Title Quantifying the impact of cascade inequalities:
a modelling study on the prevention impacts of antiretroviral therapy scale-up in Eswatini

Authors Jesse Knight[1,2,*], Huiting Ma[1], Bheki Sithole[3], Lungile Khumalo[4], Linwei Wang[1], Sheree Schwartz[5], Laura Muzart[3], Sindy Matse[6], Zandile Mnisi[6,+], Rupert Kaul[7], Michael Escobar[8], Stefan Baral[5], and Sharmistha Mishra[1,2,7,8,9]

[1] MAP Centre for Urban Health Solutions, Unity Health Toronto
[2] Institute of Medical Science, University of Toronto
[3] EpiC, FHI 360, Eswatini
[4] Voice of Our Voices, Eswatini
[5] Bloomberg School of Public Health, Johns Hopkins University
[6] Ministry of Health, Eswatini
[7] Division of Infectious Diseases, Department of Medicine, University of Toronto
[8] Dalla Lana School of Public Health, University of Toronto
[9] Institute for Clinical Evaluative Sciences, Toronto, Ontario
[+] *in memory of*
[*] jesse.x.knight@protonmail.com

Date 2024 February 16

Journal medRxiv

Status preprint

Funding The study was supported by the Natural Sciences and Engineering Research Council of Canada (CGS-D); the Ontario Ministry of Colleges and Universities (QEII-GSST); the Canadian Institutes of Health Research (FN-13455); the National Institute of Allergy and Infectious Diseases (R01AI170249).

Acknowledgements We thank: Kristy Yiu, Samantha Lo (Unity Health Toronto) for research coordination support; Amrita Rao, Carly Comins (Johns Hopkins University), Alex Whitlock, Korryn Bodner (Unity Health Toronto), and Leigh Johnson (University of Cape Town) for helpful discussions and feedback on model design.

Data & Code We used only published aggregate data, except for individual-level data from two female sex worker surveys, which were accessed under approval from the Scientific and Ethics Committee of Swaziland (Eswatini) Ministry of Health (MH/599B), and the Institutional Review Board of the Johns Hopkins Bloomberg School of Public Health (3508). All code and selected results are available at: github.com/mishra-lab/hiv-model-eswatini

Abstract **Background** Inequalities in the antiretroviral therapy (ART) cascade across subpopulations remain an ongoing challenge in the global HIV response. Eswatini achieved the UNAIDS 95-95-95 targets by 2020, with differentiated programs to minimize inequalities across subpopulations, including for female sex workers (FSW) and their clients. We sought to estimate additional HIV infections expected in Eswatini if cascade scale-up had not been equal, and under which epidemic conditions these inequalities could have the largest influence.

Methods Drawing on population-level and FSW-specific surveys in Eswatini, we developed a compartmental model of heterosexual HIV transmission which included eight subpopulations and four sexual partnership types. We calibrated the model to stratified HIV prevalence, incidence, and ART cascade data. Taking observed cascade scale-up in Eswatini as the base-case — reaching 95-95-95 in the overall population by 2020 — we defined four counterfactual scenarios in which the population overall reached 80-80-90 by 2020, but where FSW, clients, both, or neither were disproportionately left behind, reaching only 60-40-80. We quantified relative additional cumulative HIV infections by 2030 in counterfactual *vs* base-case scenarios. We further estimated linear effects of viral suppression gap among FSW and clients on additional infections by 2030, plus effect modification by FSW/client population sizes, rates of turnover, and HIV prevalence ratios.

Findings Compared with the base-case scenario, leaving behind neither FSW and nor their clients led to the fewest additional infections by 2030: median (95% credible interval) 13.3 (9.2, 18.7) % *vs* 26.7 (19.7, 33.8) % if both were left behind — a 104 (50, 167) % increase. The effect of lower cascade on additional infections was larger for clients *vs* FSW, and client population size, turnover, and prevalence ratios were the largest modifiers of both effects.

Interpretation Inequalities in the ART cascade across subpopulations can undermine the anticipated prevention impacts of cascade scale-up. As Eswatini has shown, addressing inequalities in the ART cascade, particularly those that intersect with high transmission risk, could maximize incidence reductions from cascade scale-up.

Funding Natural Sciences and Engineering Research Council of Canada; Ontario Ministry of Colleges and Universities; Canadian Institutes of Health Research; National Institute of Allergy and Infectious Diseases.

Keywords HIV, mathematical model, antiretroviral therapy, sex work, healthcare disparities

1 Introduction

Early HIV treatment via antiretroviral therapy (ART) is a lifesaving intervention increasing both the quantity and quality of life [1]. A secondary benefit of early ART given Undetectable = Untransmittable (U=U) is that transmission risks are mitigated in serodifferent partnerships [2]. To realize these benefits, massive efforts are underway to achieve the UNAIDS 95-95-95 ART cascade targets [3] — *i.e.*, to have: 95% diagnosed among people living with HIV, 95% on ART among those diagnosed, and 95% virally suppressed among those on ART. Botswana, Eswatini, Rwanda, Tanzania, and Zimbabwe have already surpassed 95-95-95 nationally [3], and and achieving these targets is expected to help reduce HIV incidence towards elimination.

Yet, there are growing concerns that inequalities in the ART cascade could undermine the population-level prevention impacts of ART anticipated from individual-level and model-based studies [4–6]. Specifically, available data suggest that cascade attainment is often lower among subpopulations at greater risk of HIV acquisition and/or transmission, including key populations, younger men and women, and highly mobile populations [5,7]. These inequalities can be driven by systemic barriers to engagement in care faced by marginalized populations, which intersect with individual, network, and structural determinants of HIV risk, such as economic insecurity, mobility, stigma, discrimination, and criminalization [8–11]. Moreover, cascade data may be lacking entirely for subpopulations experiencing the greatest barriers to care — *i.e.*, the lowest ART cascades likely remain unmeasured [7].

Numerous transmission modelling studies have sought to estimate the prevention impacts of achieving 90-90-90+ across Sub-Saharan Africa [12]. Modelled populations are often stratified by risk, including key populations like female sex workers (FSW) and their clients, to capture important epidemic dynamics related to risk heterogeneity [13]. However, these studies have generally assumed that ART cascade attainment (*i.e.*, proportions diagnosed, treated, and virally suppressed) or progression (*i.e.*, rates of diagnosis, treatment initiation, and viral suppression) were equal across modelled subpopulations. For example, among the studies in [12], key populations were usually assumed to have equal cascade progression with the population overall, or greater in some scenarios, but never lesser. Thus, the potential influence of intersecting risk and cascade inequalities on ART prevention impacts has not been explored.

We therefore examined the following questions in an illustrative modelling analysis:

1. How are estimates of population-level ART prevention impacts influenced by inequalities in ART cascade across subpopulations?
2. Under which epidemic conditions do such inequalities have the largest influence?

We examined these questions using a deterministic compartmental model of heterosexual HIV transmission in Eswatini, focusing on inequalities related to sex work. Eswatini has the highest national HIV prevalence in the world, and recently surpassed 95-95-95 with minimal inequalities across subpopulations [3,14]. As such, we used observed ART cascade scale-up in Eswatini as a *base case* reflecting exemplary and evidently attainable scale-up, and explored *counterfactual* scenarios in which scale-up was slower, and where specific subpopulations could have been left behind.

Evidence before this study We adapted and updated a scoping review of HIV transmission modelling studies examining the prevention impacts of antiretroviral therapy (ART) cascade scale-up in Sub-Saharan Africa (full details in Appendix B.3). Our search yielded (prior + update) 1367 + 373 unique studies, of which: 7 + 5 examined the prevention impacts of achieving the UNAIDS 90-90-90 goals or greater (Table B.3). Some of these studies considered differences in baseline rates of diagnosis, treatment initiation, and/or treatment failure/discontinuation — mainly by sex, age, and occasionally risk. Three studies specifically examined scenarios of unequal cascade attainment across subpopulations, predicting substantially more infections when subpopulations at higher risk were left behind. However, these studies did not maintain consistent overall cascade across scenarios — *i.e.*, each scenario reached a different cascade for the population overall — and considered only hypothetical scenarios.
Added value of this study We developed and calibrated a detailed model of heterosexual HIV transmission and observed ART cascade scale-up in Eswatini, drawing on population-level and female sex worker (FSW)-specific surveys. Eswatini achieved 95-95-95 overall and similar among FSW by 2020. We estimated what would have happened if overall cascade scale-up had been slower (reaching only 80-80-90 by 2020), and inequalities in the cascade among FSW and/or clients had not been addressed (reaching 60-40-80 by 2020). We found that such cascade inequalities would have led to 50–167% additional HIV infections in the overall population by 2030, as compared to equitable scale-up.
Implications of all the available evidence The population-level prevention benefits of achieving 95-95-95 overall are undermined by inequalities in the HIV cascade, especially when populations experiencing disproportionate transmission risks are left behind, including but not limited to FSW and their clients.

Box 1: Research in context

2 Methods

We constructed a deterministic compartmental model of heterosexual HIV transmission, stratified by subpopulations defined by sex and sexual activity, and health states reflecting HIV natural history and ART cascade of care. The model includes eight subpopulations, including FSW at higher *vs* lower risk, and likewise for clients of FSW, and four partnership types, including regular and occasional sex work (Figure A.1). We calibrated the model to reflect the HIV epidemic and ART scale-up in Eswatini (*base case*). We then explored *counterfactual* scenarios in which ART cascade was reduced among various combinations of subpopulations, and quantified ART prevention impacts by comparing *base case* and *counterfactual* scenarios.

2.1 Model Parameterization & Calibration

Complete details of the model structure, parameterization, and calibration are given in Appendix A.

HIV HIV natural history included acute infection and stages defined by CD4-count. We modelled relative rates of infectiousness by stage as an approximation of viral load [15–17], as well as rates of HIV-attributable mortality by stage [18].

Risk heterogeneity We captured risk heterogeneity through subpopulation-level factors, including subpopulation size, average duration in subpopulation, genital ulcer disease (GUD) prevalence, and rates / types of partnership formation; and partnership-level factors, including assortative mixing, partnership duration, frequency of sex, and levels of condom use. Table 1 summarizes key parameter values and sampling distribu-

tions related to risk heterogeneity. To parameterize FSW at higher *vs* lower risk, we analyzed individual-level survey data from Swati FSW in 2011 [19] and 2014 [20] (Appendix A.3.9). We parameterized the remaining subpopulations using reported data from national studies in 2006 [21], 2011 [22], and 2016 [23]. We modelled increasing condom use (Figure B.10), increasing voluntary medical male circumcision (Figure B.9), and decreasing GUD prevalence over time, We did not model other interventions (*e.g.*, current pre-exposure prophylaxis scale-up) nor non-heterosexual HIV transmission.

ART cascade We modelled rates of HIV diagnosis among people living with HIV as monotonically increasing over time. We defined a base rate for women with low/medium sexual activity, and constant relative rates for men with low/medium sexual activity (< 1), clients (< 1), and FSW (> 1), reflecting increased HIV testing access via antenatal care among women *vs* men, and enhanced screening among FSW [19]. We modelled ART initiation similarly except: the relative rate for ART initiation among FSW was < 1, reflecting specific barriers to uptake and engagement in care [8]; and we defined additional relative rates by CD4 count ($\leq$ 1) to reflect historical ART eligibility criteria [23]. We modelled viral suppression using a fixed rate for all subpopulations, corresponding to an average of 4 months from ART initiation [24]. We modelled treatment failure / discontinuation with a single monotonically decreasing rate applied to all subpopulations in the base case, reflecting improving treatment success / retention over time [23]. Individuals with treatment failure / discontinuation could re-initiate ART at a fixed rate, reflecting re-engagement in care or detection of treatment failure and initiation of alternative regimens. We modelled rapid CD4 recovery during the first 4 months of ART, followed by slower recovery while virally suppressed [25]. We modelled reduced HIV-attributable mortality among individuals on ART, in addition to mortality benefits of CD4 recovery.

Calibration We calibrated the model to reflect available data from Eswatini on HIV prevalence, HIV incidence, and ART cascade of care overall and stratified by subpopulation where possible (Tables A.6–A.10) [14,19–23,26] using an adapted version of Incremental Mixture Importance Sampling (IMIS) [27]. Full methodology is given in Appendix A.4, while calibration results are given in Appendix B.1.

2.2 Scenarios & Analysis

2.2.1 Objective 1: Influence of ART cascade differences between subpopulations

For Objective 1, We defined the *base case* scenario to reflect observed ART cascade scale-up in Eswatini, reaching 95-95-95 for the population overall by 2020 [14], and 88-98-xx among FSW specifically [26].[1] Next, we defined four *counterfactual* scenarios in which the overall population cascade reached 80-80-90 by 2020, and where FSW, clients, both, or neither were disproportionately left behind. In these counterfactual scenarios, we altered cascade attainment among FSW, clients, and/or the remaining population ("all others") by calibrating constant relative scaling factors "*R*" to subpopulation-specific rates of: diagnosis ($0 \leq R_d \leq 1$), treatment initiation ($0 \leq R_t \leq 1$), and treatment failure / discontinuation ($1 \leq R_u \leq 20$). When FSW and/or clients were left behind, we calibrated their *R*s such that these subpopulations attained approximately 60-40-80 by 2020. By contrast, we calibrated *R*s for the remaining population such that the Swati population *overall* attained

[1] Data on viral suppression for FSW were not available, which we denote as "xx"; no cascade data were available for clients specifically.

Table 1: Selected model parameters related to risk heterogeneity

Parameter	Stratification	Posterior Mean	Posterior (95% CI)	Ref.
Subpopulation size (% of total)	FSW, of women overall	4.4	(2.8, 6.7)	§ A.3.11.1
	Clients, of men overall	10.0	(4.7, 15.7)	§ A.3.11.2
	HR of FSW / clients overall	20	—	§ A.3.9
Duration in subpop (mean years)	FSW overall	3.5	(2.7, 4.9)	§ A.3.12.2
	Clients overall	8.9	(4.9,15.4)	§ A.3.12.2
	HR sex work and clients	0.7	(0.3, 1.2)	§ A.3.12.2
Sex work clients per year	HR FSW, occasional sex work	84	(53, 126)	§ A.3.13.1
	LR FSW, occasional sex work	20	(13, 27)	§ A.3.13.1
	HR FSW, regular sex work	50	(23, 86)	§ A.3.13.1
	LR FSW, regular sex work	11	(6, 21)	§ A.3.13.1
Sex work visits per year	HR clients	90	(44, 154)	§ A.3.13.1
	LR clients	44	(22, 74)	§ A.3.13.1
Total partner change rate per year	Medium activity women	0.99	(0.72, 1.31)	§ A.3.13.2
	Medium activity men	0.67	(0.28, 1.27)	§ A.3.13.2
	Lowest activity women & men	0.23	(0.15, 0.35)	§ A.3.13.2
Any GUD p12m prevalence (%) [a]	HR FSW	38	(25, 53)	§ A.3.5.3
	LR FSW & HR clients	30	(20, 41)	§ A.3.5.3
	LR clients	24	(15, 33)	§ A.3.5.3
	Medium activity	18	(8, 29)	§ A.3.5.3
	Lowest activity	7	—	§ A.3.5.3
Relative infectiousness	Acute infection	10.6	(5.3, 18.1)	§ A.3.4.1
	Any GUD p12m	2.1	(1.3, 3.2)	§ A.3.4.5
Relative susceptibility	Receptive vaginal sex	1.7	(1.2, 2.0)	§ A.3.4.2
	Receptive anal sex	10.0	—	§ A.3.4.2
	Any GUD p12m	3.0	(1.2, 6.3)	§ A.3.4.5
Sex acts per partnership-year	Main/spousal	77	(40, 125)	§ A.3.14
	Casual	120	(62, 202)	§ A.3.14
	Occasional sex work	12	—	§ A.3.14
	Regular sex work	18	(12, 30)	§ A.3.14
Anal sex acts (% of all acts)	Main/spousal & casual	11	(3, 21)	§ A.3.14
	Occasional & regular sex work	17	(4, 34)	§ A.3.14
Condom use in 2020 (% of acts protected)	Main/spousal	41	(31, 51)	§ A.3.5.2
	Casual	70	(62, 77)	§ A.3.5.2
	Occasional sex work	88	(78, 95)	§ A.3.5.2
	Regular sex work	78	(64, 87)	§ A.3.5.2
	Anal vs vaginal sex	74	(63, 88)	§ A.3.5.2
Partnership duration (years)	Main/spousal	16.4	(14.6, 18.5)	§ A.3.15
	Casual	0.78	(0.45, 1.24)	§ A.3.15
	Occasional sex work	0.83	—	§ A.3.15
	Regular sex work	0.43	(0.19, 0.83)	§ A.3.15

Low: lowest activity; Med: medium activity; LR/HR: lower/higher risk; FSW: female sex workers; Clients: of FSW; p12m: past 12 months;
[a] GUD prevalence declines universally after 2010 as described in § A.3.5.3.

Table 2: Modelling scenarios for Objective 1 defined by 2020 calibration targets

	ART cascade in 2020 [a]			Re-scaled cascade rates [b]		
Scenario	FSW	Clients	Overall	FSW	Clients	All Others
Base Case	88-98-xx	xx-xx-xx	95-95-95	—	—	—
Leave Behind: FSW	60-40-80	—	80-80-90	✓	✗	✓
Leave Behind: Clients	—	60-40-80	80-80-90	✗	✓	✓
Leave Behind: FSW & Clients	60-40-80	60-40-80	80-80-90	✓	✓	✓
Leave Behind: Neither	—	—	80-80-90	✗	✗	✓

[a] Cascade: % diagnosed among PLHIV; % on ART among diagnosed; % virally suppressed among on ART; xx: no data available; [b] Rates of: diagnosis; ART initiation; treatment failure; FSW: female sex workers; Clients: of FSW; All Others: all women and men not involved in sex work. Figure B.16 plots the modelled cascades over time.

80-80-90 in all 4 counterfactual scenarios, thus ensuring that a consistent proportion of the population overall attained viral suppression. Table 2 summarizes these scenarios, while Figure B.16 plots the modelled cascades over time. When cascade rates among FSW and/or clients were unchanged from the base case, the cascade these subpopulations attained could be lower than in the base case due to subpopulation turnover and higher incidence. All cascades continued to increase beyond 2020 due to assumed fixed rates of diagnosis, treatment initiation, and treatment failure / discontinuation thereafter.

We quantified ART prevention impacts via relative cumulative additional infections (CAI) and additional incidence rate (AIR) in the counterfactual scenarios (k) *vs* the base case (0), over multiple time horizons up to 2030, starting from $t_0 = 2000$:

$$\text{CAI, AIR}\,(t) = \frac{\Omega_k(t) - \Omega_0(t)}{\Omega_0(t)}, \qquad \Omega(t) = \begin{cases} \int_{t_0}^{t} \Lambda(\tau)\, d\tau & \text{CAI} \\ \lambda(t) & \text{AIR} \end{cases} \tag{1}$$

where: Λ denotes absolute numbers of infections per year, and λ denotes incidence rate per susceptible per year. For each scenario, we computed these outcomes (CAI and AIR) for each model fit j, and reported median (95% credible interval, CI) values across model fits, reflecting uncertainty.

2.2.2 Objective 2: Conditions that maximize the influence of ART cascade differences

For Objective 2, we estimated via linear regression: the effects of lower ART cascade among FSW and clients on relative CAI and AIR, plus potential effect modification by epidemic conditions. The hypothesized causal effects are illustrated as a directed acyclic graph in Figure 1.

For this regression, we generated 10,000 synthetic samples as follows. We explored a wider range of counterfactual scenarios *vs* Objective 1 by randomly sampling the relative rates for diagnosis and treatment initiation $R_d, R_t \sim$ Beta ($\alpha = 3.45, \beta = 1.85$) and treatment failure $R_u \sim$ Gamma ($\alpha = 3.45, \beta = 1.88$) for each of: FSW, clients, and the remaining population (9 total values). These sampling distributions had 95% CI: (0.25, 0.95) and (1.5, 15), respectively, and were chosen to obtain cascades in 2020 spanning approximately 60-60-90

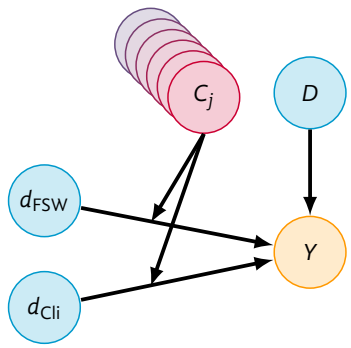

Figure 1: Directed acyclic graph (DAG) for inferring the epidemic conditions under which differential viral suppression across subpopulations matters most

Y: cumulative additional infections (CAI) or additional incidence rate (AIR) by 2030; *D*: difference in population-overall viral non-suppression in counterfactual *vs* base case scenario; d_i: difference in subpopulation-*i*-specific viral non-suppression *vs* population overall within counterfactual scenario; C_j: epidemic conditions (effect modifiers of d_i).

through 90-90-95 (Figure B.19). For each of $N_f = 1000$ model fits, we generated $N_k = 10$ counterfactual scenarios per fit using Latin hypercube sampling of *R*s, yielding $N_f N_k = 10{,}000$ total counterfactual samples for the regression.

For each of these 10,000 samples, we defined relative CAI and AIR by 2030 *vs* the base case, as in Objective 1. For each sample, we further defined U_{fki} for subpopulations $i \in \{1 : \text{FSW}, 2 : \text{clients}, * : \text{overall}\}$ as the proportions *not* virally suppressed among those living with HIV by 2020, reflecting a summary measure of ART cascade gaps. Using U_{fki}, we defined the main regression predictors as: $D_{fk} = U_{fk*} - U_{f0*} > 0$, reflecting differences in *population overall* viral non-suppression in sample $k \in [1, 10]$ *vs* the base case (denoted $k = 0$); and $d_{fki} = U_{fki} - U_{fk*} \lesseqgtr 0$, reflecting differences in *subpopulation-i-specific* viral non-suppression in sample *k vs* the population overall in sample *k* — *i.e.*, viral non-suppression inequalities.

Next, we defined the following measures of epidemic conditions (C_{fj}) related to sex work, as hypothesized modifiers of the effect of unequal non-suppression on relative CAI and RAI: FSW and client population sizes (% of population overall); average rate of turnover among FSW and clients (per year, reciprocal of duration selling / buying sex); and HIV prevalence ratios in the year 2005 among FSW *vs* other women, and among clients *vs* other men. To calculate the HIV prevalence ratios, we combined FSW at higher and lower risk, and likewise for clients at higher and lower risk. We used HIV prevalence ratios in 2005 to reflect summary measures of risk heterogeneity prior to ART, rather than including all risk factors from the transmission model, which could lead to overfitting and improper inference due to effect mediation.

Finally, we defined a general linear model for each outcome (CAI, AIR) as:

$$\text{CAI, AIR} = \beta_0 D + \sum_i \beta_i d_i + \sum_{ij} \beta_{ij} d_i C_j \tag{2}$$

such that each outcome was modelled as a sum of the effects of: differential population-level non-suppression

in the counterfactual *vs* the base scenario (D); differential non-suppression among FSW and clients *vs* the population overall within the counterfactual scenario (d_i); and effect modification of d_i by epidemic conditions (C_j). The model does not include an intercept because if $D = d_i = 0$, then we expect CAI = AIR = 0. We fitted this model for each outcome using generalized estimating equations to control for repeated use of each model fit f. We standardized all model variables (D, d_i, C_j) via $\hat{x} = (x - \text{mean}(x))/\text{SD}(x)$ to avoid issues of different variable scales and collinearity in interaction terms. This standardization does not imply that regression coefficient magnitudes can be compared to indicate variable "importance", because the standardization applied to each variable is driven by the variance before standardization — in this case reflecting arbitrary ranges (D, d_i) or uncertainty in calibration (C_j).[2] Rather, effect sizes can be interpreted as: the expected change in outcome per standard deviation change in the variable.

3 Results

We first summarize modelled patterns of HIV transmission in the base case (1000 model fits), calibrated to reflect the Eswatini epidemic up to 2020. Our model suggests that transmission within regular sex work partnerships was a dominant driver of early epidemic growth (Figures B.14 and B.15). However, from approximately 1997 onward, the majority of new yearly infections were transmitted within casual partnerships, including 44% (median) of infections in 2020 in the base case.[3] By 2020, the majority of cumulative infections acquired were among women with lowest sexual activity (Figure B.12b), while the majority of cumulative infections transmitted were from clients (Figure B.12a). Overall HIV prevalence in 2020 was median (95% CI): 24.4 (23.3, 25.4) % (Figure B.4a), and overall incidence was 8.0 (6.7, 9.4) per 1000 person-years (Figure B.4c). The prevalence ratio between FSW and women overall was 1.75 (1.69, 1.86), and between clients and men overall it was 2.14 (1.55, 2.99) (Figure B.4b). Due to turnover and higher HIV incidence among FSW, achieving similar rates of diagnosis among FSW *vs* other women (Figure B.6a) required approximately twice the rate of testing. Sex work contributed a growing proportion of infections over 2020–2040: from 40% to 52% (Figure B.14).

3.1 Objective 1: Influence of cascade differences between subpopulations

Figure B.16 illustrates ART cascade attainment over time in each of the four counterfactual scenarios (80-80-90 overall by 2020), and the base case (95-95-95 overall by 2020). Figure B.17 illustrates overall HIV incidence in each scenario. Figure 2 then illustrates cumulative additional infections (CAI) and additional incidence rate (AIR) in each counterfactual scenario *vs* the base case. Leaving behind neither FSW and nor clients led to the fewest CAI: median (95% CI) 13.3 (9.2, 18.7) % more than the base case by 2030. By contrast, leaving behind both FSW and clients led to the most CAI: 26.7 (19.7, 33.8) % more than the base case by 2030 — a 104 (50, 167) % increase. Leaving behind either FSW or clients resulted in 22.3 (16.5, 27.9) % or

[2] We verified that results were qualitatively unchanged using $\hat{x} = (x - \text{median}(x))/\text{IQR}(x)$. For further discussion on interpretation of standardized regression coefficients, see also: `stats.stackexchange.com/questions/29781` and links therein.

[3] In our model, casual partnerships can be formed by any subpopulation, and these partnerships subsume transactional partnerships.

19.1 (12.9, 25.0) % cumulative additional infections *vs* the base case, respectively. Results were similar for additional incidence rate (AIR). In all counterfactual scenarios, the majority of additional infections were transmitted via casual partnerships (Figure B.18a), and acquired among non-FSW women (Figure B.18c). Patterns of onward transmission were also similar across scenarios (Figure B.18b), though subpopulation contributions increased if they were left behind.

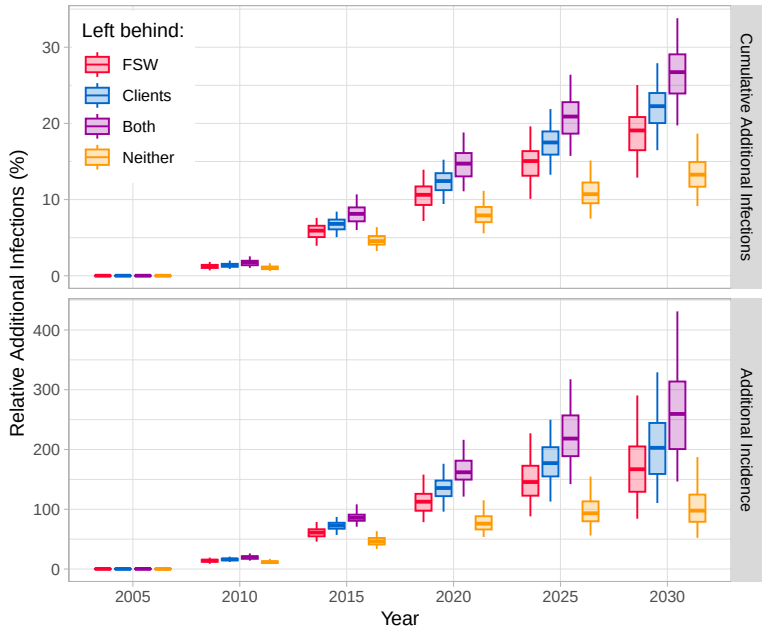

Figure 2: Relative additional infections under counterfactual scenarios *vs* the base case

Base case: 95-95-95 by 2020; counterfactual scenarios: 80-80-90 overall by 2020, with reduced cascade (60-40-80: left behind) among FSW, clients of FSW, both, or neither; whiskers, boxes, and midlines: 95% CI, 50% CI, median of model fits.

3.2 Objective 2: Conditions that maximize the influence of cascade differences

The fitted regression models Eq. (2) indicated that population-overall viral non-suppression (D) and relative non-suppression among FSW and clients (d_i) each had strong and positive effects on 2030 CAI and AIR outcomes ($p < 10^{-5}$), though the effect size was larger for clients. These effects support the results of Objective 1. Figure 3 plots the estimated effects of subpopulation-specific non-suppression d_i, and effect modification by epidemic conditions C_j. The effect of relative non-suppression among both FSW and clients on CAI and AIR increased with: larger client population size, faster client turnover, larger HIV prevalence ratio among clients *vs* other men, and smaller HIV prevalence ratio among FSW *vs* other women. The effect of relative non-suppression among FSW further increased slightly with duration in sex work, while the effect of relative non-suppression among clients further increased slightly with *smaller* FSW population size.

4 Discussion

We sought to explore how inequalities in the ART cascade that intersect with HIV risk heterogeneity may influence the model-estimated prevention impacts of ART. In our applied analysis of Eswatini, we found that slower cascade scale-up that left behind female sex workers (FSW) and their clients could have resulted in 50–167% more infections by 2030 *vs* slower scale-up alone. We also found that the impact of leaving behind FSW and/or clients was largely determined by characteristics of the client population, including: population size, turnover, and relative HIV prevalence.

Eswatini has recently surpassed 95-95-95 for the population overall [14]. These targets were achieved through numerous initiatives coordinated across sectors, including those led by the *MaxART* program [28]. Multiple stakeholders, including people living with HIV, healthcare providers, traditional and religious leaders, community groups, and researchers were engaged via multiple channels, such as Technical Working Groups, Community Advisory Boards, and specific meetings for prioritized groups (men and adolescents) [28]. Drawing on this engagement and social science research to understand barriers to care, cascade services were comprehensively strengthened via investments in training, infrastructure, anti-stigma communication, demand creation, and monitoring [28].

Among FSW living with HIV in Eswatini, data suggest that 88% were diagnosed and 86% were on ART in 2021 (*i.e.*, ART coverage was 98% among those diagnosed) [26]. Data on viral suppression among FSW were lacking, and so assumed to be similar to other women in the simulated epidemic (see § A.3.8 for more details). Although lower than 95-95-95, this cascade among FSW living with HIV is higher than in many other regions [7,10]. Strong programs are required to attain high cascades among FSW, considering that women enter and exit sex work (turnover) and likely experience highest risk of HIV acquisition during sex work. That is, programs must ensure higher rates of HIV testing, ART initiation, and retention among FSW *vs* other women to achieve similar cascades. For example, we inferred that rates of HIV testing in 2016 was 80–227% higher among FSW *vs* other women to reproduce observed cascade data during model calibration.

In Eswatini, programs for key populations include safe access to tailored services via drop-in centers (locally known as TRUE), mobile outreach, venue-based, and one-on-one options [26]. Health and clinical services are also integrated with efforts to reduce structural vulnerabilities, including experiences of harassment, violence, and fear of seeking healthcare, through community empowerment, psycho-social and legal supports, and sensitization and training for police and healthcare workers [26]. These programs have been designed and refined with ongoing community leadership and engagement, allowing them to better meet the specific needs of key populations, for whom barriers to engagement in HIV care often intersect with drivers of HIV risk, including economic insecurity, mobility, stigma, discrimination, and criminalization [4,8–11,29]. Our data-informed modeling of cascade scale-up in Eswatini confirms that such an equity-focused approach to ART cascade scale-up can maximize prevention impacts, and accelerate overall reductions in HIV incidence.

Our study highlights the importance of reaching both FSW and their clients, echoing recent modelling studies of South Africa and Cameroon [30,31]. These studies found that gaps in HIV prevention and treatment for clients were among the largest contributors to onward transmission in recent years. Such findings reiterate

the need for improved data on both FSW and clients, including estimates of population size, sexual behaviour, and ART cascade attainment. These estimates may be difficult to obtain because individuals are unlikely to report buying or selling sex in population-level surveys due to stigma and criminalization [32] and because many clients are highly mobile (including transient seasonal/occupational migration) [29]. Thus, innovative study designs, bias adjustments, and services may be needed to understand and meet clients' needs.

While numerous modelling studies have examined the potential prevention impacts of ART cascade scale-up [12] (Appendix B.3), our study is the first to explore the impact of inequalities in ART cascade across subpopulations with consistent population overall cascade across scenarios. Similar work by Marukutira et al. [33] illustrated the limited impact of achieving 95-95-95 for only citizens and not immigrants in Botswana, while Maheu-Giroux et al. [34] illustrated the high cost-effectiveness of prioritizing key populations (including clients) for ART in Côte d'Ivoire. Indeed, our findings are likely generalizable to other epidemic contexts. HIV prevalence ratios between key populations and the population overall are relatively low in Eswatini *vs* elsewhere; thus, the impact of cascade inequalities among key populations in other contexts would likely be even greater than we found for Eswatini. Moreover, as HIV incidence declines in many settings, transmissions may become concentrated among key populations [35], further magnifying the impact of cascade inequalities.

A primary strength of our analysis is the use of observed ART cascade scale-up to 95-95-95 in Eswatini as the base case, with plausible cascade inequalities explored in counterfactual scenarios. As noted above, the available data suggest that Eswatini has minimized cascade inequalities which persist elsewhere [7]. Thus, our counterfactual scenarios directly estimate the consequences of failing to address these inequalities. Second, drawing on our conceptual framework for risk heterogeneity [12, Table 1] and multiple sources of context-specific data [19–23,26], we captured several dimensions of risk heterogeneity, including: heterosexual anal sex, four types of sexual partnerships, sub-stratification of FSW and clients into higher and lower risk strata, and subpopulation turnover (Table 1, Appendix A). Accurate modelling of risk heterogeneity has been shown to mediate model-estimated ART prevention impacts [36], and is especially important when considering differential ART scale-up across subpopulations. Finally, our analytic approach to Objective 2, in which epidemic conditions are conceptualized as potential effect modifiers represents a unique methodological contribution to the HIV modelling literature.

Our study also has limitations. First, we did not model pre-exposure prophylaxis (PrEP). However, our analyses focus on the time period prior to widespread PrEP availability in Eswatini. Second, we did not consider transmitted drug resistance (TDR). However, drug resistance is more likely to emerge in the context of barriers to viral suppression; thus, lower cascade among those at higher risk would likely accelerate emergence of transmitted drug resistance, and thereby magnify our findings. Finally, our model structure did not include age, and we only considered heterosexual HIV transmission in Eswatini. Future work can explore adaptation of the model to consider PrEP, TDR, age stratification, and additional modes of HIV transmission. While the magnitude of our results may change with such adaptations, we do not expect that the qualitative interpretation would change. In fact, our findings would likely generalize to other transmission networks and determinants of risk heterogeneity, including other key populations and subpopulations such as highly mobile populations and young women [29,37].

In conclusion, the HIV response must remain rooted in context-specific understandings of inequalities in HIV risk and in access to HIV services, which often stem from common upstream factors. Thus, differences in ART cascade within and between subpopulations at higher risk of HIV must be monitored, characterized, and addressed to fully realize the anticipated benefits of ART at both the individual and population levels.

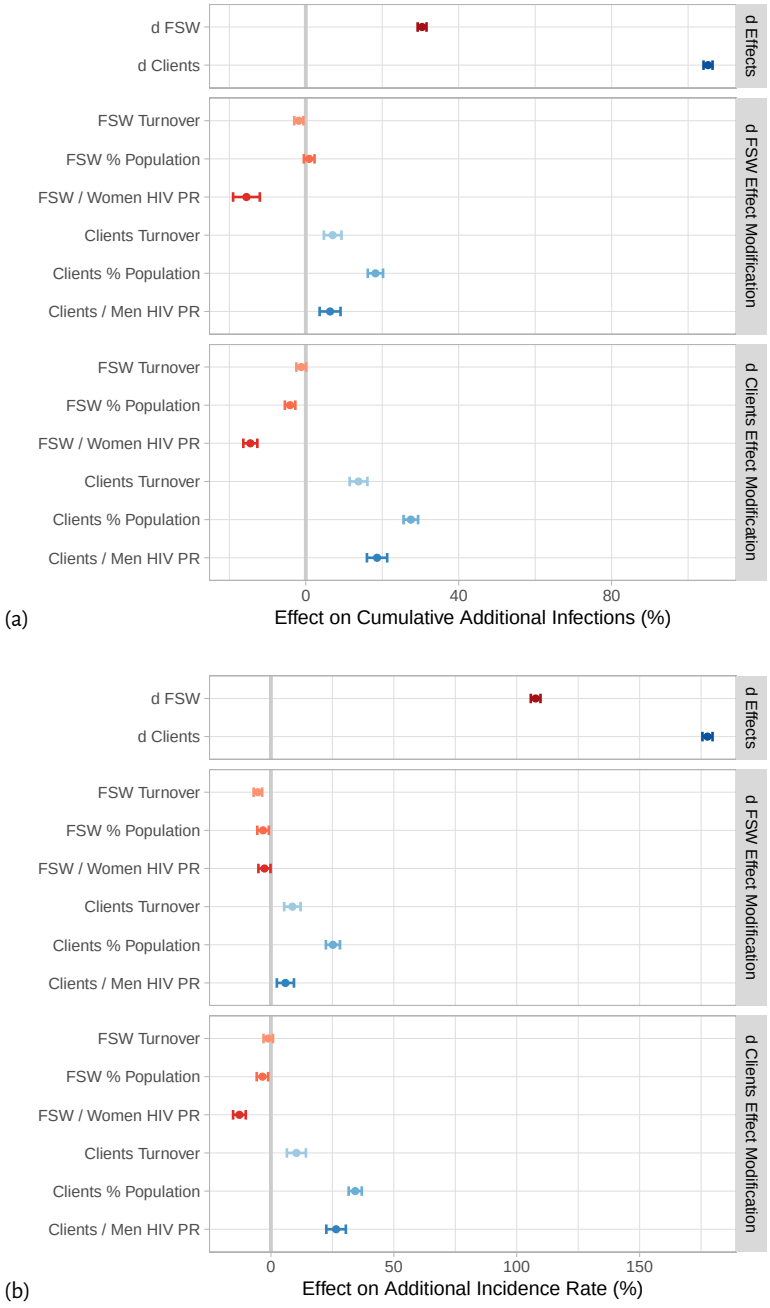

Figure 3: Estimated effects on relative additional infections of disproportionate viral non-suppression (d) among FSW and clients *vs* population overall, plus effect modification by epidemic conditions

(a) cumulative additional infections, (b) additional incidence rate by 2030 *vs* base case; FSW: female sex workers; Clients: of FSW; PR: prevalence ratio in 2005; d_i: difference in subpopulation-i-specific viral non-suppression *vs* population overall within counterfactual scenario; points and error bars: mean and 95% CI for each effect estimated via Eq. (2).

References

[1] Jens D Lundgren et al. "Initiation of antiretroviral therapy in early asymptomatic HIV infection." *New England Journal of Medicine* 373.9 (2015), pp. 795–807. https://doi.org/10.1056/NEJMoa1506816.

[2] Myron S. Cohen et al. "Antiretroviral Therapy for the Prevention of HIV-1 Transmission". *New England Journal of Medicine* 375.9 (2016), pp. 830–839. https://doi.org/10.1056/nejmoa1600693.

[3] UNAIDS. *The path that ends AIDS: 2023 UNAIDS Global AIDS Update*. Geneva, Switzerland, 2023.

[4] Stefan Baral et al. "The disconnect between individual-level and population-level HIV prevention benefits of antiretroviral treatment". *The Lancet HIV* 6 (2019), e632–638. http://doi.org/10.1016/s2352-3018(19)30226-7.

[5] Dylan Green et al. "Evidence of sociodemographic heterogeneity across the HIV treatment cascade and progress towards 90-90-90 in sub-Saharan Africa – a systematic review and meta-analysis". *Journal of the International AIDS Society* 23.3 (2020). https://doi.org/10.1002/jia2.25470.

[6] Mathieu Maheu-Giroux and Sharmistha Mishra. "Evidence from Botswana that ambitious is feasible: achieving 95-95-95 and what lies ahead". *The Lancet HIV* submitted (2024).

[7] Avi Joseph Hakim et al. "Gaps and opportunities: measuring the key population cascade through surveys and services to guide the HIV response". *Journal of the International AIDS Society* 21 (2018), e25119. http://doi.org/10.1002/jia2.25119.

[8] Kathryn E. Lancaster et al. "HIV care and treatment experiences among female sex workers living with HIV in sub-Saharan Africa: A systematic review". *African Journal of AIDS Research* 15.4 (2016), pp. 377–386. https://doi.org/10.2989/16085906.2016.1255652.

[9] Rhoda K. Wanyenze et al. ""If you tell people that you had sex with a fellow man, it is hard to be helped and treated": Barriers and opportunities for increasing access to HIV services among men who have sex with men in Uganda". *PLOS ONE* 11.1 (2016). https://doi.org/10.1371/journal.pone.0147714.

[10] Sheree Schwartz et al. "Engagement in the HIV care cascade and barriers to antiretroviral therapy uptake among female sex workers in Port Elizabeth, South Africa: Findings from a respondent-driven sampling study". *Sexually Transmitted Infections* 93.4 (2017), pp. 290–296. https://doi.org/10.1136/sextrans-2016-052773.

[11] Megan Schmidt-Sane. "Male partners of female sex workers: The intersectional risk environment of HIV/AIDS in a Kampala informal settlement". *Social Science and Medicine* 298 (2022), p. 114873. https://doi.org/10.1016/j.socscimed.2022.114873.

[12] Jesse Knight, Rupert Kaul, and Sharmistha Mishra. "Risk heterogeneity in compartmental HIV transmission models of ART as prevention in Sub-Saharan Africa: A scoping review". *Epidemics* 40 (2022), p. 100608. https://doi.org/10.1016/j.epidem.2022.100608.

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

[14] Eswatini Ministry of Health. *Eswatini Population-Based HIV Impact Assessment 3 (SHIMS3): Summary Sheet*. Mbabane, Eswatini, 2022.

[15] 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 (2009), pp. 118–129. https://doi.org/10.1016/S1473-3099(09)70021-0.

[16] Deborah Donnell et al. "Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis". *The Lancet* 375.9731 (2010), pp. 2092–2098. https://doi.org/10.1016/S0140-6736(10)60705-2.

[17] Steve E. Bellan et al. "Reassessment of HIV-1 Acute Phase Infectivity: Accounting for Heterogeneity and Study Design with Simulated Cohorts". *PLOS Medicine* 12.3 (2015), e1001801. https://doi.org/10.1371/journal.pmed.1001801.

[18] Tara D. Mangal. "Joint estimation of CD4+ cell progression and survival in untreated individuals with HIV-1 infection". *AIDS* 31.8 (2017), pp. 1073–1082. https://doi.org/10.1097/QAD.0000000000001437.

[19] Stefan Baral et al. "Reconceptualizing the HIV epidemiology and prevention needs of female sex workers (FSW) in Swaziland". *PLOS ONE* 9.12 (2014), e115465. http://doi.org/10.1371/journal.pone.0115465.

[20] PEPFAR. *Characterizing the HIV Prevention and Treatment Needs among Key Populations, including Men who Have Sex with Men and Female Sex Workers in Swaziland: From Evidence to Action*. Mbabane, Swaziland, 2015.

[21] Central Statistical Office Swaziland. *Swaziland Demographic and Health Survey 2006-07*. Mbabane, Swaziland: Central Statistical Office Swaziland, 2008. https://dhsprogram.com/.

[22] Swaziland Ministry of Health. *Swaziland HIV Incidence Measurement Survey: First Findings Report*. Mbabane, Swaziland, 2012. http://shims.icap.columbia.edu/.

[23] Eswatini Ministry of Health. *Swaziland HIV Incidence Measurement Survey 2 (SHIMS2) 2016-2017*. Mbabane, Eswatini, 2019.

[24] Andrew Mujugira et al. "HIV transmission risk persists during the first 6 months of antiretroviral therapy". *Journal of Acquired Immune Deficiency Syndromes* 72.5 (2016), pp. 579–584. https://doi.org/10.1097/QAI.0000000000001019.

[25] Delphine Gabillard et al. "Mortality, AIDS-morbidity, and loss to follow-up by current CD4 cell count among HIV-1-infected adults receiving antiretroviral therapy in Africa and Asia: Data from the ANRS 12222 collaboration". *Journal of Acquired Immune Deficiency Syndromes* 62.5 (2013), pp. 555–561. https://doi.org/10.1097/QAI.0b013e3182821821.

[26] Eswatini Ministry of Health. *2020–2021 Integrated Biological-Behavioral Surveillance Survey among Female Sex Workers and Men Who Have sex with Men in Eswatini*. Mbabane, Eswatini, 2022.

[27] Adrian E. Raftery and Le Bao. "Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling". *Biometrics* 66.4 (2010), pp. 1162–1173. http://doi.org/10.1111/j.1541-0420.2010.01399.x.

[28] MaxART Consortium. *MaxART Early Access to ART for All implementation study (2014-2018) Final Report*. Mbabane, Eswatini, 2018.

[29] Carol S. Camlin and Edwin D. Charlebois. "Mobility and its Effects on HIV Acquisition and Treatment Engagement: Recent Theoretical and Empirical Advances". *Current HIV/AIDS Reports* 16.4 (2019), pp. 314–323. https://doi.org/10.1007/s11904-019-00457-2.

[30] Jack Stone et al. "Estimating the contribution of key populations towards HIV transmission in South Africa". *Journal of the International AIDS Society* 24.1 (2021), e25650. https://doi.org/10.1002/jia2.25650.

[31] Romain Silhol et al. "Measuring HIV Acquisitions Among Partners of Key Populations: Estimates From HIV Transmission Dynamic Models". *JAIDS Journal of Acquired Immune Deficiency Syndromes* 95 (1S 2024), e59. https://doi.org/10.1097/QAI.0000000000003334.

[32] Luc Béhanzin et al. "Assessment of HIV-related risky behaviour: A comparative study of face-to-face interviews and polling booth surveys in the general population of Cotonou, Benin". *Sexually Transmitted Infections* 89.7 (2013), pp. 595–601. https://doi.org/10.1136/sextrans-2012-050884.

[33] Tafireyi Marukutira et al. "Modelling the impact of migrants on the success of the HIV care and treatment program in Botswana". *PLOS ONE* 15.1 (2020), e0226422. https://doi.org/10.1371/journal.pone.0226422.

[34] 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.

[35] Tim Brown and Wiwat Peerapatanapokin. "Evolving HIV epidemics: The urgent need to refocus on populations with risk". *Current Opinion in HIV and AIDS* 14.5 (2019), pp. 337–353. https://doi.org/10.1097/COH.0000000000000571.

[36] 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 (2013), e1001534. http://doi.org/10.1371/journal.pmed.1001534.

[37] 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 (2020), p. 10. https://doi.org/10.3389/frph.2020.00007.