedback loop via increasing HIV prevalence given like-width-like mixing (see § **??**).
- Incidence differences between the 1-year approaches (RY, PY) vs. NP grew over time — NP explicitly models the accumulation of seroconcordant partnerships wherein contacts are "wasted", or "partnership-level herd effects"; thus, by underestimating "wasted contacts" throughout the epidemic, the 1-year approaches are initially less biased vs. NP, but later overestimate incidence.
- Incidence among all risk groups was lower under full-duration approaches (RD, PD) — complete and instantaneous accounting of "wasted contacts" under these approaches effectively delays transmission in all partnership types, and contributes to a lower HIV prevalence feedback loop.
- Incidence among FSW and clients was much lower under "incidence proportion" approaches (PD, PY) — incidence proportion Eq. (4.9) treats all transmission risks as competing, and notably forces incidence $\lambda^{\text{IP}} \leq 1$, disproportionately reducing incidence among those at highest risk.
- Incidence among FSW and clients under RY approximately matched NP — competing biases due to underestimation of some "wasted contacts" (delays transmission), but not all "wasted contacts" (overestimates transmission) coincidentally yielded incidence roughly matching NP.

Overall, RD resulted in the most *consistent* bias vs. NP across groups and over time.

Figure 4.9 further illustrates the proportions of modelled yearly HIV infections transmitted via different partnership types under each approach using equal parameters (top) and approach-specific parameters for comparison (bottom). As hypothesized above, for equal parameters, the 1-year approaches (RY, PY) featured the greatest proportions of transmission via main/spousal partnerships, and the least via sex work. By contrast, the full-duration approaches (RD, PD) featured the smallest proportions transmitted via main/spousal partnerships, and the most via sex work. The distribution under the proposed approach (NP) was in between these two extremes. Interestingly, there were minimal differences in Figure 4.9 between incidence rate (RD, RY) and incidence proportion (PD, PY) approaches.

4.3.3.2 Prevention Priorities using Approach-Specific Parameters

Many models applied to assess HIV prevention priorities explicitly model specific intervention scenarios, often with a cost dimension [27–29]. However, these context-specific intervention and cost details require additional analyses and/or assumptions, and only explore a subset of the modelled transmission pathways. By contrast, the TPAF reflects an "intervention agnostic" measure of how any given transmission pathway contributes to transmission overall.

Transmission Population Attributable Fractions (TPAFs). Classic PAFs aim to quantify the relative contribution of a specific factor to a given outcome at the population level, by comparing the number of

[12] Relative differences were "paired" according to each parameter set k, and computed as $(\text{XX}_k - \text{NP}_k)/\text{NP}_k$.

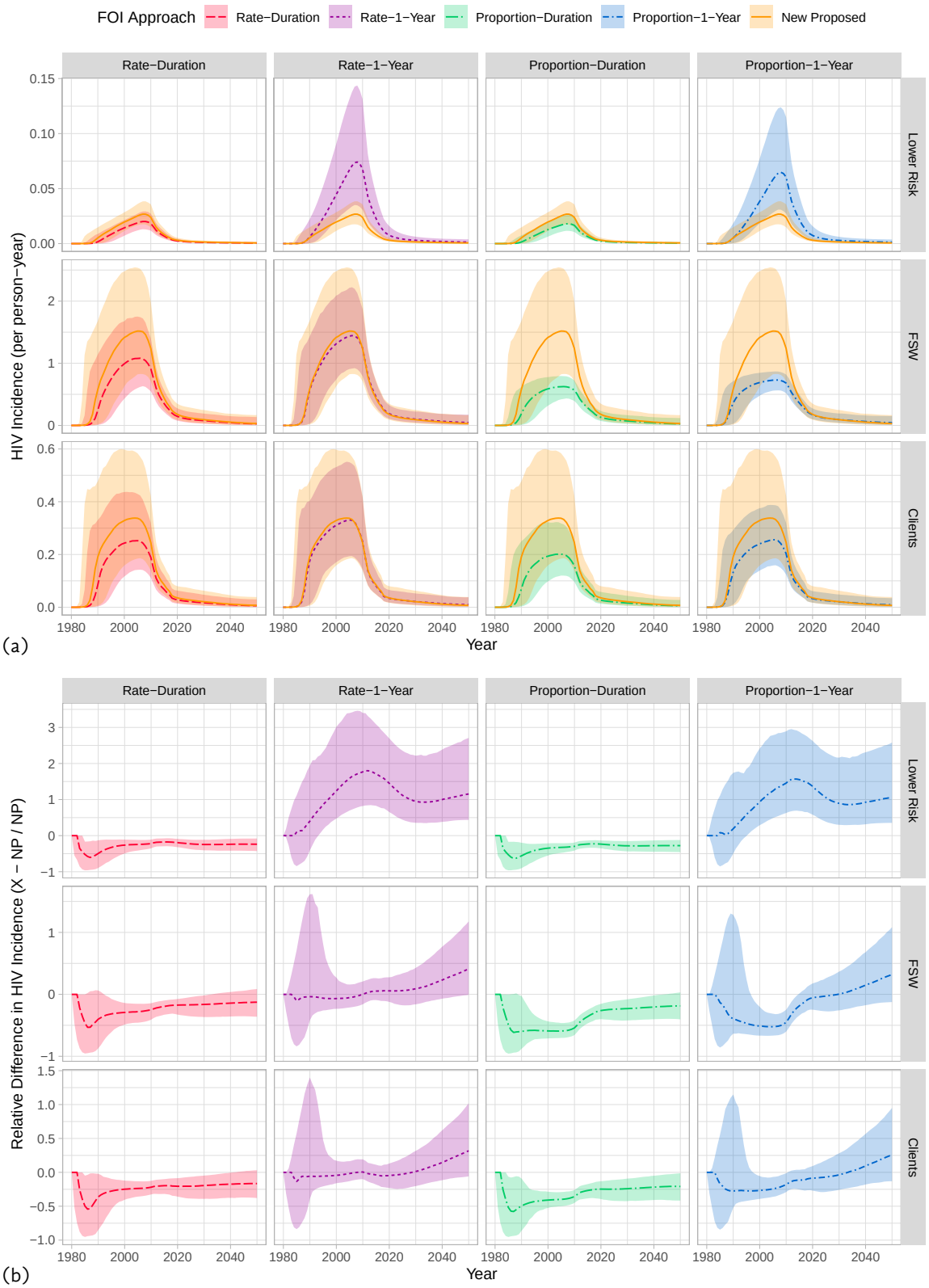


Figure 4.8: HIV incidence among selected risk groups, estimated under different prior force of infection approaches (colours) vs. the new proposed approach using equal model parameters

See Table 4.1 for approach definitions; Lower Risk: all women and men not involved in sex work; FSW: female sex workers; Clients: of FSW; ribbon and curve: range and median of model fits.

Figure 4.9: Proportions of modelled yearly HIV infections transmitted via different partnership types in Eswatini estimated under different force of infection approaches (horizontal facets) with equal vs. recalibrated parameters (vertical facets)

See Table 4.1 for approach definitions; median numbers of infections across all model fits shown.

outcomes with vs. without the factor [30, 31]. However, classic PAFs are not well-suited for infectious diseases, especially over longer time horizons, because they fail to capture the nonlinear dynamics of indirect transmission [32]. In some cases, preventing one transmission event could avert numerous downstream infections; in other cases, preventing one transmission event might only delay infection a short time for an individual at high risk. The TPAF for infectious diseases was developed as a better measure of the contribution of different transmission pathways to overall transmission [33–35].

The TPAF among population j of transmission pathway k is defined as the relative difference in cumulative infections Ω among j since a given time t_0 with vs without transmission via k:

$$\text{TPAF}_{jk}(t) = \frac{\Omega_j(t) - \Omega_{jk}(t)}{\Omega_j(t)}, \qquad \Omega_{jk}(t) = \int_{t_0}^{t} \Lambda_{j,M_k=0}(\tau)\, d\tau, \qquad t = t_0 + \Delta_t \tag{4.18}$$

Thus, TPAFs reflect hypothetical interventions with perfect prevention, ignoring practical implementation challenges associated with any real intervention. Like classic PAFs, TPAFs can sum to more than 100% [36, 37].

I computed TPAFs for the following transmission pathways: main/spousal, casual, and sex work (occasional and regular combined) partnership types; and transmission from FSW, clients, and everybody else ("lower risk"). I computed these TPAFs under each of the 5 force of infection approaches, over Δ_t = 1, 3, and 10-year time horizons, starting in t_0 = 1990, 2000, and 2010 (270 total TPAFs). I implemented scenarios without transmission via pathway k using a boolean "mask" applied to the mixing matrix $M_{pii'}$, *after* resolving the values per § **??**, such that mixing patterns were not affected.

TPAFs of Partnership Types. Figure 4.10 illustrates the TPAFs for partnership types. TPAFs generally increased over longer time horizons Δ_t but decreased with calendar year t_0 (n.b. vertical scales), especially for sex work. Such trends are consistent with prior work showing that TPAFs typically decrease as epidemics become more widespread, especially for key populations [38–40].

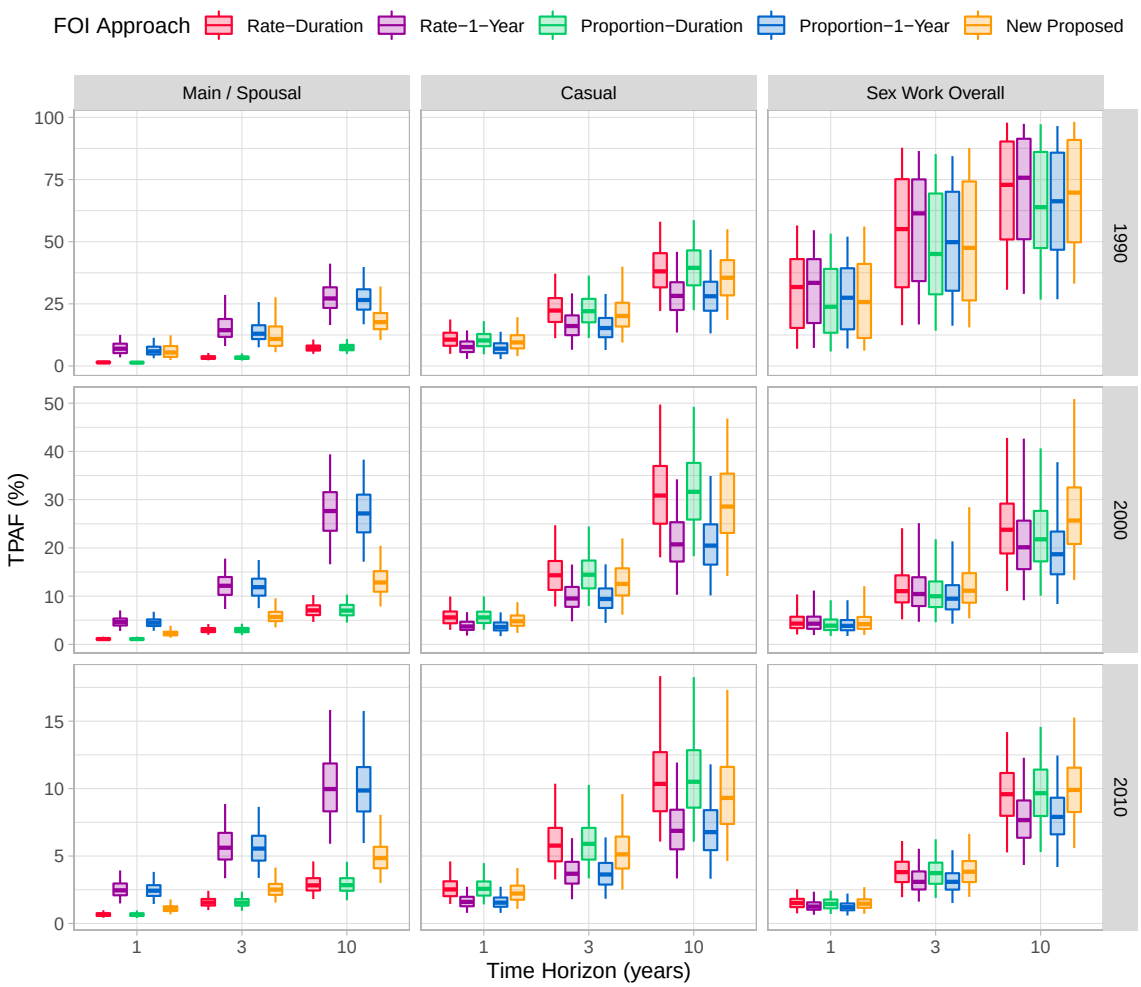


Figure 4.10: TPAF of transmission via different partnership types (horizontal facets), starting from different t_0 (vertical facets), estimated under different force of infection approaches (colours)

TPAF: transmission population attributable fraction, § 4.3.3.2; whiskers, boxes, and midlines: 95% CI, 50% CI, median of model fits.

TPAF differences across force of infection approaches were largest for main/spousal partnerships, with 1-year approaches (RY, PY) significantly larger than full-duration approaches (RD, PD). Main/spousal TPAFs under NP were similar to RY and PY from 1990 over short time horizons, beyond which TPAFs tended towards RD and PD; similar to incidence differences in § 4.3.3.1, these findings reflect reduced transmission via longer partnerships over time due to the accumulation of "partnership-level herd effects", which are only captured under NP. Relative differences in casual partnership TPAFs across approaches were exactly opposite (but less pronounced) vs. differences in main/spousal partnership TPAFs. As such, casual TPAFs were always greater than main/spousal TPAFs under RD and PD, while main/spousal TPAFs were almost always greater than casual TPAFs under RY and PY. There were minimal differences in TPAFs for main/spousal or casual partnerships *between* 1-year approaches (RY vs. PY) or between full-duration approaches (RD vs. PD).

The TPAFs of sex work were the largest of all partnership types in 1990, reflecting the disproportionate role of sex work in early transmission (Figure **??**). In 2000 and 2010, the size of sex work TPAFs remained

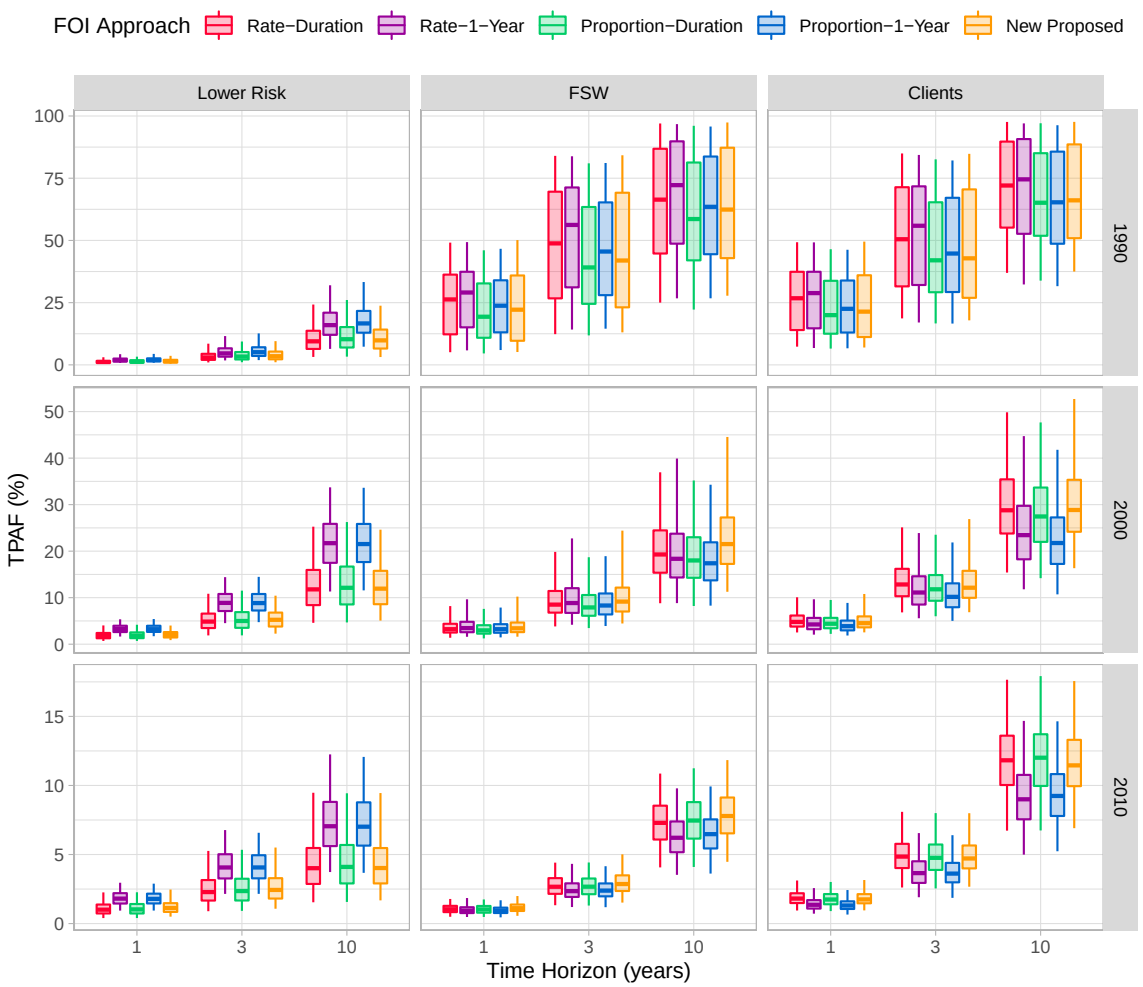


Figure 4.11: TPAF of transmission from different risk groups (horizontal facets), starting from different t_0 (vertical facets), estimated under different force of infection approaches (colours)

TPAF: transmission population attributable fraction, § 4.3.3.2; whiskers, boxes, and midlines: 95% CI, 50% CI, median of model fits.

similar to those of casual partnerships. Interestingly, differences in sex work TPAFs across force of infection approaches were less pronounced vs. for other partnership types. These smaller differences could be explained by two factors. First, HIV prevalence may saturate among FSW and their clients, such that the rate of new infections is mainly influenced by "supply" of susceptible individuals via risk group turnover, and less by differences between force of infection approaches. Second, modelled regular sex work partnership durations were not especially long or short, with posterior distribution spanning 0.5–2.0 years (see § **??** and Figure **??**). By 2010, clearer differences across approaches in sex work TPAFs emerged, which were similar to differences for casual partnerships.

TPAFs of Risk Groups. Figure 4.11 illustrates the TPAFs for transmission from selected risk groups, which can be related to the goal of perfect viral suppression via ART. In 1990, the TPAFs of FSW and clients were largest and overall similar to TPAFs of sex work partnerships. In 2000 and 2010, the TPAFs of FSW and clients remained similar to sex work partnerships, but TPAFs for clients vs. FSW grew larger and with more pronounced differences between force of infection approaches. Such results reflect the greater

onward transmission potential of clients vs. FSW due to their larger population size and greater number of casual partnerships (e.g., Figure **??**). TPAFs of lower risk groups (not engaged in sex work) had differences across approaches that were similar to those for main/spousal partnerships but less pronounced. These differences reflect the strong influence of the 1-year partners duration assumption (i.e., approaches RY, PY) on the contribution of lower risk groups to overall transmission.

4.4 Discussion

Compartmental models of sexual HIV transmission continue to support HIV epidemic response globally, including the Spectrum suite of models [41, 42], the Optima HIV model [28, 43], the Asian Epidemic Model [44], the Thembisa model for South Africa [45–47], and numerous others (e.g., Chapter **??**). Such models usually simulate sexual HIV transmission within risk- and/or age-stratified populations, possibly considering multiple partnership types and/or transmission modifiers. As I have shown in § 4.1, several existing approaches (model structures and equations) are used to define rates of HIV transmission via sexual partnerships in these models — especially with respect to heterogeneous populations, partnerships, and sex acts — each with implicit assumptions. In § 4.3, I explored the potential influence of these approaches/assumptions on: the computed probability of transmission per partnership, modelled epidemic dynamics, and on model-estimated prevention priorities. Yet, many of these assumptions can be avoided altogether using a new approach I developed in § 4.2, representing an exciting opportunity to improve the quality of compartmental model-based evidence for HIV response going forward.

4.4.1 Heterogeneity in Per-Act Probability of Transmission

In § 4.1.2 and 4.3.1, I introduced and explored the distinction between *within*-partnership heterogeneity (WPH) vs. *between*-partnership heterogeneity (BPH) in the per-act probability of transmission β. Whereas WPH reflects an assumption that all partnerships are identical, but comprise heterogeneous acts, BPH reflects an assumption that partnerships are different, but each comprise identical acts. This distinction — i.e., all of Eqs. (4.4)–(4.7) and variants thereof — is unnecessary under the proposed approach.

Numerous variations on Eqs. (4.4)–(4.9) have been used in prior models (see § **??**). For example, the Optima [28] and Goals [48] models aggregate heterogeneity due to HIV infection stage *before* aggregating sex acts within each partnership or applying transmission modifiers. Such an approach is difficult to justify, because the prevalence of each infection stage evidently reflects distinct individuals — *not* the distribution of infection stages within a given partnership (see also footnote 7.) This approach then does not allow for the *multiplicative* interaction of infection stage and other modifiers, while simultaneously allowing a high-infectivity stage with low prevalence (e.g., acute HIV) to increase transmission risk across all partnerships (see Figure 4.1b). As a result, this approach yields intermediate $B_{\text{WPH}} \geq B' \geq B_{\text{BPH}}$ (see § A.3.1). It's not clear whether these variations have systematically biased existing model-based evidence, but improved understanding of the assumptions and potential biases of each approach can help guide interpretation of existing results, and design of future models which do not adopt the proposed approach.

In some cases, modellers have noted the discrepancies between equations between models, but dismissed the differences as inconsequential because β is usually small [28]. This justification is fair when β is indeed

small. However, several combinations of transmission modifiers (e.g., condomless anal sex with GUD and acute HIV infection) [11, 49] can easily yield larger β, for which the discrepancies are *not* inconsequential (e.g., Figure 4.5). In fact, it is precisely these contexts of rapid transmission which define key epidemic dynamics, as reflected in core group theory [50]. Moreover, since the prevalence of such modifiers often varies across risk groups and transmission pathways, differences in how heterogeneous β is aggregated may ultimately yield differences in the modelled contribution of risk groups and transmission pathways to overall transmission — although some differences might be reduced via model calibration.

Lastly, statistical inference on modifiers of per-act transmission probability — e.g., relative risk with condoms, GUD, etc. — typically uses exposure-stratified individual-level data [9, 11, 51, 52]. Thus, these statistical models do not consider what *proportion* of sex acts are exposed, and need not distinguish between within vs. between partnership heterogeneity. Yet, relative risks estimated from *per-act* data have been applied to the *per-partnership* transmission probability in several models [**TOOD**]. Such an approach would then underestimate the impact of risk-reducing modifiers (e.g., condoms) and overestimate the impact of risk-increassing modifiers (e.g., GUD).[13]

4.4.2 Beyond Instantaneous Partnerships

The 2021 review by Rao et al. [13] summarizes frameworks that have been used to simulate partnership dynamics for modelling sexually transmitted infections (see also § 4.1.6 and Appendix 1 of [19]). Besides pair-based models, the review does not identify another approach which has extended the compartmental modelling framework beyond instantaneous partnerships. Pair-based models have not seen widespread adoption, likely due to exponential complexity (i.e., numbers of required compartments) [20]. However, several hybrid models have been developed [53, 54] wherein long-term pairs are explicitly modelled, but additional "one-off" partnerships are modelled as instantaneous. When long-term partnership concurrency is low, such hybrid approaches likely offer substantial improvements over fully instantaneous partnerships [55–57]. However, the high number of *regular* clients reported by Swati FSW (§ **??**) reflects precisely the kind of dense, persistent sexual network — i.e., high concurrency — which is difficult to model via a pair-based approach. The importance of partnership concurrency in HIV transmission has been debated extensively [21, 58–61]. Thus, the proposed approach offers an alternative to hybrid / pair-based models for such networks, and thereby solves a 30-year old modelling challenge [14].

4.4.2.1 Prior Comparisons of Models with vs. Without Instantaneous Partnerships

The potential biases associated with instantaneous partnerships have been explored previously, via comparison with deterministic pair-based models [55–57], a stochastic pair-based model [56], a stochastic static network-based model [56], and a stochastic dynamic network-based model [19].

Kretzschmar and Dietz [55] highlight how biases associated with instantaneous partnership increase with the true partnership durations, and conclude that: "*the number of new partners per unit time is not sufficient to predict the course of the epidemic, but that partnership duration is a quantity that is equally*

[13] Modifying the transmission probability via R — per-act: $B_a = (1 - (1 - R\beta)^A)$ vs. per-partnership: $B_p = R\,(1 - (1 - \beta)^A)$; thus: $B_a > B_p$ if $R < 1$, and $B_a < B_p$ if $R > 1$.

influential." I agree and regret that data to directly inform sexual partnership durations, especially for non-marital partnerships, remain lacking (see § **??**). Efforts to fill this data gap will likely benefit from careful consideration of different measurement approaches and sources of error [62], perhaps in conjunction with efforts to better quantify partnership formation rates, as explored in § A.2.3.

Eames and Keeling [56] and Lloyd-Smith, Getz, and Westerhoff [57] both show that instantaneous partnerships can result in overestimation of the initial epidemic growth rate and equilibrium prevalence. Such finding seem intuitive. However, in § 4.3.3 (Figure 4.8), I showed how the rate of epidemic growth under instantaneous partnerships strongly depends on the effective partnership duration used for the "wasted contacts" adjustment — if such an adjustment is applied at all. That is, when durations were capped at 1 year (approaches RY, PY), this adjustment likely had little effect, and modelled incidence was indeed overestimated relative to the proposed approach; by contrast, when full partnership durations were used (approaches RD, PD), this adjustment reduced transmission immediately in anticipation of future "wasted contacts", and modelled incidence was *underestimated* relative to the proposed approach. No adjustments for "wasted contacts" were described in [56, 57], reflecting the former case.

Johnson and Geffen [19]. This landmark study compared modelling frameworks across 6 STIs.[14] The models explored were more complex than previous works, including: population stratification by sex, age, and risk, and three partnership types within a dynamic sexual network, in a South African context. Although an adjustment for "wasted contacts" for instantaneous partnerships was applied, the adjustment considered different time period across partnership types: 1 month for main/spousal, 6 months for casual, and none for sex work partnerships; thus, regular sex work partnerships were not considered. Similar to experiments in § 4.3.3, [19] first compared model outputs from frameworks with equal parameters, and then again with recalibrated parameters.

With equal parameters, findings echoed those above [56, 57], although differences between frameworks were larger for curable STIs with faster transmission, and smaller for HIV. After recalibrating models to the same STI data in, the best-fitting parameters differed significantly across modelling frameworks, as expected — e.g., lower transmission probabilities were inferred with instantaneous partnerships. More importantly, the relative impact (infections averted after 10 years) of several illustrative intervention strategies also differed substantially. These differences were summarized as: "[instantaneous partnership models] *are likely to underestimate the importance of interventions that are targeted at high-risk groups, while overestimating the impact of interventions targeted at low-risk groups*" [19]. Such findings are similar to those in § 4.3.3.2, where the TPAFs of lower risk populations and main/spousal partnerships were overestimated, while the TPAFs of casual partnerships were underestimated under 1-year approaches (RY, PY) vs. the new proposed approach (NP). However, while [19] observed that instantaneous partnerships "*may underestimate the contribution of commercial sex*", the TPAFs of FSW, clients, and sex work partnerships were similar across approaches in § 4.3.3.2. These different findings could be explained by the fact that [19] only modelled one-off sex work partnerships, whereas transmission via sex work in the Eswatini model was dominated by regular partnerships (Figure 4.9). Indeed, many HIV models have yet to incorporate longer partnerships between higher risk groups (see § **??** and § **??**). Thus, I would offer to rephrase the conclusions of [19] as: "*without complete adjustment for 'wasted contacts', models with*

[14] The term "frequency-dependent" in [19] is synonymous with "instantaneous partnerships" here.

instantaneous partnerships may overestimate the contribution of longer partnerships, and underestimate the contribution of shorter partnerships."

4.4.2.2 Implications for Existing Model-Based Evidence

The vast majority of existing compartmental HIV transmission models — including all those listed at the start of this chapter — have used an instantaneous partnerships approach, with adjustments for "wasted contacts" of 1-year or less. The results in § 4.3.3.2 suggest that such models have likely *systematically* and *significantly* overestimated the relative contributions of longer vs. shorter partnerships to overall transmission. Such results are corroborated by [19], and have substantial implications for the existing body of model-based evidence. That is, models continue to help inform which interventions are prioritized and for whom, and existing evidence may overestimate the importance of prevention within longer partnerships for reducing overall transmission. For example:

Anderson et al. [27, 63, 64] These works explore "optimal" combinations of PrEP, early ART, behaviour change, and VMMC for MSM, other men, FSW, and other women in Kenya, under various cost constraints. Their modelling analyses indicate that early ART for non-MSM men is usually more cost effective than PrEP for FSW. Yet, for lowest risk men — modelled median [IQR] 53 [37, 70]% of non-MSM men — effective partnership duration was 4 [2.6, 6.2] years.[15] If longer partnership durations were more accurately modelled, the relative impact of PrEP for FSW vs. early ART for non-MSM would likely increase.

Optima HIV Model. [28, 43, 65, 66] As the name suggests, this model has similarly been applied to "optimize resource allocation" in over 20 countries. Recommended allocations have generally increased resources for key populations over current spending. However, the force of infection equation is defined as an incidence proportion [66] — Eq. (4.9) — wherein all sex acts over a given time period (default Δ_t = 0.2 years)[16] are modelled as competing risks, and so partnership durations are *completely ignored*. Thus, the relative impact of prioritizing key populations has likely been underestimated via both: overestimation of transmission via longer partnerships due to ignoring partnership durations, underestimation of transmission to higher risk groups due to the incidence proportion equation.

Goals Model. [41, 48] The Goals Model is part of the Spectrum suite of policy modelling tools [42], which have been widely applied in consultation with national ministries of health to estimate yearly new infections and the impact of various interventions [41]. The Goals model includes mechanistic HIV transmission among 11 total risk groups, including FSW, their clients, MSM, and PWID, without age stratification [48].[17] Yet, as in the Optima model, the force of infection is defined as an incidence proportion — using a transformation of Eq. (4.9) — with Δ_t = 1 year. Thus, for the same reasons as Optima (and worse with Δ_t = 1 vs. 0.2 years), the relative impact of prioritizing shorter partnerships and key populations for prevention have likely been systematically underestimated by the Goals Model.

In sum, decades of model-based HIV prevention evidence — for multiple countries and resource allocation questions — are built upon the instantaneous partnerships assumption and associated equations. I have

[15] Median [IQR] estimated via Monte Carlo sampling of $\bar{c}, R, \varpi$ from uniform prior distributions in Table S7; posterior parameter distribution were not given in [27].

[16] From: github.com/optimamodel/optima/blob/master/optima/parameters.py

[17] An age-stratified variant of Goals was recently developed for generalized epidemics, which subsumes key populations as proportions of age strata; the new model is called the Goals "Age-Stratified Model" (ASM), whereas the original model is now called the Goals "Risk-Stratified Model" (RSM) [41].

shown that this assumption and these equations can significantly bias model-estimated contributions of different populations and partnerships to overall transmission, and thus the model-estimated importance of different prevention strategies. Specifically, the common practice of only adjusting for up to 1-year of "wasted contacts" (or equivalently: assuming that all individuals change partners at least once per year) overestimates the importance of prevention in longer partnerships, and underestimates the importance of prevention in shorter partnerships. Moreover, defining the force of infection as an incidence proportion (vs. rate) disproportionately reduces modelled incidence among populations at higher risk, and thus underestimates the importance of prevention among these populations. Finally, I illustrated these potential biases in the context of a high prevalence HIV epidemic (Eswatini), but these biases and implications for prevention could be even greater elsewhere.

4.4.3 Transmission-Driven Emergence of Serosorting Patterns

HIV serosorting is a controversial harm reduction strategy defined as preferential selection of sexual partners with matching (perceived) HIV serostatus; related strategies can involve modified sexual practices with a given partner on the same basis [67–69]. Serosorting is often quantified using cross-sectional data, using the odds or excess fraction of seroconcordant partnerships vs. random mixing by serostatus [70, 71]. However, using an illustrative toy scenario (§ 4.2.1, Figure 4.3), I have highlighted how transmission naturally generates a disproportionate number of seroconcordant (*I-I*) partnerships as compared to random mixing by serostatus. This emergent property was also noted in [56], therein described as "*correlation of infection statuses of neighboring individuals*". This disproportionate seroconcordance would be correlated with partnership duration. Failure to consider this dynamic could then lead to overestimation of the degree to which serosorting is intentional (from cross-sectional data). Such biases in quantifying serosorting could be mitigated using longitudinal data or consideration of only new partnerships [72].

4.4.4 Future Work

Hopefully I have made a convincing case for the value added by the proposed approach. Thus, an obvious area for future work would be to integrate this approach into new and existing compartmental HIV transmission models, including widely-used models like Spectrum, Optima, Asian Epidemic Model, and Thembisa. To this end, the complete implementation of the proposed approach in the Eswatini model is available online,[18] although perhaps it would also be useful to develop a simpler example model with vs. without the proposed approach to illustrate the essential differences. It may also be useful to compare key model outputs before vs. after integrating the proposed approach in existing models and highlight notable differences, similar to the experiments in § 4.3.3. Similar work could compare and indeed validate the proposed approach vs. individual-based models, similar to the experiments in [19].

Additionally, while I developed the proposed approach in the context of HIV, accurate simulation of partnership dynamics is also relevant for compartmental models of other sexually transmitted infections [13], including gonorrhea, chlamydia, syphilis, trichomoniasis, herpes, hepatitis, papillomavirus, and mpox. However, the approach should be carefully adapted for curable infections, as I have not considered

[18] See: github.com/mishra-lab/hiv-fsw-art/blob/master/code/model/foi.py, where `foi_mode='base'`.

how transitions *out* of the newly proposed infected strata (stratification $\bar{p}$ in Figure 4.4) should be conceptualized and parameterized.

References

[1] M. Begon et al. "A clarification of transmission terms in host-microparasite models: Numbers, densities and areas". *Epidemiology and Infection* 129.1 (Aug. 2002), pp. 147–153. https://doi.org/10.1017/S0950268802007148.

[2] Jesse Knight and Sharmistha Mishra. "Beyond instantaneous partnerships: capturing partnership-level herd effects in compartmental models of sexually transmitted infections". *SMDM 44th North American Meeting*. Seattle, Washington, 2023. https://github.com/mishra-lab/hiv-fsw-art/tree/master/docs/conf/smdm.

[3] R. M. Anderson et al. "A preliminary study of the transmission dynamics of the Human Immunodeficiency Virus (HIV), the causative agent of AIDS". *Mathematical Medicine and Biology* 3.4 (Jan. 1986), pp. 229–263. https://doi.org/10.1093/imammb/3.4.229.

[4] James A Yorke, Herbert W Hethcote, and Annett Nold. "Dynamics and control of the transmission of gonorrhea". *Sexually Transmitted Diseases* 5.2 (1978), pp. 51–56. http://doi.org/10.1097/00007435-197804000-00003.

[5] Annett Nold. "Heterogeneity in disease-transmission modeling". *Mathematical Biosciences* 52.3-4 (Dec. 1980), pp. 227–240. https://doi.org/10.1016/0025-5564(80)90069-3.

[6] Herbert W. Hethcote, James A. Yorke, and Annett Nold. "Gonorrhea modeling: a comparison of control methods". *Mathematical Biosciences* 58.1 (Feb. 1982), pp. 93–109. https://doi.org/10.1016/0025-5564(82)90053-0.

[7] Valerie Isham. "Mathematical Modelling of the Transmission Dynamics of HIV Infection and AIDS: A Review". *Journal of the Royal Statistical Society. Series A: Statistics in Society* 151.1 (1988), p. 5. https://doi.org/10.2307/2982179.

[8] Edward H. Kaplan. "Modeling HIV infectivity: Must sex acts be counted?" *Journal of Acquired Immune Deficiency Syndromes* 3.1 (1990), pp. 55–61. https://pubmed.ncbi.nlm.nih.gov/2293643/.

[9] Ronald H Gray et al. "Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda". *The Lancet* 357.9263 (Apr. 2001), pp. 1149–1153. https://doi.org/10.1016/S0140-6736(00)04331-2.

[10] John Arne Røttingen and Geoffrey P. Garnett. "The epidemiological and control implications of HIV transmission probabilities within partnerships". *Sexually Transmitted Diseases* 29.12 (Dec. 2002), pp. 818–827. https://doi.org/10.1097/00007435-200212000-00014.

[11] Marie Claude Boily et al. "Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies". *The Lancet Infectious Diseases* 9.2 (Feb. 2009), pp. 118–129. https://doi.org/10.1016/S1473-3099(09)70021-0.

[12] Seyed M. Moghadas et al. "Could condoms stop the AIDS epidemic?" *Journal of Theoretical Medicine* 5.3-4 (Sept. 2003), pp. 171–181. https://doi.org/10.1080/10273660412331315147.

[13] Darcy White Rao et al. "Partnership dynamics in mathematical models and implications for representation of sexually transmitted infections: a review". *Annals of Epidemiology* 59 (July 2021), pp. 72–80. https://doi.org/10.1016/j.annepidem.2021.04.012.

[14] K. Dietz and K. P. Hadeler. "Epidemiological models for sexually transmitted diseases". *Journal of Mathematical Biology* 26.1 (Feb. 1988), pp. 1–25. https://doi.org/10.1007/BF00280169.

[15] Robert Allard. "A family of mathematical models to describe the risk of infection by a sexually transmitted agent". *Epidemiology* 1.1 (Jan. 1990), pp. 30–33. https://doi.org/10.1097/00001648-199001000-00007.

[16] Foteini K Giannou et al. "Condom effectiveness in reducing heterosexual HIV transmission: a systematic review and meta-analysis of studies on HIV serodiscordant couples". *Expert Review of Pharmacoeconomics and Outcomes Research* 16.4 (July 2016), pp. 489–499. https://doi.org/10.1586/14737167.2016.1102635.

[17] J. M. Aldaz. "Self-improvement of the inequality between arithmetic and geometric means". *Journal of Mathematical Inequalities* 2 (2009), pp. 213–216. https://doi.org/10.7153/jmi-03-21.

[18] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.

[19] Leigh F. Johnson and Nathan Geffen. "A Comparison of two mathematical modeling frameworks for evaluating sexually transmitted infection epidemiology". *Sexually Transmitted Diseases* 43.3 (Mar. 2016), pp. 139–146. http://doi.org/10.1097/OLQ.0000000000000412.

[20] Mirjam Kretzschmar and Janneke C.M. Heijne. "Pair formation models for sexually transmitted infections: A primer". *Infectious Disease Modelling* 2.3 (Aug. 2017), pp. 368–378. https://doi.org/10.1016/j.idm.2017.07.002.

[21] Larry Sawers. "Measuring and modelling concurrency". *Journal of the International AIDS Society* 16.1 (Jan. 2013), p. 17431. http://doi.wiley.com/10.7448/IAS.16.1.17431.

[22] Lorenzo Pellis et al. "Eight challenges for network epidemic models". *Epidemics* 10 (Mar. 2015), pp. 58–62. https://www.sciencedirect.com/science/article/pii/S1755436514000334#bib0190.

[23] Jan A.C. C. Hontelez et al. "Elimination of HIV in South Africa through Expanded Access to Antiretroviral Therapy: A Model Comparison Study". *PLOS Medicine* 10.10 (Oct. 2013). Ed. by Nathan Ford, e1001534. http://doi.org/10.1371/journal.pmed.1001534.

[24] C. Marijn Hazelbag et al. "Calibration of individual-based models to epidemiological data: A systematic review". *PLOS Computational Biology* 16.5 (May 2020). Ed. by Roger Dimitri Kouyos, e1007893. https://doi.org/10.1371/journal.pcbi.1007893.

[25] Samuel M. Jenness, Steven M. Goodreau, and Martina Morris. "EpiModel: An R package for mathematical modeling of infectious disease over networks". *Journal of Statistical Software* 84 (2018). /pmc/articles/PMC5931789/%20/pmc/articles/PMC5931789/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5931789/.

[26] Duncan A. Clark and Mark S. Handcock. "Comparing the real-world performance of exponential-family random graph models and latent order logistic models for social network analysis". *Journal of the Royal Statistical Society. Series A: Statistics in Society* 185.2 (Apr. 2022), pp. 566–587. https://doi.org/10.1111/rssa.12788.

[27] Sarah Jane Anderson et al. "Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: A modelling study". *The Lancet* 384.9939 (July 2014), pp. 249–256. https://doi.org/10.1016/S0140-6736(14)61053-9.

[28] Cliff C. Kerr et al. "Optima: A Model for HIV Epidemic Analysis, Program Prioritization, and Resource Optimization". *Journal of Acquired Immune Deficiency Syndromes* 69.3 (July 2015), pp. 365–376. https://doi.org/10.1097/QAI.0000000000000605.

[29] Mathieu Maheu-Giroux et al. "Cost-Effectiveness of Accelerated HIV Response Scenarios in Cote d'Ivoire". *Journal of Acquired Immune Deficiency Syndromes* 80.5 (2019), pp. 503–512. https://doi.org/10.1097/QAI.0000000000001956.

[30] B. Rockhill, B. Newman, and C. Weinberg. "Use and misuse of population attributable fractions." *American Journal of Public Healthournal of public health* 88.1 (Oct. 1998), pp. 15–19. https://doi.org/10.2105/AJPH.88.1.15.

[31] Charles Poole. "A history of the population attributable fraction and related measures". *Annals of Epidemiology* 25.3 (Mar. 2015), pp. 147–154. https://doi.org/10.1016/j.annepidem.2014.11.015.

[32] Sharmistha Mishra and Stefan D. Baral. "Rethinking the population attributable fraction for infectious diseases". *The Lancet Infectious Diseases* 20.2 (Feb. 2020), pp. 155–157. https://doi.org/10.1016/S1473-3099.

[33] Kathleen N. Deering et al. "The impact of out-migrants and out-migration on the HIV/AIDS epidemic: A case study from south-west India". *AIDS* 22.S5 (Dec. 2008). https://pubmed.ncbi.nlm.nih.gov/19098476/%20https://journals.lww.com/aidsonline/Fulltext/2008/12005/The_impact_of_out_migrants_and_out_migration_on.15.aspx.

[34] Sharmistha Mishra et al. "Distinguishing sources of HIV transmission from the distribution of newly acquired HIV infections: Why is it important for HIV prevention planning?" *Sexually Transmitted Infections* 90.1 (Feb. 2014), pp. 19–25. http://doi.org/10.1136/sextrans-2013-051250.

[35] Sharmistha Mishra et al. "Data and methods to characterize the role of sex work and to inform sex work programs in generalized HIV epidemics: evidence to challenge assumptions". *Annals of Epidemiology* 26.8 (Aug. 2016), pp. 557–569. https://doi.org/10.1016/j.annepidem.2016.06.004.

[36] Alexander K. Rowe, Kenneth E. Powell, and W.Dana Dana Flanders. "Why population attributable fractions can sum to more than one". *American Journal of Preventive Medicine* 26.3 (Apr. 2004), pp. 243–249. https://www.sciencedirect.com/science/article/pii/S0749379703003775.

[37] Sharmistha Mishra et al. "Estimating the epidemic consequences of HIV prevention gaps among key populations". *Journal of the International AIDS Society* 24.S3 (July 2021), e25739. https://onlinelibrary.wiley.com/doi/full/10.1002/jia2.25739%20https://onlinelibrary.wiley.com/doi/abs/10.1002/jia2.25739%20https://onlinelibrary.wiley.com/doi/10.1002/jia2.25739.

[38] Leigh F. Johnson and Peter J. White. "A review of mathematical models of HIV/AIDS interventions and their implications for policy". eng. *Sexually Transmitted Infections* 87.7 (Dec. 2011), pp. 629–634. https://doi.org/10.1136/sti.2010.045500.

[39] Sharmistha Mishra et al. "Impact of High-Risk Sex and Focused Interventions in Heterosexual HIV Epidemics: A Systematic Review of Mathematical Models". *PLOS ONE* 7.11 (Nov. 2012), e50691. http://doi.org/10.1371/journal.pone.0050691.

[40] Marie Claude Boily et al. "What really is a concentrated HIV epidemic and what does it mean for West and Central Africa? Insights from mathematical modeling". *Journal of Acquired Immune Deficiency Syndromes* 68 (Mar. 2015), S74–S82. https://doi.org/10.1097/QAI.0000000000000437.

[41] John Stover et al. "Modeling the epidemiological impact of the UNAIDS 2025 targets to end AIDS as a public health threat by 2030". *PLOS Medicine* 18.10 (Oct. 2021), e1003831. https://doi.org/10.1371/journal.pmed.1003831.

[42] Avenir Health. *Spectrum Models*. M, 2022. https://avenirhealth.org/software-spectrum.php.

[43] Optima Consortium for Decision Science. *Optima HIV User Guide Vol. VI: Parameter Data Sources*. 2021. http://optimamodel.com/user-guide.

[44] T. Brown and W. Peerapatanapokin. *The Asian Epidemic Model: A process model for exploring HIV policy and programme alternatives in Asia*. Aug. 2004. https://doi.org/10.1136/sti.2004.010165.

[45] Leigh F. Johnson and Rob E. Dorrington. "Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions". *Demographic Research* 14.22 (2006), pp. 541–574. https://doi.org/10.4054/DemRes.2006.14.22.

[46] Leigh F. Johnson et al. "Prospects for HIV control in South Africa: a model-based analysis". *Global Health Action* 9.1 (Dec. 2016). https://doi.org/10.3402/gha.v9.30314.

[47] Leigh F. Johnson et al. "The effect of HIV programmes in South Africa on national HIV incidence trends, 2000-2019". *JAIDS Journal of Acquired Immune Deficiency Syndromes* (June 2022). https://doi.org/10.1097/qai.0000000000002927.

[48] John Stover et al. "How Can We Get Close to Zero? The Potential Contribution of Biomedical Prevention and the Investment Framework towards an Effective Response to HIV". *PLOS ONE* 9.11 (2014), e111956. https://doi.org/10.1371/journal.pone.0111956.

[49] Julie Fox et al. "Quantifying sexual exposure to HIV within an HIV-serodiscordant relationship: Development of an algorithm". *AIDS* 25.8 (May 2011), pp. 1065–1082. https://doi.org/10.1097/QAD.0b013e328344fe4a.

[50] Charlotte Watts et al. "Remodelling core group theory: the role of sustaining populations in HIV transmission". *Sexually Transmitted Infections* 86.S3 (Dec. 2010), pp. iii85–iii92. https://doi.org/10.1136/sti.2010.044602.

[51] Nicholas P. Jewell and Stephen C. Shiboski. "Statistical Analysis of HIV Infectivity Based on Partner Studies". *Biometrics* 46.4 (Dec. 1990), p. 1133. https://doi.org/10.2307/2532454.

[52] Maria J. Wawer et al. "Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda". *Journal of Infectious Diseases* 191.9 (May 2005), pp. 1403–1409. https://doi.org/10.1086/429411.

[53] Maria Xiridou et al. "The contribution of steady and casual partnerships to the incidence of HIV infection among homosexual men in Amsterdam". *AIDS* 17.7 (2003), pp. 1029–1038. https://doi.org/10.1097/00002030-200305020-00012.

[54] Kimberly A. Powers et al. "The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: A modelling study". *The Lancet* 378.9787 (July 2011), pp. 256–268. https://doi.org/10.1016/S0140-6736(11)60842-8.

[55] Mirjam Kretzschmar and Klaus Dietz. "The effect of pair formation and variable infectivity on the spread of an infection without recovery". *Mathematical Biosciences* 148.1 (Feb. 1998), pp. 83–113.

[56] Ken T.D. Eames and Matt J. Keeling. "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases". *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (Oct. 2002), pp. 13330–13335. https://www.pnas.org/content/99/20/13330%20https://www.pnas.org/content/99/20/13330.abstract.

[57] James O. Lloyd-Smith, Wayne M. Getz, and Hans V. Westerhoff. "Frequency-dependent incidence in models of sexually transmitted diseases: Portrayal of pair-based transmission and effects of illness on contact behaviour". *Proceedings of the Royal Society B: Biological Sciences* 271.1539 (Mar. 2004), pp. 625–634. https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2003.2632.

[58] Timothy L. Mah and Daniel T. Halperin. "Concurrent sexual partnerships and the HIV epidemics in Africa: Evidence to move forward". *AIDS and Behavior* 14.1 (Feb. 2010), pp. 11–16. https://doi.org/10.1007/s10461-008-9433-x.

[59] Frank Tanser et al. "Effect of concurrent sexual partnerships on rate of new HIV infections in a high-prevalence, rural South African population: A cohort study". *The Lancet* 378.9787 (July 2011), pp. 247–255. https://doi.org/10.1016/S0140-6736(11)60779-4.

[60] Steven M. Goodreau et al. "Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe". *AIDS and Behavior* 16.2 (Feb. 2012), pp. 312–322. https://link-springer-com.myaccess.library.utoronto.ca/article/10.1007/s10461-010-9858-x.

[61] Marie Claude Boily, Michel Alary, and Rebecca F. Baggaley. "Neglected issues and hypotheses regarding the impact of sexual concurrency on HIV and sexually transmitted infections". *AIDS and Behavior* 16.2 (Feb. 2012), pp. 304–311. https://doi.org/10.1007/s10461-011-9887-0.

[62] Bart Burington et al. "Estimating duration in partnership studies: Issues, methods and examples". *Sexually Transmitted Infections* 86.2 (Apr. 2010), pp. 84–89. https://doi.org/10.1136/sti.2009.037960.

[63] Sarah-Jane Anderson et al. "HIV prevention where it is needed most: Comparison of strategies for the geographical allocation of interventions: Comparison". *Journal of the International AIDS Society* 20.4 (2017), e25020. https://doi.org/10.1002/jia2.25020.

[64] Sarah-Jane Anderson et al. "Frontloading HIV financing maximizes the achievable impact of hiv prevention". *Journal of the International AIDS Society* 21.3 (2018), e25087. https://doi.org/10.1002/jia2.25087.

[65] Robyn M. Stuart et al. *How should HIV resources be allocated? Lessons learnt from applying optima HIV in 23 countries*. Apr. 2018. https://doi.org/10.1002/jia2.25097.

[66] Cliff C. Kerr et al. "Optima HIV Methodology and Approach". *Tackling the World's Fastest-Growing HIV Epidemic: More Efficient HIV Responses in Eastern Europe and Central Asia*. Ed. by Feng Zhao, Clemens Benedikt, and David Wilson. Washington, DC: The World Bank, 2020. https://doi.org/10.1596/978-1-4648-1523-2_ch13.

[67] Kelly Jean Heymer and David P. Wilson. *Available evidence does not support serosorting as an HIV risk reduction strategy*. Mar. 2010. https://doi.org/10.1097/QAD.0b013e328337b029.

[68] Susan Cassels et al. "Available evidence does not support serosorting as an HIV risk reduction strategy: Author's reply". *AIDS* 24.6 (Mar. 2010), pp. 936–938. http://www.ncbi.nlm.nih.gov/pubmed/24850995.

[69] Susan Cassels and David A Katz. "Seroadaptation among men who have sex with men: Emerging research themes". *Current HIV/AIDS Reports* 10.4 (Dec. 2013), pp. 305–313. http://doi.org/10.1007/s11904-013-0188-2.

[70] Susan Cassels et al. "HIV serosorting as a harm reduction strategy: Evidence from Seattle, Washington". *AIDS* 23.18 (Nov. 2009), pp. 2497–2506. http://www.ncbi.nlm.nih.gov/pubmed/19834319.

[71] Linwei Wang et al. "Population-level sexual mixing according to hiv status and preexposure prophylaxis use among men who have sex with men in montreal, canada: Implications for hiv prevention". *American Journal of Epidemiology* 189.1 (Jan. 2020), pp. 44–54. https://doi.org/10.1093/aje/kwz231.

[72] Hae Young Kim et al. "HIV seroconcordance among heterosexual couples in rural KwaZulu-Natal, South Africa: a population-based analysis". *Journal of the International AIDS Society* 23.1 (Jan. 2020), e25432. https://doi.org/10.1002/jia2.25432.

Appendix A

Supporting Mathematics

A.1 Distributions

A.1.1 Fitting Distributions

Uncertainty distributions for all parameters and calibration targets were estimated by fitting a parametric distribution to specified quantiles. Let $f(x \mid \theta)$ be the probability density function of random variable x (parameter or target) given distribution parameters θ. Then $F(x \mid \theta) = \int_0^x f(\tau)\, d\tau$ is the cumulative distribution function, and $Q(p \mid \theta) = F^{-1}(p \mid \theta)$ is the quantile function. Our objective is to estimate θ, given a set of quantiles (e.g., $q = \{q_{2.5}, q_{97.5}\}$ for the 95% CI). For each estimation, I minimized[1] the the following error function:

$$J(\theta) = \sum_i \left| q_i - Q(p_i \mid \theta) \right|^\omega \tag{A.1}$$

where ω can specify absolute differences ($\omega = 1$) or squared differences ($\omega = 2$) to improve convergence. Distribution fit was validated visually using a plot of the distribution quantiles $Q(p_i \mid \theta)$ vs. the target quantiles q_i, overlaid on the density distribution $f(x \mid \theta)$; e.g., Figure A.1.

A.1.2 Beta Approximation of the Binomial (BAB) Distribution

Numerous model parameters and calibration targets represent population proportions. Such proportions can be estimated as $\rho = n/N$, where N is the sample size and n is the number of individuals with the characteristic of interest. The uncertainty around n is then given by the binomial distribution:

$$p(n) = \binom{N}{n} \rho^n (1-\rho)^{N-n} \tag{A.2}$$

However, Eq. (A.2) is only defined for discrete values of n. It is more convenient to have a continuous distribution for ρ, for sampling parameters and evaluating the likelihood of calibration targets, since compartmental models can have non-whole-number population sizes. For this purpose, I use a beta

[1] Using docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html

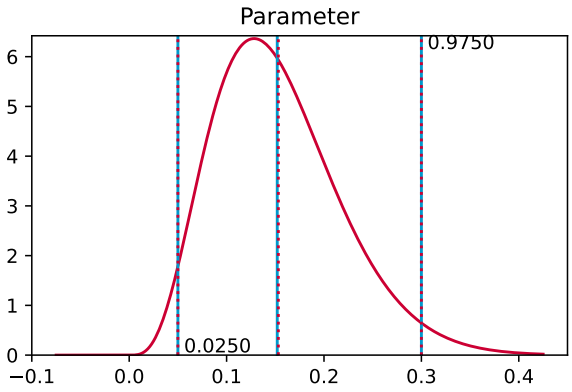


Figure A.1: Example distribution fitting validation plot

BAB distribution fit to $\{q_{2.5} = .05, q_{97.5} = .30\}$; blue solid lines: target quantiles q_i; red dotted lines: distribution quantiles $Q(p_i \mid \theta)$; red solid line: density distribution $f(x \mid \theta)$.

approximation of the binomial distribution (BAB):

$$p(\rho) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\rho^{\alpha-1}(1-\rho)^{\beta-1} \tag{A.3}$$

with $\alpha = N\rho$ and $\beta = N(1-\rho)$. Unlike the approximation by a normal distribution, the beta distribution ensures that $\rho \in [0, 1]$. Figure A.2 illustrates the approximation for $N = \{10, 20, 40\}$ and $\rho = \{0.01, 0.1, 0.5\}$.

A.1.3 Joint Sampling with Relational Constraints

Figure A.3 illustrates the posterior (sampled) distributions for variables X_1, X_2, X_3, having uniform priors but subject to $X_1 < X_2 < X_3$. Three approachs to enforcing $X_1 < X_2 < X_3$ were explored:

- **joint:** sample X_1, X_2, X_3 simultaneously; then discard any samples failing $X_1 < X_2 < X_3$.
- **forward:** sample X_1; then sample X_2 until $X_1 < X_2$; then sample X_3 until $X_2 < X_3$.
- **backward:** sample X_3; then sample X_2 until $X_2 < X_3$; then sample X_1 until $X_1 < X_2$.

All three methods result in a different posterior vs. the prior, but the forward and backward methods severely distort the distributions for X_3 and X_1, respectively, while leaving the distributions for X_1 and X_3 unchanged. By contrast, the joint method influences the posterior distributions of each variable in a more "equitable" way, which is preferred.

A.2 Durations

A.2.1 Exponential Duration Assumption in Compartmental Modedls

Let λ be the fixed exit rate from compartment A, which is assumed to be homogeneous. Then $\delta \sim \lambda e^{-\lambda\delta}$ is the exponentially distributed duration time in the group.

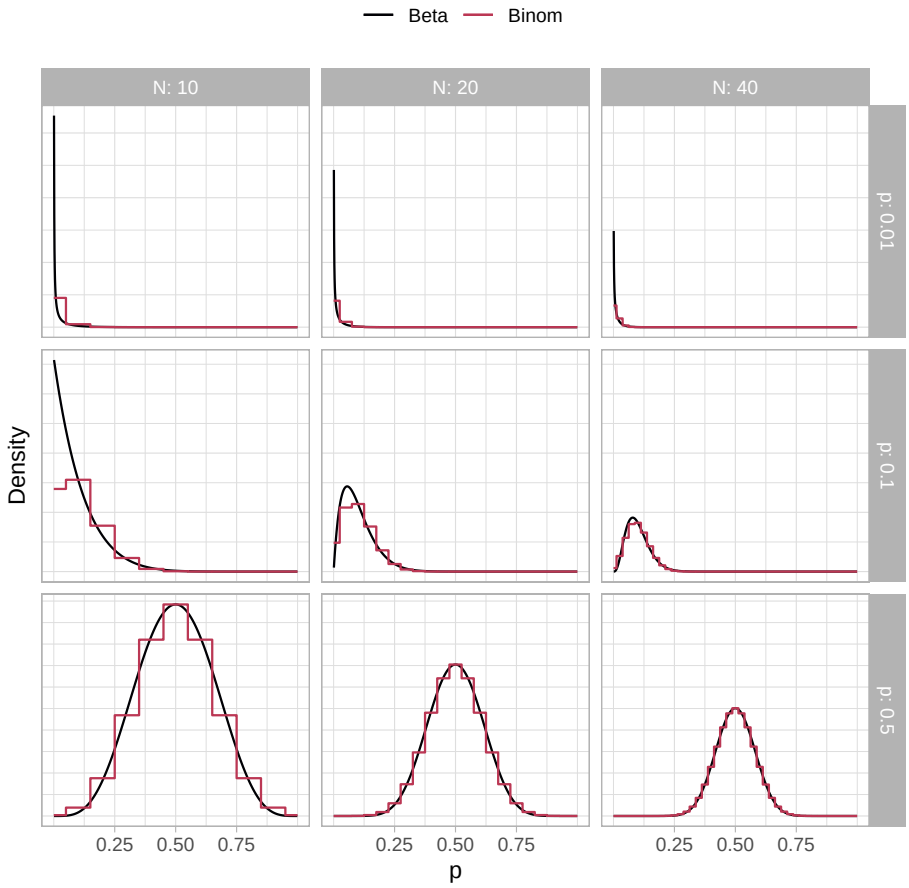


Figure A.2: Beta approximation of the binomial distribution (BAB)

Mean & Median Duration. The mean duration is $\mu = 1/\lambda$ and the median is $m = \log(2)/\lambda \approx 0.69\,\mu$. Thus, if 50% of individuals progress from compartment A to B by time τ (median duration), the exit rate λ is given by $\log(2)/\tau$.

Collapsing Compartments in Series. Let compartments A and B be in series, with exit rates λ_A and λ_B respectively. Collapsing A and B into AB will sum the mean durations: $\delta_{AB} = 1/\lambda_A + 1/\lambda_B$; thus, the exit rate from AB will be $\lambda_{AB} = 1/(1/\lambda_A + 1/\lambda_B)$.

Collapsing Compartments in Parallel. Let compartments A and B be in parallel, with exit rates λ_A and λ_B respectively. Collapsing A and B into AB will sum the exit rates: $\lambda_{AB} = \lambda_A + \lambda_B$; thus, the mean duration in AB will be $\delta_{AB} = 1/(\lambda_A + \lambda_B)$.

A.2.2 Estimating Duration in Sex Work from Cross Sectional Data

Cross sectional sex work surveys will often ask respondents about their duration in sex work. These durations might then be taken to be the average durations in sex work; however, this will be an underestimate,

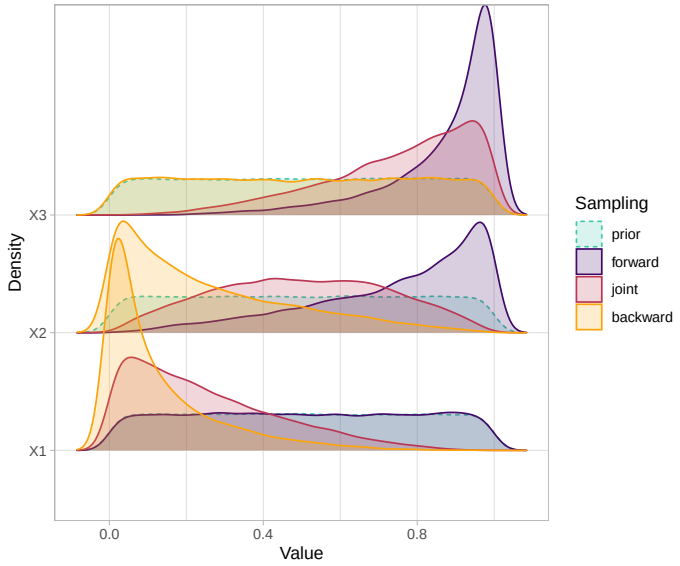


Figure A.3: Illustration of different sampling biases when enforcing $X_1 < X_2 < X_3$

because respondents will continue selling sex after the survey [1].[2]

Figure A.4 illustrates a steady-state population with 7 women selling sex at any given time. The steady-state assumption implies that a women leaving sex work after δ years will be immediately replaced by a women entering sex work whose eventual duration will also be δ years. Let δ be this true duration, and δ_s be the duration reported in the survey. If we assume that the survey reaches women at a random time point during the duration δ, then $\delta_s \sim \text{Unif}(0, \delta)$, and the mean reported duration is $E(\delta_s) = \frac{1}{2}E(\delta)$. Thus, $E(\delta) = 2E(\delta_s)$ would be an estimate of the true mean duration from the sample. In reality, sex work surveys may be more likely to reach women who have already been selling sex for several months or years, due to delayed self-identification as sex worker [2]. Thus, we would expect that $f = E(\delta)/E(\delta_s) \in (1, 2)$, which we can use to compute the mean exit rate as described in § A.2.1.

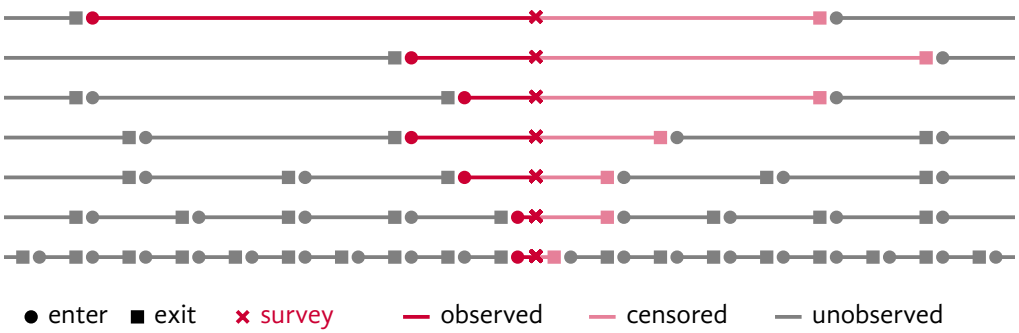


Figure A.4: Illustrative steady-state population of 7 FSW, with varying true durations in sex work δ, vs. the observed durations in sex work δ_s via cross-sectional survey.

[2] An alternate example would be to take the mean age of a population as the life expectancy! Thanks to Saulius Simcikas and Dr. Jarle Tufto for help identifying and discussing this bias: stats.stackexchange.com/questions/298828.

Another observation we can make from Figure A.4 is that women who sell sex longer are more likely to be captured in the survey. That is, while the sampled durations are representative of women who *currently* sell sex, these durations are biased high vs. the population of women who *ever* sell sex. It's not clear whether this observation is widely understood and kept in mind when interpreting sex work survey data.

A.2.3 Quantifying Partnerships

Similar to § A.2.2, sexual partnerships are often quantified using cross-sectional surveys. In this case, respondents are typically asked to report the numbers of unique partners during a standardized recall period γ — e.g., "*How many different people have you had sex with during the past year?*" Such data can then be used to inform modelled rates of partnership change Q and/or numbers of concurrrent partnerships K.

If partnership duration is long and the recall period is short — including $\gamma \approx 0$ for "*Are you currently in a long-term sexual partnership?*" — the reported partnerships mostly reflect *ongoing* partnerships, and thus $C \approx K$. If partnership duration is short and the recall period is long, — including $\delta \approx 0$ for "*How many one-off sexual partners have you had during the past year?*" — the reported partnerships mostly reflect *complete* partnerships, and thus $C/\gamma \approx Q$. However, if partnership duration and recall period are similar in length, the reported partnerships reflect a mixture of tail-ends, complete, and ongoing partnerships, and thus C overestimates K, but C/γ also overestimates Q. In summary:

- $\gamma \ll \delta$: mostly ongoing partnerships; $C \approx K$ (concurrrent)
- $\gamma \gg \delta$: mostly complete partnerships; $C/\gamma \approx Q$ (change rate)
- $\gamma \approx \delta$: some tail-ends, some complete, some ongoing; $C > K$, $C/\gamma > Q$ (neither)

I developed an approach to estimate Q and K from C and γ. The approach draws on a similar assumption as in § A.2.2: that survey timing is effectively random with respect to partnership duration. Then, if either end of the recall period would capture an ongoing partnership, the intersection point would be, on average, at the partnership mid-point. Thus, the recall period is effectively extended by half the partnership duration $\delta/2$ on each end, and δ overall, as illustrated in Figure A.5. As such, we can define Q and K as:

$$Q = \frac{C}{\gamma + \delta} \tag{A.4}$$

$$K = \frac{C\delta}{\gamma + \delta} = Q\delta \tag{A.5}$$

As an example, Figure A.5 illustrates a recall period of $\gamma = 6$ years, for which $C = 9$ partnerships are reported, having durations of $\delta = 3$ years. Thus, we can compute $Q = 9/(6 + 3) = 1$ and $K = 1(3) = 3$, which can be verified by careful examination of Figure A.5.

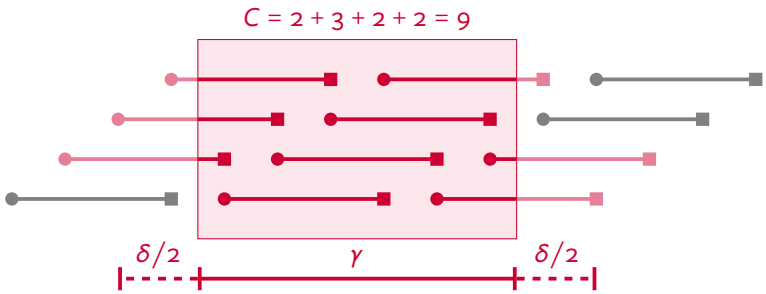


Figure A.5: Illustration of conceptual framework for quantifying partnerships from the number reported during a given recall period

Circle: partnership start; line: ongoing partnership; square: partnership end; red: reported partnership; grey: partnership not reported; γ/red: recall period; δ: partnership duration; C: number of reported partnerships for γ.

A.3 Miscellaneous

A.3.1 Proof that $B_{\text{WPH}} \geq B_{\text{BPH}}$

In § 4.1.1, I claimed that the per-partnership probability of transmission B is larger for within- vs. between-partnership heterogeneity — $B_{\text{WPH}} \geq B_{\text{BPH}}$, from Eqs. (4.5) and (4.6), respectively — given the same set of transmission modifiers R_f, α_f. Here is a proof of that claim:

$$\begin{gathered} B_{\text{WPH}} \geq B_{\text{BPH}} \\ 1 - \prod_f (1-\beta_f)^{A\alpha_f} \geq 1 - \sum_f \alpha_f (1-\beta_f)^A \end{gathered} \tag{A.6}$$

Let $x_f = (1-\beta_f)^A$; then

$$\prod_f {x_f}^{\alpha_f} \leq \sum_f \alpha_f x_f \tag{A.7}$$

Since $\sum_f \alpha_f = 1$ and $\alpha_f \in [0, 1]$ are effectively weights, Eq. (A.7) is the weighted arithmetic mean–geometric mean (AM-GM) inequality [3]. In fact, Aldaz [3] further shows that the the gap between B_{WPH} and B_{BPH} increases with the α_f-weighted variance in ${\beta_f}^{\frac{1}{2}}$ (although the increase is not exact), which supports the results of § 4.3.1 mathematically.

Using Jensen's inequality [4] we can also show that the approach in [5] (and others) to aggregate heterogeneity by HIV infection stage first produces an intermediate per-partnership probability B':[3]

$$B_{\text{WPH}} \geq B' \geq B_{\text{BPH}}, \qquad B' = 1 - \left(1 - \sum_f \alpha_f \beta_f\right)^A \tag{A.8}$$

References

[1] Erika Fazito et al. "Analysis of duration of risk behaviour for key populations: a literature review." *Sexually transmitted infections* 88.S2 (Dec. 2012), pp. i24–i32. https://doi.org/10.1136/sextrans-2012-050647.

[3] math.stackexchange.com/questions/4660409

[2] Eve Cheuk et al. "Transitions: Novel Study Methods to Understand Early HIV Risk Among Adolescent Girls and Young Women in Mombasa, Kenya, and Dnipro, Ukraine". *Frontiers in Reproductive Health* 2 (Sept. 2020), p. 10. https://doi.org/10.3389/frph.2020.00007.

[3] J. M. Aldaz. "Self-improvement of the inequality between arithmetic and geometric means". *Journal of Mathematical Inequalities* 2 (2009), pp. 213–216. https://doi.org/10.7153/jmi-03-21.

[4] J. L.W.V. Jensen. "Sur les fonctions convexes et les inégalités entre les valeurs moyennes". *Acta Mathematica* 30.1 (Dec. 1906), pp. 175–193. https://doi.org/10.1007/BF02418571.

[5] Cliff C. Kerr et al. "Optima: A Model for HIV Epidemic Analysis, Program Prioritization, and Resource Optimization". *Journal of Acquired Immune Deficiency Syndromes* 69.3 (July 2015), pp. 365–376. https://doi.org/10.1097/QAI.0000000000000605.