Marital Shopping and Epidemic AIDS: Web Appendix

Jeremy R. Magruder

1 Data Description and Summary statistics

Four data sources are used in this article.

1) The Cape Area Panel Study[1] (Lam et al 2006) is a random sample of 4758 young adults aged 14-22 in 2002 who live in the Cape Town Metropolitan Area. These young adults were interviewed first in 2002. A subset of 1360 young adults were reinterviewed in 2003, with the remainder reinterviewed in 2004, and all were reinterviewed in 2005. This paper utilizes information on annual partners collected in 2002 and 2005.

2) DHS data: this refers to the 1998 South Africa, 2003 Kenya, 2004 Lesotho, 2004 Malawi, 2003 Tanzania, and 2005/06 Zimbabwe DHS surveys. DHS survey information is available at www.measuredhs.com. Table 1 provides summary statistics from these data.

3) Labour Force Survey: This is a twice yearly rotating panel of households in South Africa designed to examine employment status and demographic variables across South Africa, with sampling done from 1996 census blocks. More information and data requests are available at the South African Data Archive, www.nrf.ac.za/sada

4) NSFG data: this refers to the 2002 National Survey of Family Growth, collected by the National Center of Health Statistics (NCHS). The survey population is designed to be representative of the United States, and more females than males were surveyed (7643 versus 4928). More information is available at www.cdc.goc/nchs/nsfg.htm.

Simulated Data is created according to parametric assumptions given in table 2, while summary stats (from 2500 simulated observations) are contained in table 3.

2 Age-Death Algebra

Let $S(t)$ be the survival rate after t years of infection and $I(a,t)$ be the overall incidence of HIV at age a in time period t. The number of deaths from HIV at age a in year t is given by

$$D(a,t) = \sum_{r=0}^{a} I(a-r, t-r)\left(S(r) - S(r+1)\right) \tag{1}$$

[1]The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation.

The marital shopping model suggests that a sharp change in risk behaviors occurs at time at marriage, so to infer death rates it is convenient to partition individuals into married versus single. That is, if $I^S(a,t)$ is the fraction of individuals who are both single and infected at age a and time t, and $I^M(a,t)$ is the same for married individuals, we have

$$D(a,t) = \sum_{r=0}^{a} (S(r) - S(r+1)) \left(I^S(a-r,t-r) + I^M(a-r,t-r)\right) \tag{2}$$

Suppose $\iota^S(t)$ is the raw incidence rate for single individuals at time t, or the fraction of single individuals who become infected in that time period. Under the matching model, single individuals all behave the same, and hence are at equal risk, with one caveat. HIV is an absorbing state, and so individuals who have already been infected cannot become infected again. Hence let $i^S(k,t)$ represent the risk of becoming infected for an individual who has been sexually active for k years. Moreover, due to HIV's absorbing nature, the raw incidence rate has to be multiplied by the fraction of individuals who can still be infected, that is, $(1-\psi^s(t))$ if $\psi^s(t)$ is the single prevalence rate at time t in order to generate the true risk that a single individual faces. Then

$$i^s(k,t) = \frac{\iota^s(t)}{1-\psi^S(t)} \prod_{r=0}^{k} \left(1 - \frac{\iota^S(r)}{1-\psi^S(r)}\right)$$

and

$$I^S(a,t) = \sum_{k=0}^{a} i^S(k,t)\,\xi^S(a,k)$$

if $\xi^S(a,k)$ is the percentage of women who are single and searching at age a and who have been searching for k years.

In the marital shopping model, married individuals face an incidence rate which declines exponentially at the annual transmission probability, ρ. In particular, let $i^M(\mu,k,t)$ be the risk of infection for an individual who married μ years earlier after k years of search in period t. then

$$i^M(\mu,k,t) = \psi^m(t-\mu)\,\rho\,(1-\rho)^{\mu-1} \prod_{r=0}^{k} \left(1 - \frac{\iota^s(t-\mu-r)}{1-\psi^S(t-\mu-r)}\right) \tag{3}$$

where $\psi^m(t)$ is the prevalence rate among newlyweds, which is in general different from the single prevalence due to the declining reservation quality with age. Hence

$$I^M(a,t) = \sum_{\mu=0}^{a} \sum_{k=0}^{a-\mu} i^M(\mu,k,t)\,\xi^M(a-\mu,k)$$

where $\xi^M(a-\mu,k)$ is the percentage of women who married at age $\alpha-\mu$ after having been single for k years. I assume independence between age of sexual onset and age of marriage; in the South African DHS data these are uncorrelated. Hence if $M(a)$ represents the fraction of individuals who are married at age a, $\xi^S(a,k) = (1-M(a))\,\tilde{\xi}^S(k)$, and $\xi^M(a,k) = M(a)\,\tilde{\xi}^M(k)$, where $\tilde{\xi}^S(k)\left(\tilde{\xi}^M(k)\right)$ is the proportion of single (married) individuals who have actively searched for k years. Since this is unobservable, I assume it is proportional to the percentage of women who report having had sex for the first time at age k years earlier. Search intensity while single seems likely to be different for currently married individuals – in particular, for age-a married individuals, $\int_0^a \tilde{\xi}^M(k) = 1$, whereas there is no such implication for $\tilde{\xi}^S(k)$; so if $X(k)$ is the distribution of individuals who report sexual onset at age k, then I assume that $\xi^S(a,k) = \gamma^S(1-M(a))X(a-k)$ and $\xi^M(a,k) = \gamma^M(M(a))X(a-k)$. Therefore

$$\frac{D(a,t)}{Popn} = \sum_{r=0}^{a}(S(r) - S(r+1)) * \tag{4}$$
$$\begin{pmatrix} \gamma^S(1-M(a))\sum_{k=0}^{a} i^S(k,t)(X(a-k) - X(a-k-1)) + \\ \gamma^M M(a)\sum_{\mu=0}^{a}\sum_{k=0}^{a-\mu} i^M(\mu,k,t)(X(a-\mu-k) - X(a-\mu-k-1)) \end{pmatrix}$$

Using DHS data, I estimate a Kaplan-Meier survival function out of singlehood for African Women in South Africa, and a similar survival function into sexual activity. For men, for whom there is no South African DHS data, tabulations of percent never married are taken at each age from the September 2001 South African Labour Force Survey, and beginning sexual search is calibrated in two ways: as being identical to the female distribution of coital onset, and as being the female distribution plus five years (as the average married male is five years older than his spouse in South Africa). The survival function for HIV is taken from UNAIDS(2002). ρ, the incidence per year of relationship, is set to .20, similar to Gray's (2001) finding for young couples, and non-AIDS deaths are taken to be identical to those in 1996. Time-paths of single incidence rates, single prevalence rates, and newlywed prevalence rates are simulated with the model, allowing identification of everything but γ^S and γ^M. In other words, at time t, we know the shape of deaths from infections which the married incur and the shape of deaths from infections incurred by the single but not how to weight those curves in adding them. At time t, we also know the ratio of the total deaths from married infections to that of single infections. That is,

$$D^m(t) = \sum_{p=1}^{t-1}\sum_{j=p}^{t-1}\mu(p)\psi^m(p) * (1-\psi^m(p))\rho(1-\rho)^{j-p} * (S(t-j) - S(t-j+1))$$

Where $\mu(p)$ represents the number of individuals who are married in year p. In turn,

$$D^S(t) = SandS * \sum_{j=1}^{t}\iota^S(j)(S(t-j) - S(t-j+1))$$

where $SandS$ identifies the number of single and searching individuals. If search lasts on average nine years, as in South Africa, then this corresponds to about 8% of single and searching people being married per year. Therefore, $\mu(p) \cong .08 * SandS\ \forall p$, and we have all of the information to determine $D^S(t)/D^M(t)$, which in turn identifies γ^S/γ^M,meaning that I can identify the death rate up to a constant. In year 15, $D^S/D^m \cong 1.85$.

3 Analysis in Other African Countries

There are two challenges in exporting the analysis presented in the main text to countries other than South Africa. First, there is (much) greater prevalence both of concurrent and of consecutive marriages in the other countries of sub-Saharan Africa (14% for women over 30 in South Africa, vs. e.g. 28% in Zimbabwe and 46% in Malawi). The spousal search model is easily adjustable to accommodate these, the first through offering multiple search draws and the second through relabeling "marriage." However, it would require strong modelling assumptions to generate analogous theoretical predictions on single and married risk to those used earlier (for example, are polygynous men constantly searching?). In principle, one could use a hazard rate into the final marriage. However, date of most recent marriage would be needed construct such a hazard rate, and that data is not collected in the DHS surveys. Thus, the empirical focus in this section is on implications of spousal search which are less sensitive to these sorts of assumptions than the age-infection distribution. Secondly, the DHS represent the best available comparable data across countries with HIV tests. However, these data are compromised by 12-25% testing refusal/failure rates, which are surely non-random. In particular, an individual's refusal decision seems likely to be affected by her subjective probability of infection. Results presented here are robust to the alternate extreme assumptions that all test-refusers are positive or that all are negative (results under these assumptions are available from the author), suggesting that test-refusal is not strongly correlated with the trends presented here. However, if test refusers are heterogeneous in their likelihood of infection, and that heterogeneity is correlated with the variables of interest, endogenous refusal remains a problem and all analysis below must be considered in that light.

This section elaborates further on two trends summarized in the main paper. First, the paper discusses that HIV infection risk declines with marital tenure, and suggests that the process of becoming married may be dangerous. As a test, probability of joint negative infection is estimated by probit in Table 4 for the high prevalence countries in Southern and Eastern Africa with publicly available DHS data, with fixed effects controlling for 5-year age group and spousal age group. Analysis is restricted to monogamous couples (except in Lesotho where the polygamy question was not asked); throughout, the omitted group is couples in their first 5 years of marriage. In every country, couples who have been married for at least ten years are substantially less likely to be infected than their recently married peers, with precisely estimated coefficients for every country save Lesotho. Ten years of marriage reduces likelihood of infection by 50%-100% of the mean prevalence level, relative to peers who have been married 0-5 years.

A second prediction of the marital shopping model which is summarized in the paper is that length spent single should be correlated with HIV infection. That is, a longer time period spent single indicates less luck in finding a match, and should be correlated with the amount of risk borne. In fact, across all 5 of these countries, being single longer is associated with greater HIV risk, even while two other likely candidates for HIV risk, polygamy and spousal age differences, are not consistently statistically significant (results available from the author; similar results are presented in Bongaarts 2006). However as both the decision to commence sexual activity and the decision to get married are endogenous, we may remain concerned that the length of singlehood is picking up other correlated omitted variables like preferences. I address this issue by observing that infections from pre-marital behavior should disappear with marriage tenure. That is, if singlehood itself is risky, then the effect of years single should be strongest over the recently married, as those who have been married longer are unlikely to have survived pre-marital infections. In contrast, if a long period of singlehood is simply a signal of preferences, then those with long singlehoods should still be at risk years after marriage. Table 5 presents marginal effects from a probit of HIV prevalence on years of singlehood and dummies for each marital tenure category, where the effect of singlehood on HIV is allowed to change with marital tenure. Across countries, years of singlehood are most strongly related to HIV status in the first ten years of marriage; in all countries, the point estimates go down substantially after these years (and for Lesotho and Zimbabwe, we can statistically distinguish the effects of years 0-5 from years 15+ at least the 5% level). Ten years is a logical turning point, as it is roughly the median life expectancy after infection. These results suggest that being single longer is risky in large part due to behaviors which take place when single rather than correlated behaviors which last a lifetime, consistent with the marital shopping model.

4 Gonorrhea

An external additional test of the model would be to input transmission dynamics of a different sexually transmitted disease in a different context and evaluate the age profile. Unfortunately, transmission probabilities of most sexually transmitted diseases are little understood and cases are often undocumented. Gonorrhea in the US, however, provides a good case study (though extremely low prevalences may cause some concern over selectivity). Unlike HIV, gonorrhea is extremely infectious, with transmission probabilities very high for a single contact and approaching 1 for a month-long relationship. Gonorrhea is also a transient infection, with most people spontaneously recovering without treatment in a few months, or experiencing quicker recovery with an anti-biotic. The marital search model, then, would predict a constant incidence for single, sexually active adults and zero incidence for married adults. Figures 1 and 2 illustrate the predicted versus observed gonorrhea prevalences by age in the US. For both women and men I overpredict prevalences at older ages. Nonetheless, the predictions do exhibit a similar pattern to the data despite the very different biological and geographical context.

5 HIV transmission rates

HIV studies have measured transmission rates in a variety of ways. The most recent studies, all African, choose to report per-partner-year transmission rates, sometimes accompanied with per-coital-act rates. While all measurements are highly variable, PPY rates seem to be the most stable. Table 6 lists the methods, sample sizes, and estimated transmission rates from a survey of these studies.

References

Central Bureau of Statistics (CBS) [Kenya], Ministry of Health (MOH) [Kenya], and ORC Macro. Kenya demographic and health survey 2003. Calverton, Maryland: CBS, MOH, and ORC Macro, 2004.

Central Statistical Office (CSO) [Zimbabwe] and Macro International Inc. Zimbabwe demographic and health survey 2005-06. Calverton, Maryland: CSO and Macro International Inc., 2007.

Angela M. Downs and Isabelle De Vincenzi. Probability of heterosexual transmission of HIV: Relationship to the number of unprotected sexual contacts. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11(4):388–395, 1996.

David Lam, Jeremy Seekings, and Meredith Sparks. *The Cape Area Panel Study: Overview and Technical Documentation for Waves 1-2-3.* The University of Cape Town, December 2006.

Y. Laurian, J. Peynet, and F. Verroust. HIV infection in sexual partners of HIV-seropositive patients with hemophilia. *New England Journal of Medicine*, 320(3):183, 1989.

Ministry of Health and Social Welfare (MOHSW) [Lesotho], Bureau of Statistics (BOS) [Lesotho], and ORC Macro. Lesotho demographic and health survey 2004. Calverton, Maryland: MOH, BOS, and ORC Macro, 2005.

National Center for Health Statistics. National Survey of Family Growth, Cycle 6. http://www.cdc.gov/nchs/nsfg.htm, 2002.

National Statistical Office (NSO) [Malawi], and ORC Macro. Malawi demographic and health survey 2004. Calverton, Maryland: NSO and ORC Macro, 2005.

Nancy S. Padian, Stephen C. Shiboski, and Nicholas P. Jewell. Female-to-male transmission of human immunodeficiency virus. *Journal of the American Medical Association*, 266(12):1664–67, 1991.

Thomas A. Peterman, Rand L. Stoneburner, James R. Allen, Harold W. Jaffe, and James W. Curran. Risk of human immunodeficiency virus transmission from heterosexual adults with transfusion-associated infections. *Journal of the American Medical Association*, 259(1):55–58, 1988.

Margaret V. Ragni, Lawrence A. Kingsley, Pamela Nimorwicz, Phalguni Gupta, and Charles R. Rinaldo. HIV heterosexual transmission in hemophilia couples: Lack of relation to t4 numebr, clinical diagnosis, or duration of HIV exposure. *Journal of Acquired Immune Deficiency Syndromes*, 2(6):557–563, 1989.

Alberto Saracco, Massimo Musicco, Alfredo Nicolosi, Gioacchino Angarano, Claudio Arici, Giovanna Gavazzeni, Paolo Costigiola, Sergio Gafa, Cristina Gervasoni, Roberto Luzzati, Felice Piccinino, Francesco Puppo, Bernardino Salassa, Alessando Sinicco, Roberto Stellini, Umberto Tirelli, Giuseppe Turbessi, Gian Marco Vigevani, Giuseppe Visco, Roberto Zerboni, and Adriano Lazzarin. Man-to-woman sexual transmission of HIV: Longitudinal study of 343 steady partners of infected men. *Journal of Acquired Immune Deficiency Syndromes*, 6(5):497–502, 1993.

South Africa Department of Health and ORC Macro. South African Demographic and Health Survey 1998. www.measuredhs.com, 2001.

Statistics South Africa (Producer). South Africa Labour Force Survey, September 2002, September 2003, September 2004. Pretoria, South Africa: South African Data Archive (distributor). www.nrf.ac.za/sada, 2002-2004.

Tanzania Government and ORC Macro. Tanzania HIV/AIDS indicator survey 2003-04. Dar es Salaam: Tanzania Commission for AIDS and National Bureau of Statistics, 2005.

Isabelle De Vincenzi. A longitudinal study of human immunodeficiency virus by heterosexual partners. *New England Journal of Medicine*, 331(6):341–46, 1994.

Table 1: Means from DHS data

	(1)	(2)	(3)	(4)	(5)	(6)
Individual Sum Stats	Kenya	Lesotho	Malawi	Tanzania	Zimbabwe	South Africa
HIV	0.073	0.259	0.106	0.046	0.164	
	(0.008)	(0.012)	(0.009)	(0.004)	(0.007)	
Age	30.202	30.801	27.658	29.232	29.294	34.270
	(0.224)	(0.239)	(0.236)	(0.151)	(0.153)	(0.144)
Age First Sex	17.074	16.652	16.392	17.183	17.930	16.908
	(0.085)	(0.069)	(0.068)	(0.053)	(0.051)	(0.042)
Age 1st Marriage	19.155	18.358	17.338	18.371	18.705	21.287
	(0.104)	(0.080)	(0.077)	(0.066)	(0.060)	(0.970)
Spousal Age Gap	6.379	5.986	5.508	6.826	6.367	6.153
	(0.129)	(0.135)	(0.116)	(0.113)	(0.094)	(0.103)
N	1415	1425	1256	2824	3045	2915
Polygynous (in cluster)	0.162		0.160	0.092	0.107	0.048
	(0.008)		(0.006)	(0.005)	(0.006)	(0.004)
Cluster N	394	402	499	345	398	935
Couple Sum Stats						
Both Negative	0.901	0.665	0.860	0.923	0.764	
	(0.010)	(0.020)	(0.011)	(0.006)	(0.011)	
Years Married	10.217	11.857	9.680	8.996	10.227	
	(0.271)	(0.400)	(0.259)	(0.157)	(0.214)	
Couples N	897	558	928	1780	1427	
Year	2003	2004	2004	2003	2005/06	1998

Table 2: Parameter Inputs

$\theta \in \{0, 1, 2, ..., 50\}$

$F(\theta) = N(25.5, 25/4)$

(Acute Transmission): $\rho_1^m = \rho_1^f = .2$

(Late Transmission): $\rho_2^m = \rho_2^f = .02$

(Mature Transmission): $\rho_3^m = \rho_3^f = .073$

$$\Pr(\theta'|\theta) = \left\{ \begin{array}{c} .3|\theta' = \theta \\ .2|\theta' = \theta \pm 1 \\ .15|\theta' = \theta \pm 2 \end{array} \right\}$$

$T = 480$ months

25 Cohorts, with 100 men and women in each

Search time per individual = 240

Transmission parameters are taken from Gray et al(2000), Pilcher et al (2004), and Wawer et al (2005)

Table 3: Simulated Data Summary Statistics

variable	median	mean	sd	min	max
lifetime partners	11	13.30	9.82	1	65
# partners, last 12 mos.	1	1.57	1.64	1	12
total years single	8	8.63	6.85	0.08	19.92

Table 4: HIV Negative Couples by Marriage Tenure

	(1)	(2)	(3)	(4)	(5)
Yrs Married	Kenya	Lesotho	Malawi	Tanzania	Zimbabwe
5-9	0.032	0.019	0.056*	0.047***	0.110***
	(0.02)	(0.07)	(0.03)	(0.02)	(0.03)
10-14	0.076***	0.121	0.143***	0.052***	0.136***
	(0.02)	(0.08)	(0.02)	(0.02)	(0.03)
15+	0.100***	0.192*	0.121***	0.086***	0.235***
	(0.03)	(0.10)	(0.04)	(0.02)	(0.04)
N	897	558	928	1780	1427
Pseudo R-Sq	0.047	0.045	0.055	0.038	0.065
DepVar Mean	0.901	0.665	0.86	0.922	0.764
Year	2003	2004	2004	2003	2005/06

Presents marginal effects from probits of a couple testing jointly negative for HIV on several categories of marriage tenure. Fixed effects for each 5-year age group and spousal age group are also included.

Table 5: Years Single on HIV by Marital Tenure

	(1)	(2)	(3)	(4)	(5)
Years Single	Kenya	Lesotho	Malawi	Tanzania	Zimbabwe
Years Single*	0.006**	0.035***	0.009*	0.007***	0.028***
Married 0-4 yrs	(0.002)	(0.009)	(0.005)	(0.002)	(0.005)
Years Single *	0.009***	0.012	0.017**	0.007***	0.013***
Married 5-9 yrs	(0.003)	(0.014)	(0.007)	(0.002)	(0.005)
Years Single *	0.004	0.018	0.006	0.000	0.012*
Married 10-14 yrs	(0.006)	(0.012)	(0.009)	(0.004)	(0.006)
Years Single *	0.004	-0.017	0.004	0.004	-0.002
Married 15+ yrs	(0.004)	(0.020)	(0.011)	(0.004)	(0.008)
Observations	1422	1425	1256	2824	3045
Pseudo R2	0.046	0.019	0.017	0.023	0.027
Mean HIV	0.068	0.259	0.106	0.046	0.164

Presents marginal effects from a probit of HIV status on marital tenure for women. Dummies for each marital tenure group are also included as covariates

Table 6: Review of Medical Studies

Study	Location	Year	Report	N (M/F)	Per Partner	PPY	PPY-at risk	Per Coital Act
Gray et al	Uganda	2001	PCA	174	21.84%	11.60%	11.60%	0.0011
M/F Comparison				97/77	17.5%/27.3%	9.1%/14.8%	9.1%/14.8%	0.0009/0.0013
Fideli et al	Zambia	2001	PPY	1022	1022	15.8%	7.70%	
M/F Comparison				535/487	12.33%/8.8%	8.28%/7.07%		
Quinn et al	Uganda	2000	PPY	415	21.70%	11.80%	11.80%	
M/F Comparison				228/187	21.4%/21.9%	12.0%/11.6%	12.0%/11.6%	
Pedraza et al	Spain	1999	PP	38	26.30%	5.76%	5.76%	0.0005
M/F Comparison				27/11	25.9%/27.3%			
Downs and De Vincenzi	Europe	1996	PP	121	9.90%	5.00%	5.00%	0.0004
M/F Comparison				73/48	11.0%/8.3%			0.0005/0.0003
Rockstroh et al	USA	1995	PP	198 (M)	0.100			
De Vincenzi	Europe	1994	PPY	245	4.70%	2.30%	4.80%	0.001
M/F Comparison				157/89	5.09%/4.6%			
Saracco et al	Europe	1993	PPY	343 (M)	0.055	3.90%	7.20%	
Padian et al	USA	1991	PP	379	16.40%			
M/F Comparison				307/72	19.9%/1.4%			
Ragni et al	USA	1989	PP	45 (M)	0.130			
Laurian et al	Europe	1989	PP	31	0.097	0.0480	0.0880	
Peterman et al	USA	1988	PP	80	15%			0.0099
M/F Comparison				55/25	18.2%/8%			0.0012/0.0005

Report indicates the chosen method of reporting, where PCA = per coital act, PP = partnership, PPY = per partnership-year. At-risk partnerships are defined to be those who do not report perfect condom use.

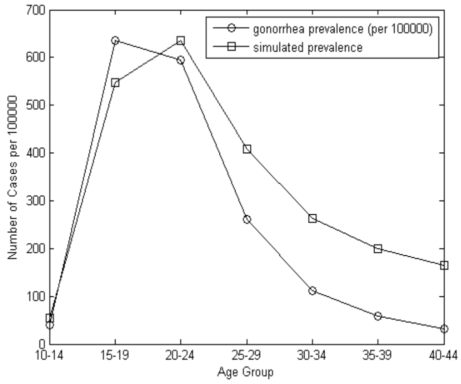

Figure 1: US Female Gonorrhea prevalence by Age

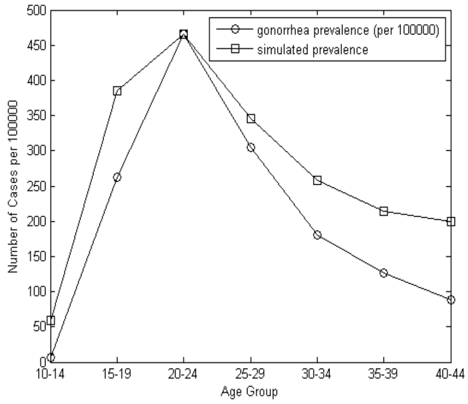

Figure 2: US Male Gonorrhea by Age