**Supplementary Appendix**

Supplement to: Topazian HM, Stoner MCD, Edwards JK, et al. Variations in HIV risk by young women’s age and partner age-disparity in rural South Africa (HPTN 068).

**Appendix: Technical Methods**

**Parametric g-formula**

The parametric g-formula1,2 was utilized to simulate how HIV probability varies over AGYW age and partner age. We first ran pooled regression models using all person-visits from the observed data to determine the conditional probabilities of time-varying outcome, exposure, and covariates: maximum partner age difference over the next one-year (linear), acquisition of HIV over the next one-year (logistic), multiple partners (>1, logistic), ever had sex (logistic), alcohol use (logistic), and wealth (quartiles, multinomial logistic). Maximum partner age difference was modeled using a restricted quadratic spine and included interaction terms for current AGYW age and time (indicator variable for year since enrollment), and maximum partner age difference and time.

**Monte Carlo Simulation**

Next, we drew a large Monte Carlo sample of AGYW drawn with replacement from the observed data. HIV risk estimates stabilized once the simulated sample size reached 950,000 AGYW, contributing 4,443,683 visits. We retained only each AGYW’s baseline age and randomized controlled trial cash transfer intervention assignment. Coefficients from the pooled regression models above were used to simulate time-varying confounders, maximum partner age difference, and HIV outcome data by calculating the weighted probability of a new covariate given an AGYW’s presently simulated covariate pattern. Values were sampled from a Bernoulli distribution for binary variables and a multinomial distribution for the categorical wealth variable. To align with the age pairings in our original dataset, simulated maximum partner ages were drawn from a normal distribution with a minimum of 10 years. Maximum partner ages were only allocated to AGYW who had ever had sex, as assigned through the model. Conditional probabilities were estimated at up to five time points to simulate outcome, exposure, and covariate values over the study period, one for each annual study visit. Once a participant turned 21 years of age or tested positive for HIV, she was simulated as exiting the risk set. We transformed log-odds of HIV values into predicted probabilities and computed averages at intervals by AGYW age and partner age difference, and by AGYW age and partner age group in the simulated dataset.

**Bootstrapping**

Bootstrapping was employed to calculate 95% confidence intervals. To bootstrap, we took 400 samples of 950,000 AGYW drawn with replacement from the observed data with the original sample size. We repeated the Monte Carlo simulation described above for each of the 400 samples, obtaining 400 predicted probabilities for each age pairing between AGYW and sexual partners. The standard deviation of the 400 predicted probabilities was used to calculate 95% confidence intervals around our originally simulated results, retaining the uncertainty from the original dataset.

**References**

1 Edwards JK, McGrath LJ, Buckley JP, Schubauer-Berigan MK, Cole SR, Richardson DB. Occupational radon exposure and lung cancer mortality: estimating intervention effects using the parametric g-formula. *Epidemiology* 2014; **25**: 829–34.

2 Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The Parametric g-Formula for Time-to-event Data. *Epidemiology* 2014; **25**: 889–97.

**Appendix: Table 1.** The total number of HIV events recorded (numerator) and the total number of partnerships reported (denominator) for each age pairing between AGYW and sexual partners. AGYW age and partner age are taken from the same visit at which an HIV test was performed. This table represents all age pairings captured in Figure 1.

|  |  |  |
| --- | --- | --- |
|  | **Partner Age** |  |
| AGYW Age | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **>35** | **Overall** |
| 13 | 0/1 | 0/1 | - | 0/6 | 0/1 | 0/1 | 0/1 | 0/1 | 0/1 | - | - | 0/1 | - | - | 0/1 | - | - | - | - | - | - | - | - | - | - | - | **15** |
| 14 | 0/1 | - | - | 0/2 | 0/7 | 0/12 | 0/16 | 0/9 | 0/12 | 0/5 | 0/1 | - | - | 0/1 | - | - | - | - | - | - | - | - | - | - | - | - | **66** |
| 15 | - | 0/1 | - | 0/4 | 0/24 | 1/41 | 0/44 | 4/70 | 0/43 | 0/17 | 0/7 | 0/8 | 0/4 | - | 0/1 | - | - | - | - | 0/3 | - | - | 0/3 | - | - | 0/3 | **273** |
| 16 | 0/2 | - | - | 0/1 | 0/5 | 0/40 | 1/112 | 6/125 | 4/122 | 0/78 | 0/51 | 0/31 | 0/17 | 0/8 | 0/7 | 1/3 | 0/3 | 0/3 | 0/1 | 0/1 | - | - | - | - | - | 0/1 | **611** |
| 17 | 0/3 | 0/4 | - | 0/1 | 0/10 | 1/20 | 0/78 | 3/175 | 2/201 | 3/184 | 4/119 | 0/52 | 1/45 | 1/9 | 4/22 | 0/6 | 2/8 | 3/5 | 0/6 | 0/2 | - | 0/3 | 0/1 | - | - | 1/3 | **957** |
| 18 | 0/2 | 0/4 | - | 0/1 | 0/5 | 1/7 | 0/25 | 2/110 | 2/199 | 3/200 | 6/151 | 3/112 | 3/74 | 2/38 | 4/29 | 1/13 | 0/6 | 1/3 | 0/4 | - | - | 0/1 | - | 0/1 | 0/1 | 0/1 | **987** |
| 19 | 0/1 | 0/1 | 0/1 | - | - | 0/2 | 0/14 | 0/29 | 1/94 | 2/143 | 2/148 | 5/138 | 4/129 | 4/58 | 6/61 | 3/35 | 2/13 | 2/13 | 0/6 | - | 0/2 | 1/4 | 0/1 | 0/1 | 0/1 | 0/3 | **898** |
| 20 | 0/3 | 0/1 | - | - | - | 0/1 | 0/2 | 3/18 | 2/16 | 3/76 | 9/128 | 9/105 | 5/117 | 5/89 | 10/91 | 5/45 | 5/27 | 2/17 | 1/6 | 1/8 | 0/2 | 0/2 | 0/3 | 0/2 | - | 0/4 | **763** |
| 21 | - | - | - | - | - | - | - | 0/3 | 0/10 | 0/17 | 1/42 | 4/67 | 4/73 | 2/58 | 3/75 | 2/39 | 2/26 | 3/14 | 1/17 | 2/11 | 1/2 | 1/5 | - | - | - | 0/5 | **464** |
| 22 | - | - | - | 0/1 | - | 1/1 | 2/2 | 0/1 | 0/2 | 1/7 | 0/5 | 2/26 | 2/38 | 1/25 | 5/25 | 2/28 | 1/10 | 1/14 | 0/8 | 1/8 | 0/6 | 1/3 | 0/1 | 0/3 | - | 0/1 | **215** |
| 23 | - | - | - | - | - | - | 1/1 | - | 0/1 | 0/2 | - | 0/3 | 1/10 | 1/13 | 4/26 | 1/12 | 0/7 | 0/5 | 1/3 | 0/2 | - | 0/1 | 1/1 | - | - | 0/1 | **88** |
| 24 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 0/3 | 0/1 | 0/3 | 0/1 | - | - | - | 0/1 | - | 0/1 | - | - | **10** |
| 25 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 0/1 | 0/1 | 0/1 | - | - | - | - | - | - | - | - | - | **3** |
| 26 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 1/1 | - | - | - | - | - | - | - | - | - | - | - | **1** |
| Overall | **13** | **12** | **1** | **16** | **52** | **125** | **295** | **541** | **701** | **729** | **652** | **543** | **507** | **299** | **343** | **183** | **104** | **75** | **51** | **35** | **12** | **20** | **10** | **8** | **2** | **22** | **5,351** |

**Appendix: Table 2.** The proportion of the total number of partnerships at each age pairing in which an AGYW became infected with HIV. AGYW age and partner age are taken from the same visit at which an HIV test was performed. This table represents all age pairings captured in Figure 1.

|  |  |  |
| --- | --- | --- |
|  | **Partner Age** |  |
| AGYW Age | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** | **26** | **27** | **28** | **29** | **30** | **31** | **32** | **33** | **34** | **35** | **>35** | **Overall** |
| 13 | 0 | 0 | - | 0 | 0 | 0 | 0 | 0 | 0 | - | - | 0 | - | - | 0 | - | - | - | - | - | - | - | - | - | - | - | **0** |
| 14 | 0 | - | - | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - | - | 0 | - | - | - | - | - | - | - | - | - | - | - | - | **0** |
| 15 | - | 0 | - | 0 | 0 | 0·02 | 0 | 0·06 | 0 | 0 | 0 | 0 | 0 | - | 0 | - | - | - | - | 0 | - | - | 0 | - | - | 0 | **0·02** |
| 16 | 0 | - | - | 0 | 0 | 0 | 0·01 | 0·05 | 0·03 | 0 | 0 | 0 | 0 | 0 | 0 | 0·33 | 0 | 0 | 0 | 0 | - | - | - | - | - | 0 | **0·02** |
| 17 | 0 | 0 | - | 0 | 0 | 0·05 | 0 | 0·02 | 0·01 | 0·02 | 0·03 | 0 | 0·02 | 0·11 | 0·18 | 0 | 0·25 | 0·6 | 0 | 0 | - | 0 | 0 | - | - | 0·33 | **0·03** |
| 18 | 0 | 0 | - | 0 | 0 | 0·14 | 0 | 0·02 | 0·01 | 0·02 | 0·04 | 0·03 | 0·04 | 0·05 | 0·14 | 0·08 | 0 | 0·33 | 0 | - | - | 0 | - | 0 | 0 | 0 | **0·03** |
| 19 | 0 | 0 | 0 | - | - | 0 | 0 | 0 | 0·01 | 0·01 | 0·01 | 0·04 | 0·03 | 0·07 | 0·1 | 0·09 | 0·15 | 0·15 | 0 | - | 0 | 0·25 | 0 | 0 | 0 | 0 | **0·04** |
| 20 | 0 | 0 | - | - | - | 0 | 0 | 0·17 | 0·13 | 0·04 | 0·07 | 0·09 | 0·04 | 0·06 | 0·11 | 0·11 | 0·19 | 0·12 | 0·17 | 0·13 | 0 | 0 | 0 | 0 | - | 0 | **0·08** |
| 21 | - | - | - | - | - | - | - | 0 | 0 | 0 | 0·02 | 0·06 | 0·05 | 0·03 | 0·04 | 0·05 | 0·08 | 0·21 | 0·06 | 0·18 | 0·5 | 0·2 | - | - | - | 0 | **0·06** |
| 22 | - | - | - | 0 | - | 1 | 1 | 0 | 0 | 0·14 | 0 | 0·08 | 0·05 | 0·04 | 0·2 | 0·07 | 0·1 | 0·07 | 0 | 0·13 | 0 | 0·33 | 0 | 0 | - | 0 | **0·09** |
| 23 | - | - | - | - | - | - | 1 | - | 0 | 0 | - | 0 | 0·1 | 0·08 | 0·15 | 0·08 | 0 | 0 | 0·33 | 0 | - | 0 | 1 | - | - | 0 | **0·11** |
| 24 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 0 | 0 | 0 | 0 | - | - | - | 0 | - | 0 | - | - | **0** |
| 25 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 0 | 0 | 0 | - | - | - | - | - | - | - | - | - | **0** |
| 26 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 1 | - | - | - | - | - | - | - | - | - | - | - | **1** |
| Overall | **0** | **0** | **0** | **0** | **0** | **0·03** | **0·01** | **0·03** | **0·02** | **0·02** | **0·03** | **0·04** | **0·04** | **0·05** | **0·11** | **0·08** | **0·12** | **0·16** | **0·06** | **0·11** | **0·08** | **0·15** | **0·10** | **0** | **0** | **0·05** | **0·04** |

**Appendix: Figure 1.** One-year risk of HIV among AGYW by age and by the age of their reported sexual partners. The AGYW age line shows one-year risk in the overall cohort, regardless if partner age was reported. The maximum partner age line shows the likelihood of an AGYW with a partner of that age recording an HIV infection at an annual study visit. Only pairings with more than 15 observations are displayed.





**Appendix: Table 3**.AGYW risk of HIV in the next one-year at 0, 5, 10, and 15 year maximum partner age differences in the simulated dataset (N=950,000), as shown in Figure 3. Risk estimates are stratified by AGYW age group (years). Absolute estimates of the risk of HIV in each AGYW age group are calculated only among AGYW who were assigned partners in the simulation, regardless of partner age. Estimates from pairings in the 13 to 14 years stratum with partner age differences of 10 and 15 years, and estimates from pairings in all strata with a partner age difference of 15 years are extrapolated from sparse data in the observational cohort and should be interpreted with caution.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Young Woman’s Age |  | **Absolute one-year risk of HIV** |  | **One-year risk by partner age** |  | **One-year risk difference** |
|  |  | 0 years | +5 years | +10 years | +15 years |  | 0 years | +5 years  | +10 years  | +15 years |
| 13 to 14 |  | 0·05 |  | 0·01 | 0·05 | 0·21 | 0.24 |  | *ref* | +0·04 | +0·20 | +0.23 |
| 15 to 16 |  | 0·04 |  | 0·01 | 0·05 | 0·11 | 0·10 |  | *ref* | +0·03 | +0·10 | +0·09 |
| 17 to 18 |  | 0·04 |  | 0·02 | 0·05 | 0·09 | 0·07 |  | *ref* | +0·03 | +0·07 | +0·05 |
| 19 to 21 |  | 0·07 |  | 0·05 | 0·07 | 0·09 | 0·02 |  | *ref* | +0·02 | +0·04 | -0·03 |
| Overall |  | 0.05 |  | 0·03 | 0·06 | 0·10 | 0·08 |  | *ref* | +0·03 | +0·07 | +0·05 |