Supplementary Digital Content (SDC) for:

Title: Do age-disparate relationships drive HIV incidence in young women? Evidence from a population cohort in rural KwaZulu-Natal, South Africa

Running Head: Age-disparities & HIV in rural South Africa

Authors: Guy Harling ScD [1], Marie-Louise Newell PhD [2,3], Frank Tanser PhD [2], Ichiro Kawachi MD, PhD [1], SV Subramanian PhD [1], Till Bärnighausen MD, ScD [2,4].

1. Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Avenue, Boston MA, 02115, USA
2. Africa Centre for Health and Population Studies, University of KwaZulu-Natal, P.O. Box 198, Mtubatuba, 3935, South Africa
3. Faculty of Medicine, University of Southampton, Southampton, SO171BJ, UK
4. Department of Global Health and Population, Harvard School of Public Health, 677 Huntington Avenue, Boston MA, 02115, USA

Corresponding author: Guy Harling. Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Avenue, Boston MA, 02115, USA. Telephone: +1 617 432 1135. Fax: +1 617 432 3123. Email: gharling@hsph.harvard.edu.

Table, SDC 1: Cox proportional hazards models for appropriate functional form (hazard ratios, 95% confidence intervals)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Respondent's age (in years, centred at age 15)							
Linear	0.94	1.20	1.18	1.11	1.18	1.18	1.18
	(0.91 - 0.97)	(1.09 - 1.33)	(1.07 - 1.29)	(0.95 - 1.30)	(1.07 - 1.29)	(1.07 - 1.29)	(1.07 - 1.29)
Quadratic†		0.98	0.97	0.97	0.97	0.97	0.97
		(0.97 - 0.99)	(0.96 - 0.98)	(0.96 - 0.98)	(0.96 - 0.98)	(0.96 - 0.98)	(0.96 - 0.98)
Cubic†			2.81	10.94	2.81	2.81	2.81
			(1.65 - 4.78)	(0.51 - 234.5)	(1.65 - 4.78)	(1.65 - 4.78)	(1.65 - 4.78)
Quartic†				0.44			
				(0.07 - 2.74)			
Most recent partner's age-disparity (in years, centred at 0)							
Linear					1.00	1.01	1.00
					(0.97 - 1.03)	(0.96 - 1.06)	(0.94 - 1.07)
Quadratic†						1.00	1.00
						(1.00 - 1.00)	(0.99 - 1.01)
Cubic†							0.95
							(0.73 - 1.23)
Akaike Information Criteria	6,510	6,490	6,478	6,480	6,480	6,482	6,484

For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. All models contain indicator variables for year of observation (not shown). † Coefficients for squared, cubic and quartic terms represent 10, 100 and 1000 unit change in the respective variables.

Table, SDC 2: Descriptive statistics for multiply imputed datasets

	Multiply imputed dataset	Complete case dataset	Missingness (%)	Change (%) in incidence rate**
All women aged 15-29 years old				
Time at risk (person-years)	7,722	5,913	23.4	
Number of respondents at baseline	3,358	2,444	27.2	
Number of seroconversions	584	458		
Incidence rate (per 100 person-years)	7.56	7.75		-2.5
95% confidence interval	(6.97 - 8.20)	(7.07 - 8.49)		
15-19 years old				
Time at risk (person-years)	2,599	1,746	32.8	
Number of respondents at baseline	1,770	1,112	37.2	
Number of seroconversions	222	136		
Incidence rate (per 100 person-years)	8.54	7.79		8.8
95% confidence interval	(7.49 - 9.74)	(6.59 - 9.22)		
20-24 years old				
Time at risk (person-years)	3,668	2,907	20.7	
Number of respondents at baseline	1,128	982	12.9	
Number of seroconversions	285	251		
Incidence rate (per 100 person-years)	7.77	8.63		-11.1
95% confidence interval	(6.92 - 8.73)	(7.63 - 9.77)		
25-29 years old				
Time at risk (person-years)	1,455	1,260	13.4	
Number of respondents at baseline	460	350	23.9	
Number of seroconversions	77	71		
Incidence rate (per 100 person-years)	5.29	5.63		-6.4
95% confidence interval	(4.23 - 6.62)	(4.46 - 7.11)		

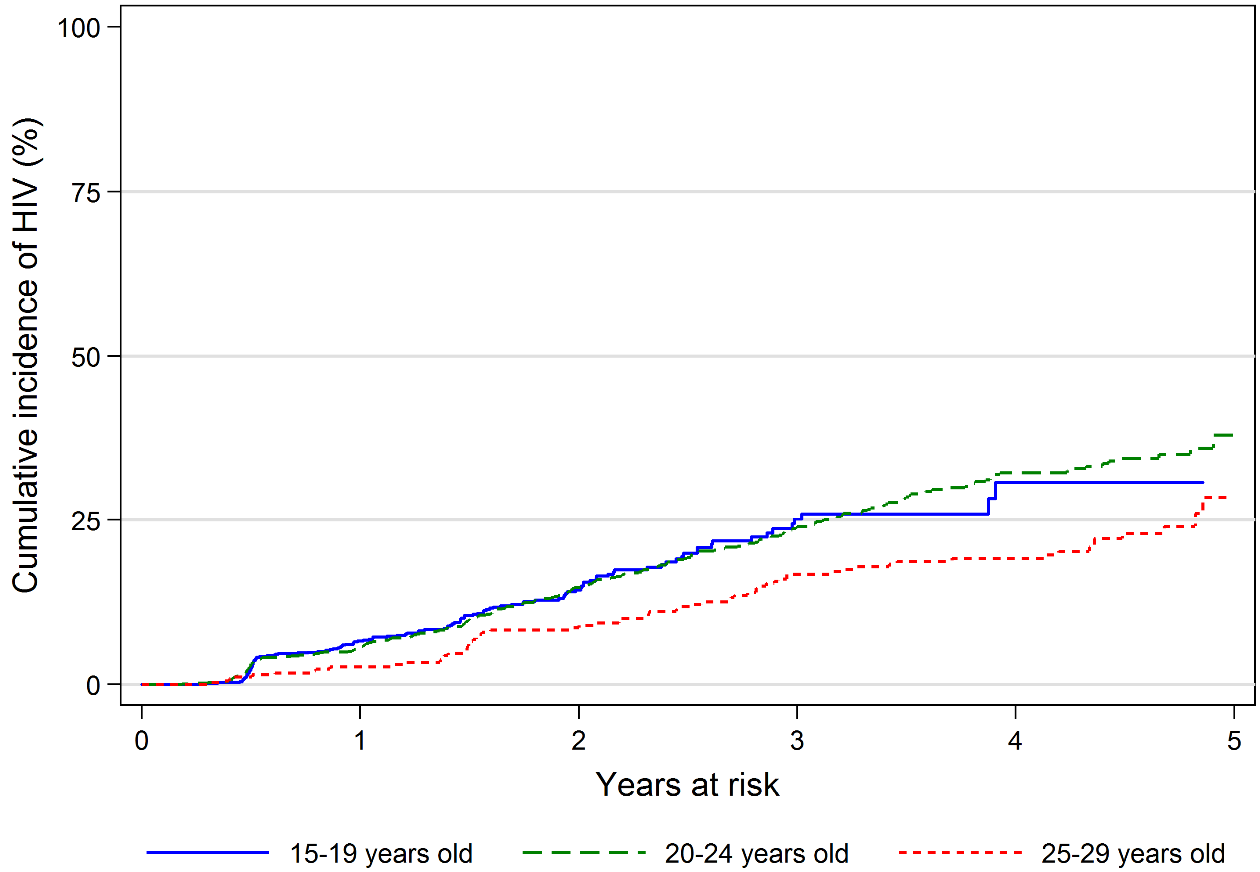
	Incidence rate for all valid responses	Missingness (%)*	Change (%) in incidence rate**
Age disparity of partner in most recent relationship	7.69	17.6	-1.7
Highest educational attainment	7.56	-	0.0
Household wealth quintile	7.38	4.9	2.4
Marital status	7.50	0.4	0.8
Age at sexual debut	7.59	0.7	-0.4
Multiple partners in past 12 months	7.50	5.0	0.8
Casual partner in past 12 months	7.56	-	0.0
Lowest condom use level in relationships in past 12 months	7.49	0.9	0.9

* Missingness for covariates is the amount of person-time with missing values prior to imputation for each variable, as a proportion of the 7,722 years of person-time in the Multiply Imputed dataset. ** Change in the incidence rate comparing all person-time in the imputed dataset to all person-time in the complete-case dataset (in the top half of the table this is the complete-case dataset for all independent variables; in the bottom half it is the complete-case dataset for each individual variable).

The dataset for multiple imputation consisted of all person-time for which: (i) there was a valid HIV test result at the beginning and end of the observation period; (ii) the respondent had participated in the General Health module by the beginning of the period; and (iii) we were unable to exclude the possibilities that they had yet to reach their sexual debut (i.e. we excluded women who reported never having had sex, or who had never reported any sexual partner up to the beginning of the period). Practically, this meant adding to the complete case dataset observation time for which: (a) partner age-disparity data was missing (1,406 person-years); or (b) covariate information was missing (an additional 403 person-years).

Missing data were imputed 20 times using chained imputations. The chained imputation process used all variables included in analyses in this paper, plus several other variables that might be expected to be correlated with sexual and socio-demographic variables: urban/suburban/rural household; distance to nearest health clinic, major road, minor road, primary school and secondary school; Isigodi (traditional Zulu area) of household; currently economically active (yes/no); and year of visit.

Figure, SDC 3: Kaplan-Meier cumulative incidence curve for HIV incidence during follow-up, stratified by respondent's baseline age



Table, SDC 4: Regression results from multiply imputed dataset (hazard ratios and 95% confidence intervals)

	Model 1		Model 2†		Model 3		Model 4	
Age-disparity (one year increase in partner's age)	0.96	(0.85 - 1.09)			0.98	(0.86 - 1.12)	0.97	(0.87 - 1.12)
Respondent aged 15-19			1.06	(0.86 - 1.30)				
Respondent aged 20-24			0.89	(0.73 - 1.08)				
Respondent aged 25-29			0.98	(0.75 - 1.27)				
Age of respondent (centred at 15 years old)*								
Age	0.85	(0.67 - 1.07)	0.85	(0.66 - 1.06)	0.88	(0.69 - 1.10)	0.88	(0.69 - 1.10)
Age squared	1.15	(0.78 - 1.69)	1.13	(0.76 - 1.65)	1.13	(0.76 - 1.65)	1.13	(0.77 - 1.65)
Age cubed	0.93	(0.77 - 1.12)	0.94	(0.77 - 1.14)	0.94	(0.77 - 1.13)	0.94	(0.77 - 1.13)
Highest educational attainment								
None or Primary (0-7 years)					1.25	(0.95 - 1.64)	1.24	(0.93 - 1.63)
Secondary (8-12 years)					1.00		1.00	
Tertiary					0.74	(0.46 - 1.18)	0.74	(0.46 - 1.17)
Household wealth quintile								
Lowest					1.32	(0.92 - 1.87)	1.33	(0.93 - 1.88)
2nd lowest					1.25	(0.90 - 1.72)	1.27	(0.91 - 1.76)
Middle					1.42	(1.02 - 1.96)	1.44	(1.04 - 1.99)
2nd highest					1.31	(0.92 - 1.83)	1.32	(0.93 - 1.85)
Highest					1.00		1.00	
Current marital status of respondent								
Never Married					1.00		1.00	
Engaged					0.91	(0.68 - 1.20)	0.93	(0.69 - 1.23)
Married					0.33	(0.11 - 0.90)	0.33	(0.12 - 0.92)
Divorced/Separated/Widowed					1.56	(0.21 - 11.2)	1.65	(0.22 - 11.9)
Age at sexual debut (one-year increment)							1.00	(0.94 - 1.05)
Any casual partner in past 12 months							1.11	(0.90 - 1.37)
Multiple partners in past 12 months							1.78	(1.01 - 3.11)
Lowest condom use level in relationships in past 12 months								
Never							1.00	
Sometimes							0.92	(0.75 - 1.12)
Always							1.11	(0.90 - 1.37)

For all models, n=2,972, time at risk = 7,722 person-years and there were 584 seroconversions All models contain indicator variables for year of observation (not shown). * The coefficient on the variable "age squared" represents a 10-unit change in this variable; the coefficient on the variable "age cubed" represents a 100-unit change in this variable. † This model also contains indicator variables for age categories. A joint test for equality on the three age by age-disparity interaction terms was not statistically significant (χ^2 = 1.42, p-value: 0.492).

Table, SDC 5: Sensitivity analyses using binary age-disparity categories

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age of respondent						
Linear	1.18	1.18	1.18	1.19	1.18	1.18
	(1.07 - 1.29)	(1.07 - 1.30)	(1.07 - 1.29)	(1.08 - 1.30)	(1.07 - 1.29)	(1.07 - 1.29)
Quadratic	0.73	0.73	0.73	0.73	0.73	0.73
	(0.67 - 0.80)	(0.67 - 0.80)	(0.67 - 0.80)	(0.67 - 0.80)	(0.67 - 0.80)	(0.67 - 0.81)
Cubic	1.11	1.11	1.11	1.10	1.11	1.11
	(1.05 - 1.17)	(1.05 - 1.17)	(1.05 - 1.17)	(1.05 - 1.17)	(1.05 - 1.17)	(1.05 - 1.17)
Age-disparity ≥5 years	0.98					
	(0.81 - 1.20)					
Respondent aged 15-19		1.00				
		(0.71 - 1.42)				
Respondent aged 20-24		0.98				
		(0.75 - 1.28)				
Respondent aged 25-29		0.96				
		(0.60 - 1.52)				
Age-disparity ≥10 years			0.98			
			(0.67 - 1.43)			
Respondent aged 15-19				1.38		
				(0.68 - 2.81)		
Respondent aged 20-24				0.77		
				(0.42 - 1.41)		
Respondent aged 25-29				1.05		
				(0.54 - 2.04)		
Age-disparity ≥20 years					0.61	
					(0.15 - 2.46)	
Respondent aged 15-19						1.69
						(0.24 - 12.10)
Respondent aged 20-24						0.52
						(0.07 - 3.70)
Respondent aged 25-29						-
Akaike Information Criteria	6,474	6,480	6,474	6,478	6,476	6,478

For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. Values are hazard ratios and 95% confidence intervals from Cox proportional hazards moldes. . All models contain indicator variables for year of observation (not shown). Hazard ratios for age-disparity in each age-group in models 2, 4 and 6 compare those in age-disparate relationships in each age-group to all those in non age-disparate relationships, after allowing for respondents' own age.

Table, SDC 6: Sensitivity analyses using interactions of socio-demographic variables and age-disparity measures

Functional form of age-disparity	Continuous (one-year increase in partner's age)	Binary, partner ≥5 years older	Binary, partner ≥10 years older	Binary, partner ≥20 years older
Current marital status				
Never married	1.01	1.09	1.21	0.29
	(0.93 - 1.09)	(0.57 - 2.08)	(0.40 - 3.69)	(0.02 - 4.77)
Engaged, married or formerly married	1.00	0.93	0.87	1.59
	(0.92 - 1.08)	(0.51 - 1.72)	(0.31 - 2.47)	(0.22 - 11.73)
χ^2 test of equality	0.02	0.06	0.09	0.56
p-value	0.90	0.81	0.76	0.45
Highest educational attainment				
None or Primary (0-7 years)	0.99	1.14	0.98	
	(0.92 - 1.07)	(0.62 - 2.11)	(0.38 - 2.54)	
Secondary (8-12 years)	1.01	1.03	1.11	1.01
	(0.98 - 1.04)	(0.83 - 1.28)	(0.73 - 1.69)	(0.25 - 4.08)
Tertiary	0.84	0.42		
	(0.70 - 1.01)	(0.14 - 1.27)		
χ^2 test of equality	3.54	2.63	0.05	-
p-value	0.17	0.27	0.82	
Household wealth quintile				
Lowest	0.97	0.58	0.75	
	(0.90 - 1.04)	(0.34 - 1.00)	(0.30 - 1.86)	
2nd lowest	0.99	1.04	1.35	1.28
	(0.93 - 1.06)	(0.70 - 1.54)	(0.68 - 2.68)	(0.18 - 9.30)
Middle	1.02	1.56	1.32	1.29
	(0.98 - 1.07)	(1.09 - 2.23)	(0.69 - 2.53)	(0.18 - 9.28)
2nd highest	0.99	1.01	0.83	
	(0.92 - 1.07)	(0.63 - 1.60)	(0.30 - 2.28)	
Highest	0.87	0.52	0.39	
	(0.77 - 0.98)	(0.24 - 1.12)	(0.05 - 2.86)	
χ^2 test of equality	5.77	12.74	2.66	0.00
p-value	0.22	0.01	0.62	1.00

Values are hazard ratios and 95% confidence intervals from Cox proportional hazards models. This table contains summary results for 12 separate regression models (four outcome forms for each of three interacted covariates). For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. All models contain indicator variables for year of observation, age of respondent (linear, quadratic and cubic terms centred at age 15) and all socio-demographic and behavioural covariates from the main analysis. Empty cells in this table reflect categories in which there were so few observed individuals with age-disparate partners that they had to be removed in order for models to converge.