LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR 90 1710

Received

JUN 11 1990

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

LA-UR--90-1710

DE90 011968

TITLE THE IWG MODEL FOR THE HETEROSEXUAL SPREAD OF HIV
AND THE DEMOGRAPHIC IMPACT OF THE AIDS EPIDEMIC

AUTHOR(S) E. A. Stanley, S. T. Scitz, P. O. Way, P. D. Johnson,
and T. F. Curry

SUBMITTED TO UN/WHO Workshop on Modelling the Demographic Impact of the
AIDS Epidemic in Pattern II Countries

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

MASTER

THE IWG MODEL FOR THE HETEROSEXUAL SPREAD OF HIV AND THE DEMOGRAPHIC IMPACT OF THE AIDS EPIDEMIC*

E. A. Stanley¹, S. T. Seitz², P. O. Way³, P. D. Johnson³, T. F. Curry⁴

Prepared for the UN/WHO workshop on Modelling the Demographic Impact of the
AIDS Epidemic in Pattern II Countries - Progress to Date and Policies for the Future

¹*Los Alamos National Laboratory, Los Alamos, NM;* ²*University of Illinois, Champaign, IL;* ³*US Bureau of the
Census, Washington, DC;* ⁴*US Air Force Academy, Colorado Springs, CO*

*address requests for further information to S.T. Seitz, Merriam Laboratory, University of Illinois, 512 E.
Chalmers St., Champaign, IL 61820 or E. A. Stanley, MS B284, Los Alamos National Lab, Los Alamos NM 87545*

I. Introduction

This paper presents the State Department's Interagency Working Group (IWG) model for the spread of HIV. The model is fully operational for Pattern II (heterosexual & blood transmission) countries and soon will be extended to Pattern I (primarily homosexual & IV drug transmission) and Pattern III (heterosexual, homosexual, & IV drug transmission) countries. This model was developed for various uses, including technical research, policy analysis, and support for decision making. Research uses include studying patterns of HIV spread, assessing the relative effect of different processes on the spread of HIV, examining the demographic impact of HIV infections, and comparing the potential impact of behavioral versus medical intervention strategies. The model will be used in workshops where policy makers and health officials can do hands-on scenario analyses, gain qualitative insights into the possible long-term epidemiological and demographic impact of HIV, gauge the uncertainties in predictions for the future, and study the impact that intervention strategies are likely to have. These workshops will highlight the magnitude of the problems that local decision makers are facing and help them focus on the need to collect data about a few specific parameters that most affect this epidemic. The computational model uses a deterministic system of differential equations and runs on on a 286- or a 386-based IBM-compatible machine under Microsoft Windows. The program requires an input ASCII (text) file to run; all parameters used by the model are input through this file and, therefore, are user-accessible. The software is user-friendly, mouse-driven, and allows for interactive manipulation of input data and visualization and processing of model outputs.

* This is an interagency research effort under the auspices of the U.S. Department of State Interagency Working Group on AIDS Models and Methods.

II. Highlight of Model Outcomes

II.1. Model Objectives

We have developed a model of the spread of HIV by heterosexual sex [both short-term (casual) and long-term (married) sexual partners], blood transfusions, and perinatal transmission. This deterministic model takes assumptions about the social and biological factors involved in the spread of the epidemic and converts them into qualitative rules which specify their logical consequences. It accounts in a realistic manner for all factors which are known to affect the Pattern II epidemic in any major way, as well as many factors which are somewhat more speculative. Since these rules can be manipulated by turning them on and off or by changing parameters through simulated interventions, the model is a flexible tool that, unlike other methodologies, can be used to answer a great many questions about the influence of and interaction among these factors.

Population studies provide us with some insight into the prevalence and incidence of HIV infection, and of the behaviors which spread HIV, but they leave many unanswered questions about the epidemiology of the disease and its eventual demographic impact. Statistical extrapolation, which can use historical data to estimate the number of AIDS cases or HIV infections for a few years into the future, cannot incorporate the effects of changing behaviors or the saturation of subpopulations by HIV infection because it reduces the dynamic processes that drive the epidemic to a static snapshot. Because the IWG model relies on processes to create its results, it is immediately responsive to changes in social behavior, medical factors, and the saturation of population groups. However, we are building a statistical extrapolation package into the code which will allow model results and extrapolations to be compared.

The IWG model is designed to provide qualitative insights into specific demographic and epidemiological questions about the epidemic. Demographic questions include the age pattern of infections and the effect of the epidemic on the age-structure of the population (e.g. dependency ratios), on infant and child mortality, and on overall population growth rates. Epidemiological questions include the relative importance of each infection route, how the virus spreads between age groups, the role that very high risk takers play in spreading the epidemic, how the epidemic spreads from urban to rural regions (why does it seem to spread rapidly under some conditions and not spread at all under other conditions), the effect of other STDs, and the impact of interventions on the epidemic. Intervention studies might include the effectiveness of promoting condom use, especially given that this affects other STDs; comparison of condom use strategies with strategies to make blood screening more effective; the treatment of other STDs; and the impact on HIV (and AIDS) incidence and prevalence of treatment strategies like AZT.

A demographic model provides the framework for the AIDS-related processes. This model distributes the population by age, includes a fertility component that calculates birth rates, and uses life tables to calculate death and widowhood rates. The population is divided into two regions (urban and rural), so that some aspects of interregional spread can be studied. We then include the major population divisions and processes important in HIV spread.

The questions which we are interested in understanding are questions about the behavior of the population aggregate. Thus, we have not used Monte Carlo simulations, which require the specification of rules for the behavior of individuals: we know much less about the rules which build aggregate behaviors than we do about the aggregate rules themselves. To study questions about how individual life histories create the sexual

distributions seen in cross-sectional population studies is a major research project in itself. There is also the difficulty that, given the large amount of population heterogeneity involved in the spread of HIV, random fluctuations will be large and it would thus require a large number of simulations for each set of parameters in order to obtain population means and variances. Not only would this be confusing to the nonmathematician users of our code, who would want to rely on results from single runs, but studies that varied parameters would then become prohibitively time-consuming and increasingly indefinite.

II.2. Population Divisions

Our model contains two types of population divisions. The first set, status groups, is shown in Fig. 1. These status groups are in some sense the definable, or measurable, characteristics of the individuals in the population: their age, sex, home region, marital status (single, paired to uninfected or paired to infected), and infection class (uninfected, infected, or AIDS). Age is a continuous variable and the rest are discrete variables.

We distribute the population continuously in age for several reasons. AIDS may have a large impact on population structures in some developing nations. Fertility rates, sexual contacts, blood transfusions, and many other processes are strongly age-structured, and it has been convincingly demonstrated that these age-structures have an impact on the spread of a disease. One of the questions we would like to investigate is the distribution of infection in the male and female population that arises in light of our epidemiological assumptions, and the manner in which this changes over time. The continuous age structure, while being more realistic than the bin structure commonly used in demographic models, provides us with a very stable infrastructure, with smooth transitions between groups, and allows us to have very flexible age-structured pairing.

In order to capture the major impacts of internal migration and sexual contacts between the regions, we have chosen to treat the population as living in two distinct regions of the country. Behavior and health may be very different between the two regions, which may represent urban and rural, two large cities, or any other major geographical divisions. For example, blood transfusions may depend on access to medical care, and people with large numbers of casual sex partners may be more common in the relative anonymity of urban areas. This geographic division could be generalized to class or ethnic/tribal differences, if the appropriate "two group" data were to become available.

In separating the population by pairing status, we have chosen to follow Dietz (1988) and Bongaarts (1989). This allows us to study the degree to which the virus may be spread in and dampened by long-term relationships (with repeated sexual contacts), the role that divorce and remarriage may play in spreading the virus, and the impact that the age of marriage may have on virus spread. These factors are particularly important in countries where HIV is infecting monogamous women of child-bearing age. Pairing status is reflected in sexual behavior, with those involved in long-term relations generally having lower rates of casual sex partners than single people of the same age. Note that, unlike Dietz, we have followed individuals rather than pairs: the pair structure becomes too complex for a comprehensive model once we start accounting for polygamy.

The divisions in Fig. 1 are not sufficient to reflect the sexual behavior and biology that is important to the spread of HIV. Within the same regional, age, and sexual preference groups, there is a continuum of partner-acquisition rates, from those who have a single partner in their lifetime to those who average many partners per day. The more partners that a person has, the more likely that one of them will be infected. The work of a number of

authors [Anderson, et al (1986), Hyman and Stanley (1988), Jacquez, et al (1988)] shows that this distribution is crucial in the spread of HIV and that two or three "risk activity" levels are not sufficient to account for its effects.

Population Status Groups

age	sex	infection class	marital status	region
0	Female	Uninfected	Single	Rural
:		Infected	Paired to Uninfected	
apub	Male		AIDS	Paired to Infected
:				
amax				

Figure 1

The population is divided into status groups, which are their definable characteristics

There is not a constant rate of developing AIDS during the course of infection; instead the rate of developing AIDS depends on the duration of infection. There also appears to be a wide diversity in the probability of transmitting infection with a single contact: some people appear to shed much more virus than others, while at the same time, it is possible that some people are much less resistant to infection than others. We do not know all the factors that come into play, but there is growing evidence that the presence of other STDs increases transmission rates, probably by both increasing susceptibility to infection and by increasing viral shedding rates. However, even when all other factors appear to be equal, there may be large differences in infectiousness. As intervention programs promoting condom use begin to have an effect on behavior, some people will use them for all contacts, but many will use them with only a fraction of their contacts.

Our model accounts for these considerations by using distribution functions within each status group (Figure 2). Each status group is distributed over "risk," which is proportional to partner-acquisition rates. Infected status groups are distributed over duration of infection. Each status group is distributed over the fraction of times condoms are used during casual sex. The populations in each status group are also divided into those with and without other STDs and those who are and are not circular migrants. These distributions are dynamic, and we solve equations that determine how the distributions change over time and how they vary between status groups. To keep the calculations in the numerical code within reason, they are not all cross-linked with each other. For example, the duration of infection distribution is assumed to be independent of all the other distributions. However, the probability of having another STD is clearly not independent of risk and we thus take it to depend on risk. Likewise, the distribution over transmissibility depends on the distribution over STD presence and the use of condoms.

Status Group Characteristics

partner acquisition rate	infection duration	circular migrants	condom use frequency	STDs	transmis- sibility
0	0	Migrant	0	noSTD	0
:	:		:		:
:	:		:		:
fmax	τ max	nonMigrant	1	STD	1

Figure 2

Each status group is further distributed and divided according to a number of other characteristics. Partner acquisition rate, infection duration, condom use frequency, and the fraction who are circular migrants are independent of each other. The presence of an STD and the transmissibility depend on the partner acquisition rate, the use of condoms, and the presence of an STD

II.3. Model Usage, Outputs, Etc.

The model runs from a user-provided ASCII input file ("baseline file") that is processed to transform the data into the functional forms used in the code ("equations of state" file). This processed file can be examined by the scientific user, who can modify the file interactively, both graphically (with the mouse) and with a text editor. Input to the model consists of the age, sex, and risk group specific population data as well as demographic and epidemiological parameters for each modeling application (see Appendix 3 for the input files used in section IV).

The code has a default projection time of 10 years, but the user can interactively specify much longer times. The calculation time varies with the number of years being projected, the size of the time step, the order of the numerical solver being used (the user can choose among five solvers, three with automatic time steps and two with user-defined time steps), and the number of time-step solutions being saved to disk. (The program defaults to storing a full accounting trace for the time step nearest each year, but the user may wish to store all of these traces to improve graphics accuracy.) A simple 1st-order Euler run with ten 1-year time steps takes between 2 and 3 minutes on a 386-based machine. This simple solution provides reasonable qualitative projections because the long duration of infection and the small infectivity per contact imply that processes happen slowly. Exploring the effects of intervention programs may take calculations of 20 - 25 years or more and may require the use of one of the user-selectable higher-order solvers. Some scientific questions, such as the study of the timing and magnitude of the epidemic peak and the systematic study of social system transformations, may take calculations of 100 years or more.

Even for 10-year runs, the model is differentially sensitive to input parameters. It is especially sensitive to the probability of transmission per sexual contact and the mean casual sex rates in the high-risk population. A number of implications can be drawn from this: for

example, different countries may have different seroprevalences not only because of the date of introduction of the virus, but also because their social and health conditions and behavioral practices differ. This sensitivity may also imply that intervention programs may vary in their effectiveness, depending upon the social and behavioral context.

The code can produce output in four forms, which allows the user to examine the results from different perspectives. Graphical output provides population tables (integrating over the age distribution) and other relevant outputs, such as fertility rates, percent of deaths due to AIDS, and the proportion of infections due to transfusions. These tables can be obtained in several dimensions, allowing time slices, age groupings, etc. Graphical output allows the user to look at population trends over time or at the age structure at any given time slice. Up to six graphs of the population using different calculations, times, or status groups can be overlaid in a single plot. For trends over time, the code automatically adjusts the axes so that different calculations can be compared. Three-dimensional surface plots, color coded by height, allow the population age structures to be viewed over time. These plots can be interactively rotated and studied. File output contains the same information as the input ("equation of state") file (see the Appendix). This functional equivalence between the "equations of state" file and any slice from the "time history file" makes it possible to graphically view any time slice, to back-generate new input (baseline) files from the last frame in a time history file, or to start a new simulation with the last frame in a time history file.

II.4. Input Files and Data Availability

An input file is used to specify all parameters for each model run. This file contains information about the status of the population at the initial time and the demographic and HIV-related parameters. For country-specific studies, we would like the values in the files to be as accurate as possible. There are large uncertainties in many of the parameters. This does not diminish the value of using the model to gain qualitative insights into the spread of the epidemic, and the relative utility of intervention programs, but it does mean that the ranges for our quantitative predictions are potentially very broad.

Some information does exist which allows each parameter, and its possible range, to be estimated. Population distribution by age and sex, and the demographic parameters of fertility, mortality, and marriage rate are generally available from both censuses and surveys and have been combined to make country-specific estimates (see, for example, United Nations, 1989). Less information than often estimates for these rates can be obtained by combining soft information about population movements and proxy data from similar countries [Castro and Rogers, 1981]. For HIV-related parameters, surrogate values from studies in developed countries may have to be used for the time being, but a number of ongoing studies in developing nations will soon begin to let us estimate how these parameters vary according to circumstances. Cohort studies of couples are already beginning to provide some information on transmissibility with and without STDs [Latif, 1989]. Knowledge, attitude, behavior, and practice (KABP) studies are being conducted in many countries, and will give us data on sexual activity rates [Mande-Lule, 1989]. However, error bounds on HIV parameters and sexual activity are bound to remain large, given that data is uncertain even for the United States.

Exploration of the changes in the behavior of the model with different assumptions about parameters will give us insights into the sensitivity of the model to uncertainties in

input data. Unreasonable model behaviors can indicate when our conceptions about a society are inconsistent with the observed epidemic spread.

III. Synthesis of the Model: Model Processes

III.1. Demography

Fertility and Mortality. Births are calculated based on initial age-specific fertility rates for females and total fertility rates specified for subsequent periods. NonAIDS deaths are calculated from initial and final life tables, input in the form of age and sex-specific death rates (nmx). Levels of life expectancy for intermediate years can be input by the user to control the pace of nonAIDS mortality change.

Migration. We model several different types of interregional migration. Circular migration, where people periodically leave home for weeks at a time in order to work in the other region, is currently treated by assuming that people can have long-term partners in only their home region and behave as if single while away and by assuming that the circular migrant shifts residence several times per year. This construction assumes that circular migration of families (as opposed to individuals) has a negligible impact on the spread of HIV.

Permanent migration is treated by modeling migration of single people, pairs moving as a unit, and paired people leaving their spouse at home. Migration rates depend on age, sex, pairing status, and region and are independent of other population characteristics. Permanent migrants influence the rates of partner acquisition of their new region because a user-specified fraction retain the behavior of their initial region.

Because there is little information on infection prevalence among international migrants, we currently assume that the composition of the population migrating in and out of the country differs only by age and sex from the population in the country. Thus, in our model, external migration can have an impact on population size and on its age and sex structure but it can only indirectly impact infection prevalence.

Long-term Pairs. Pairing status is determined from age-specific marriage, infection, divorce and widowhood rates for the opposite sex using an algorithm that balances pairing between the two sexes. Partners are chosen randomly with regard to all characteristics except age. Pairings are age-biased between the sexes, with the age-structure determined by male preferences. Male preferences are given by a desired age shift and a variance in that age shift, so that their female spouses come from an age range centered about this shift in ages and with roughly the width given by this standard deviation. When there is a shortage of women in the desired age range, women from outside the range are used and pairing rates are somewhat decreased for men in the age groups that would like to pair to these depleted female age ranges. Pair formation does not occur between people with different home regions.

In some studies, polygamy has shown up as a risk factor for HIV infection. Assuming that the details of the age-structure of polygamy are secondary, we have incorporated a simplified version of polygamy. Polygamy is accounted for by assuming that each man marries "x" women and has sexual contacts with "x" women at all times, where "x" is independent of age and is the ratio of the number of paired women to the number of paired men. When there is a shortage or excess of women in the most desirable age groups, there may be a change in this ratio, "x". Divorce affects one woman at a time. Each woman marries only one man at a time. This structure, while it only partially accounts for the fact that wives are successively

accumulated, allows us to examine the effects of polygamy without excessive complexity in the model formulation.

III. 2. Epidemiology

We model the critical epidemiological processes, including perinatal transmission, blood transmission, heterosexual spread by casual sexual contacts, and spread in long-term partnerships, the spread of a generic cocirculating treatable STD, and changes in sexual behavior. Equations for the infection rates are given in Appendix 1.

Infection by blood and mothers. A user-specified fraction of babies born to infected mothers are infected at birth. These infected infants will die at an accelerated rate. There is recent data that indicates an increase in the death rate for the infants of infected mothers which is greater than anticipated from direct infection [Lallemont, et al,1989], and we are considering incorporating this possibility into the model, although it could be accounted for in the present model by increasing the fraction assumed to be infected. The rate of blood transfusion infection is determined by the fraction of blood infected in the region of residence (calculated from the fraction donated by each age and sex group and their infection prevalence plus the fraction of blood that is screened), the rate of transfusion for the group, and the average number of units per transfusion.

Infection by casual sex. One of the major ways that HIV is spreading from person to person is through short-term sexual relations. We treat sexual relationships that last less than a few months as single contacts between individuals (with the rate of transmission per contact the same as for a single encounter in a long-term pairing). The rate of infection by casual sexual contacts depends on the partner-acquisition rate, the probability that the person with whom a contact occurs is infected, and the probability that a single contact with an infected person transfers infection. Casual sex contacts between regions, which may be a major route of urban/rural spread, are a user specified fraction of all contacts.

Our expression for the transmission rate per unit time for a member of a given uninfected status group, detailed in Appendix 1, is the product of the average rate of partner acquisition, the mean susceptibility to infection, and the sum over all possible groups from which the sex partner can come of the probability that the partner comes from that group, and the mean infectiousness per contact of a person from that group. The probability of transmission, described in the transmissibility section, is thus taken to be the product of the susceptibility of the uninfected person and the infectiousness of the infected contact.

We have developed an algorithm that ensures heterosexual contact matching and also allows for arbitrary age-structures of contacts to be specified. Contact rates between groups must be matched mathematically, so that if group A has x contacts with group B per unit time, then group B has x contacts with group A per unit time. We also know that contacts will be age-structured, so that people of a certain age will tend to have partners that are from another, possibly different, age group. Partners are distributed over a range of ages, which plays a crucial role in determining the amount and pattern of spread between age groups. This mixing strategy automatically adjusts to changes in population structures which occur over time.

We do not model social learning as an endogenous process. For example, the program does not assume that those who progress to AIDS will automatically reduce their sexual activity or change their condom use to protect their partners. These modifications can, however, be introduced through user-defined behavioral intervention strategies. The program

supports sets of interventions drawn within or across the behavioral, social, and medical arenas.

Infection by long-term partners. Infection by a long-term partner occurs for uninfected people paired to an infected partner. The rate of transmission is the product of the rate of sex within the pair and the probability that each sexual contact transfers infection. We currently assume that long-term partners do not use condoms for disease prevention. We anticipate generalizing the code to permit changing behaviors among pairs with infection status. When polygamy occurs, men are considered paired to infected if any spouse is infected. Contact rates are matched between the partners (men of age a and women of age a') using a harmonic mean between the age-dependent rates for men and for women, prorated by the number of wives per man. Circularly migrating spouses are assumed to have no sex with their spouses while out of the region.

Transmissibility. As discussed in section II.2, we distribute the population in our model by the amount of condom use. Those using condoms are infected much less frequently, and hence, the model dynamics will create a condom use distribution that differs for the uninfected and the infected populations. We distinguish in our model between susceptibility and infectiousness, taking their product to be the probability that infection transfer occurs in a single sexual contact between an uninfected and an infected person.

The presence of genital ulcers appears to increase both infectivity and susceptibility. This increase in transmissibility may partially explain why HIV has spread very rapidly in some regions of the world with high rates of STDs. By controlling them it may be possible to have a significant impact on the spread of HIV. In order to study this possibility, we determine the fraction of each status group with a "generic" STD, which represents a combination of syphilis, gonorrhea, and chancroid. This STD is spread by the same sexual contacts as HIV. Since the infectivity of these diseases is large and the recovery times are relatively short compared to HIV, we assume that the STD is in equilibrium at all times and that transmission is 100% in long-term relationships. These assumptions give us a steady-state equation for the fraction with an STD in each status group, distributed over partner-acquisition rates and condom use, which varies over time as the conditions in the population change. As condom use increases, or partner-acquisition rates decrease, these diseases will thus automatically become less prevalent. By increasing STD recovery rates over time, we can examine the effect of improved STD recognition and treatment.

There may be other reasons for individual variability in susceptibility and infectiousness - some circumstantial evidence exists for the concept of super-spreaders, people who are extremely infectious at all times during the disease yet who seem to have no distinguishing characteristics. Similarly, uncircumcised males may be more susceptible to HIV than circumcised males (Cameron, et al, 1989). We assume a distribution of susceptibility or infectiousness in the status group that is dependent on the presence of STDs and condom use, and on the duration of infection. This generic treatment of transmissibility is fairly flexible, and it is easily modified as new information becomes available.

Partner Acquisition Rates. The distribution of partner acquisition rates ("risk") varies with the characteristics of the status group: in particular the infected populations will tend to have higher risk than the uninfected populations since infection rates increase with risk, and with increasing age the risk distribution in the population shifts. As people mature, they marry and may move to lower risk and when they divorce they take on the behavior of their new status group. Education about AIDS may cause a decrease in the mean risk in a status group over

time. This decrease can be implemented using the intervention strategy feature of the program. There is a feedback loop in the model which causes decreasing risk behaviors as the size of the epidemic gets large and those with many partners are infected and die.

Social Learning. The user can impose behavioral changes on top of those that will naturally occur in two ways. First, the user can introduce expected changes in fertility and mortality rates in the input file. Second, the user can introduce behavior changes through simulated policy interventions for such things as casual sex rates, condom use, net urban migration ratios, and more.

III.3. Neglected Considerations

Despite the amount of detail in our model, we neglect some aspects of social structure in Pattern II countries that may be important to HIV spread. These include ethnic, economic and social barriers between population groups, and spatial spreading other than between urban and rural regions. We do not separate the never-married from the divorced and widowed, who may have higher rates of remarriage in many populations, and our treatment of polygamy is simplified. Developing countries in Africa and elsewhere are experiencing social upheavals caused by nationalism and urbanization, economic change, ethnic rivalries, modern communication and transportation, and natural disasters. The situation is aggravated when the medical and technical infrastructure lags behind the rate of social change. We do not currently account for transmission by homosexual contacts, medical injection independent of blood transfusions, or IV needle sharing, all of which are thought to have a very minor impact on HIV spread in Pattern II countries.

III.4. Model Evaluation

There are a number of procedures for evaluating the model and its usefulness for various purposes. These include investigating its sensitivity to changes in parameters (see Appendix 2), verifying that the numerical solutions are consistent with the model formulation, validating that the model results are consistent with our current data on the epidemic, and investigating and understanding the behavior of the model under a wide variety of scenarios. We are systematically applying these procedures, to develop our understanding of the epidemic and the insights that can be gained from using it. Our analysis indicates that the model is indeed sensitive to changes in a few key parameters. Given the large uncertainties in a few of the key parameters, we cannot expect the bounds on point estimates of future AIDS cases and infections to be tight, but many other behaviors and results from the model may prove to be more robust. For example, the time that it takes the epidemic to affect population sizes and growth rates is relatively large (>10 years), even under dramatic scenarios such as the worst case described in the next section. Some intervention strategies may be ineffective under all reasonable circumstances.

IV. Results

We present results in this section from our model, based on the standardized data provided by the UN/WHO for this meeting. There are five processes (permanent migration, circular migration, divorce, interregional sexual contacts, and cocirculating STDs) included in our model for which no data was provided for the comparative studies. Besides presenting results where these processes are turned off, we use this section to investigate the impact that they have on the epidemic.

Best-case and worst-case results below come from assuming that the contact rates for males and the probability of transmission from a sexual contact simultaneously take on their upper and lower bounds. Unfortunately, these two bounds are so extreme that in all runs, with or without the processes, the best case parameters imply that the epidemic is not sustained and the worst case parameters imply that the adult population quickly becomes nearly 100% infected above a fairly young age: the population does not entirely die out only because there are women of fertility age who can have uninfected babies. We thus also investigate an intermediate case. For this intermediate case, we averaged the sexual contact rates between the best and worst cases and took the probability of transmission per sexual contact to be 0.03 for infected males and 0.01 for infected females and the amount of condom use to be the average of the two cases (11%).

IV. 1. Comments on input data

The information specified for the standardized set of runs for this meeting does not allow us to fully utilize our input file. We therefore have made some assumptions in order to take fuller advantage of our model. The resulting input files are given in Appendix 3. Our assumptions are detailed below.

Initial infection prevalence and epidemic age. We have the capability to specify the initial infection prevalences by sex, risk, and marital status, but we take prevalence to be independent of these characteristics for the simulations below. In order to estimate the initial distribution of the population by duration of infection, we chose 10 years as the age of the epidemic.

We take the urban and rural populations to have identical initial conditions and parameters except for population size and infection prevalences: as requested 25% of the population is urban and the initial prevalence in the urban population is 11%. For the rural population the initial seroprevalence is only 0.9% (giving a total population infection prevalence of 3.4%).

Age distribution of casual sex. The data provided for the fractions of the population with various types of "promiscuous behavior" are independent of age. For married people, we assumed that the percent high risk is independent of age, but for single people we used a distribution that starts at zero for age 10. To obtain this distribution, we assumed that the proportion of single women who are high risk has the same shape as the proportion married, with a maximum of 80%, and with the minimum age 10 instead of the age 12 specified for marriage. For single men, the proportion high risk is skewed more to younger ages, with a maximum of 89% reached at 27, after which it decreases down to 80%.

Casual sex contact rates. We do not use the simple distinctions of "prostitute", "male who has sex with prostitute", etc in our model. Instead, we transform these numbers into mean

frequencies for casual sex among different marital status groups in the male and female populations. To keep this simple, we assume for these calculations that mean rates are age-independent. For single males we use the given rates for males with prostitutes and for married males we take a weighted sum of 4/7 of the rates with promiscuous women and 3/7 the rates with prostitutes.

We do not endogenously calculate the female contact rates, but instead specify in the input file both male and female contact rates by age for both sex within marriage and casual (high risk) sex. These are adjusted within the code simultaneously for both sexes to ensure consistency as the population changes. To set up female contact rates for our input file from the specified male contact rates, we first assume that we have female prostitutes and the given male populations. We next assume that married female behavior is independent of age, but that the fraction of high-risk women is age-dependent, with the 35 year old single population being 100% prostitutes. We then solve an equation for the prostitution rates (min 160/year, max) and the promiscuous rates (min 2/year, max). This then gives the mean casual sex rates given in the input file for the best and worst case scenarios.

Demographic Parameters. The mortality and fertility rates change continuously in time in our code. We thus take the first set of mortality and fertility rates to be valid at the midpoint of the 1985-1990 time interval, and the last set to apply to 2007.5. Then we linearly extrapolate the life expectancy at age and the total fertility rates between the two points to obtain our 1985.0 and 2010.0 life tables and fertility rates.

Pair Formation. Partner ages are distributed across a range of values in our model. We take the standard deviation in partner age (mean ages are females 5 years younger than males) to be 3.3 years.

Divorce Divorce is added as a process by taking divorce rates to start at 7% per year for men age 15 and decrease slowly to 3% per year. This process was interactively turned off within the code, or set to 0 in the input file to study the effect of divorce and remarriage on the epidemic. Note that divorce rates, initial marriage distributions, and widowing rates are combined in the initial file processing to estimate age-structured marriage rates. Since 95% of the older population remains married at all times, marriage rates for the small fraction of unmarried people at these older ages are extremely high.

Migration. Both permanent migration and circular migration can be investigated by turning the process on or off for the numerical simulations. We investigate the migration processes simultaneously here. Migration rates are assumed to be twice as high for men as for women and to be small for both the very young and the very old. The circular migrants are assumed to be 10 years younger than permanent migrants, but to be roughly the same fraction of the population. Pair migrants are assumed to be five years older than single migrants.

Cocirculating STDs. We assume that the probability of transmission per contact of our generic STD is 0.3, that men are infected an average of 1.5 months and that women recover after an average of 2 months. The use of condoms has the same effect as for HIV, an 80% decrease in transmission probability. We then compare the case where STDs increase transmission probabilities by a factor of 6 to the case where they do not increase transmission at all.

Interregional Sexual Contacts. In our "full process" simulations, one out of every 100 casual sex contacts is between regions. For the case where divorce, migration, and the cofactor effect of STDs are turned off, we have no interregional sex contacts and the two regions are entirely decoupled.

IV. 2. Results

Best case. In the best case simulation the epidemic was not sustained in the population and gradually dies out. Figure 3a shows the decay in the number infected, the number of AIDS cases and the number of AIDS deaths that occur over time. Figure 3b shows a contour plot of the age structure of the population for the 25 year projection period. The population age structure in this extreme "best" case is modestly affected by the epidemic, particularly in the middle of the projection period. The infected population distribution over age is shown versus time in figure 3c. Given the inability of the epidemic to sustain itself over time, the infected population can be seen to be aging and slowly to being depleted by natural and AIDS mortality. Except for perinatal infection of infants, there is little evidence of infection transfer.

The results drawn from the best case scenario are not surprising, because the default parameters are sufficiently optimistic that they guarantee that the epidemic cannot be sustained in the population. Even with a large body of people initially infected, as with these initial conditions, the number infected decays over time. Stated differently, the disease reproductive number (R_0) is well below 1.0, and each infected person is infecting on average much less than one other person, so that the disease cannot reproduce itself epidemically.

The addition of extra processes to the model does not affect these results: the disease transmission parameters and sexual behavior specified by the best-case parameter set imply very little transfer of infection, other than to infants of infected mothers.

Worst case. Results from the worst-case scenario are shown in Figure 4. For the worst-case scenario, the transmission probabilities and sexual activity rates given as the defaults are so extreme that the epidemic spreads explosively, especially among singles. Infection prevalences in single urban females aged 25-30 are close to 100% within a year of the start of the calculation, and several additional age-specific subpopulations approach 100% infection by 1990. Figure 4a shows the projected infection prevalence, number of AIDS cases, and AIDS deaths. Figure 4b shows the dramatic impact of the epidemic on the age structure of the population over time. This rapidly creates a population with oscillations in births and some sustained population depletion. The rate of natural increase steadily declines during the first decade of the simulation and becomes increasingly negative during the second decade. By the end of the first decade, the average age is younger than in 1985, but with AIDS deaths among mid-age adults in the second decade of the simulation, the average age of the population has jumped considerably by the end of the projection period in 2010. Figure 4c shows the infected population distributed over age and time. The graph has been simplified by clipping hidden surfaces. It shows rapid and sustained infection levels in the most sexually active age groups, coupled with considerable perinatal transmission.

The addition of extra processes to the model (divorce, migration, and cocirculating STDs) hasten the slippage in natural rate of increase and introduce even steeper population depletions among mid-age adults. HIV prevalence levels are also generally higher.

Intermediate case. Intermediate-case results are shown in Figure 5. This case is much more interesting and realistic than the best and worst case scenarios. Figure 5a shows the projected infection prevalence, AIDS cases, and AIDS deaths. The graph indicates a slow curvilinear infection spread in the first half of the projection period followed by an explosive growth of the epidemic after 1995. The population mat in Figure 5b flattens at the left boundary, corresponding to a decline in the young child population, like that seen earlier in Figure 3b (best case) but unlike the crumpled population mat in Figure 4b (worst case). In Figure 5c we see an infection wave that has a broader age base at the beginning of the simulation and then ends with a massive concentration in the sexually active adult population by the end of the simulation period, coupled with renewed levels of perinatal transmission.

A comparison of the three scenarios (best, middle, and worst) is shown in Figure 6. The projected populations are illustrated in Figure 6a. The population continues to grow for the best-case and intermediate-case, but there is negative population growth by the end of the projection period for the worst-case scenario. The percent infected in urban areas for each of the three scenarios is shown in Figure 6b. Although the epidemic dies out in the best-case scenario, the intermediate-case shows around 7% of urbanites infected in 2010 and over 50% infected for the worst case. Figure 6c shows the percent infected among the single or unmarried population. The best case scenario again shows extinction, the intermediate case exceeds 3% by 2010 and the worst case shows about 35% of singles infected by projection's end.

The default intermediate-case does not include such demographic processes as migration, divorce, and interregional sex contacts, and it does not include STDs as an HIV cofactor. These additional processes have dramatic effects on the epidemic when compared to the default intermediate case, as portrayed in Figure 7. We offer this more complete scenario as a second but less optimistic intermediate case. Taken together, the additional processes spike the epidemic, as illustrated in a comparison of the default file's (UNMID) projected prevalence, a second projection including permanent and circular migration, interregional sex contacts, divorce, and STD cofactors (UNMIDDLE), a third projection including all of these extra processes except divorce (UNMIDTD), a fourth projection including all but migration (UNMIDTM), and a fifth projection including all but cofactors (UNMITDS) in Figure 7a.

The graph reveals that the extra processes (UNMIDDLE) change the epidemic's contour from concave to convex growth by spiking the infection prevalence midway through the projection. With migration (UNMIDTM) turned off, the infection concentrates in the urban area, thereby leading to even higher HIV prevalence by the end of the projection period. Conversely, this means that migration and circular migration slightly dilute the disease's transmission, at least in this 25 year simulation period. There is a non-obvious interaction between divorce and cofactors revealed in Figure 7a. No divorce (UNMIDTD) or no cofactor tick (UNMIDTS) alone each result in a decline in HIV prevalence. But the original intermediate file (UNMID) contains neither process but still results in increased HIV prevalence. Why? Our model distinguishes between condom use in marriage and condom use in casual sex, with the former focusing on family planning and the latter focusing on safe sex practices. With no divorce and no safe sex practices in marriage, coupled with some extra-marital sexual contacts, the initial levels of HIV infection eventually produce growing

levels of HIV prevalence. With divorce (and hence greater safe sex prophylaxis) or with a cofactor kick but no divorce (and hence greater transmissibility within marriage that depletes the infected married population), there is a long-term decline in HIV prevalence.

Figures 7b and 7c look at the distribution of mean risk over population age and time. Unlike HIV prevalence, the mean risk indicates the source pools from which HIV incidence arises. The contour plot for the intermediate case without migration, divorce, and cofactors (UNMID) shows a rather stable distribution of mean risk over time for the sexually active adult population. Figure 7c provides mean risk for the intermediate case that includes divorce, cofactors, and migration (UNMIDDLE). Over time, these additional processes show the mean risk concentrating in three segments of the age range, the largest for young sexually active adults and the second and third for age waves among middle age and older parts of the sexually active population. Clearly, the differences in structure of mean risk distribution between Figures 7b and 7c suggest longer term generational differences in the spread of HIV. The more complete population picture (with divorce, migration, and cofactors) leads to a more refined picture of long-term age-based risks for HIV infection.

Summary

We have developed a mathematical model of the spread of the AIDS epidemic in developing nations. This model is a tool to help scientists and policy makers gain qualitative insights into the spread of the epidemic, explore scenarios about the future, and evaluate and compare intervention strategies.

Our deterministic differential equation model has a number of special characteristics that we believe are unique when combined together. These include matching of sexual contacts between age groups for both casual sex and long-term pairs (marriage) which gives an age-structured mixing that is neither random nor same-age only, the use of continuous ages and time since infection, and a split into urban and rural regions. We have tried to build a basic model of the spread of HIV. We have made several explicit assumptions about AIDS-related processes that help us reduce the model's overall complexity.

Our effort has advanced rapidly because of a close collaboration between mathematicians, demographers, computer experts, political scientists, operations researchers, and physicians. This has allowed us to divide our efforts in a way that uses the talents of each investigator and that allows us to rapidly exchange ideas. Ultimately, the pragmatic value of a model is whether those using it understand more and can do better in their tasks than those without the model. We believe that this model, in collaboration with field research, can help the world to proceed more rapidly in its efforts to control this disease. The model can provide a window into the future of the epidemic, bolster arguments for targeted data collection aimed at evaluating important parameters, and help field researchers demonstrate to policy-makers where the epidemic is going and why.

References

- Anderson, R. M., May, R. M., Medley, G. F., and Johnson, A. (1986), "A preliminary study of the transmission dynamics of HIV, the causative agent of AIDS," *IMA Journal of Applied Medicine and Biology*, **3**, 229-263.
- Bongaarts, J. (1989), "A model of the spread of HIV infection and the demographic impact of AIDS", *Statistics in Medicine*, **8**, 103-120
- Cameron, D. W., Simonsen, J.N., d'Costa, L. J. (1989), "Female to male transmission of HIV 1: risk factors for seroconversion in men", *Lancet*, 19 August, 1989, 403-407
- Castro, L. J., and Rogers, A. (1981), "A model migration schedule: a simplified formulation and an alternative parameter estimation method", International Institute for Applied Systems Analysis, working paper 81-63. Lacenburg, Austria
- Dietz, K. (1988), "On the transmission dynamics of HIV", *Mathematical Biosciences*, **90**, 397-414.
- Hyman, J. M., Stanley, E. A. (1988), "Using mathematical models to understand the AIDS epidemic", *Mathematical Biosciences*, **90**, 415-474.
- Jacquez, J. A., Simon, C. P., Koopman, K., ... (1988) "Modeling and analyzing HIV transmission: the effect of contact patterns", *Mathematical Biosciences*, **92**, 119-199
- Konde-Lule, J.K., Berkley, S. F., Downing, R., (1989), "Knowledge, attitudes and practices concerning AIDS in Ugandans", *AIDS*, **3**, 513-518
- Lallemont, M., Lallemont-Le-Coeur, S., Cheynier, D., ..., "Mother-child transmission of HIV-1 and infant survival in Brazzaville, Congo", *AIDS*, **3**, 643-646
- Latif, A.S., Katzen, D.A., Bassett, MT , ..., (1989), "Genital ulcers and transmission of HIV among couples in Zimbabwe", *AIDS*, **3**, 519-524
- United Nations Department of International Economic and Social Affairs. (1989), World Population Prospects 1988, Population Studies No. 106. St/ESA/SER.A/106. New York

Appendix 1 : Infection rates

In this appendix we detail the infection rates in our model. For those who are interested we will provide the full equations as a supplement.

Definitions. Let the independent variables for the status groups shown in figure 1 in the main text be designated as follows

t	time
a	age
s	sex
i	infection class
p	pairing status
r	home region

Let the population with these characteristics be denoted by $P(t, a, s, i, p, r)$ (since a is a continuous variable, the integral from a_1 to a_2 is the number of people between those ages). Also define

$P(t, a, s, i, r)$	:	the sum over all pairing groups of $P(t, a, s, i, p, r)$
$c(t, a, s, i, p, r)$		the fraction of people of age a , sex s , infection class i , pairing status p from the region r that are circular migrants
$c(t, a, s, r)$		the fraction of people of age a , sex s , from the region r that are circular migrants
$d(s, r)$		the average fraction of time a circular migrant spends away from home region r
$P_{res}(t, a, s, i, r)$		the number of people age a , sex s , infection status i , residing in region r

where

$$P_{res}(t, a, s, i, r) = P(t, a, s, i, r) (1 - d(s, r) c(t, a, s, r)) + P(t, a, s, i, R) d(s, R) c(t, a, s, R)$$

with R the opposite region ($r \neq R$).

We have three sources of infection after birth in our model: blood transfusions; sex within long-term heterosexual pairings (marriage); and casual heterosexual sex.

Infection by blood. The rate of blood transfusion infection for any group is determined by the fraction of blood that is infected in the region of residence (calculated from the fraction donated by each age and sex group in the region, the infection prevalence in the group, and the fraction of blood that is screened), the rate of transfusion for the group and the average number of units given per transfusion. The rate of infection for those whose home region is r is the sum of the rate of infection for those who are residing in their home region and that for those residing in the other region, R , weighted by the proportions of the population from r residing in those regions:

$$\lambda_{\text{blood}}(t, a, s, r) = [1 - (1 - i_T Q_{\text{Bld}}(t, r))]^{\text{Units}(t, a, s, r)} B(t, a, s, r) (1 - d(s, r) c(t, a, s, r)) \\ + [1 - (1 - i_T Q_{\text{Bld}}(t, R))]^{\text{Units}(t, a, s, R)} B(t, a, s, R) d(s, r) c(t, a, s, r)$$

where r is the region affected by this term, and R is the opposite region from region r .

i_T	the probability that a unit of infected blood transmits infection;
$Q_{\text{bld}}(t, r)$	the fraction of blood infected in the region;
$\text{Units}(t, a, s, r)$	the average number of units of blood per transfusion;
$B(t, a, s, r)$	the rate of transfusion per person.

The fraction of blood units infected in region r is

$$Q_{\text{Bld}}(t, r) = \frac{(1 - D_s(t, r)) \int_{a_{\min}}^{a_{\max}} D(t, a, s, r) P_{\text{res}}(t, a, s, \text{Inf}, r) da}{\int_{a_{\min}}^{a_{\max}} D(t, a, s, r) P_{\text{res}}(t, a, s, \text{Unif}, r) da + (1 - D_s(t, r)) \int_{a_{\min}}^{a_{\max}} D(t, a, s, r) P_{\text{res}}(t, a, s, \text{Inf}, r) da}$$

Here

$D(t, a, s, r)$	the fraction of blood in the region denoted by the age and sex group;
and	
$D_s(t, r)$	the fraction of blood screened, assumed to be a perfect process.

Infection by long-term partners. Infection by a long-term partner occurs for uninfected people paired to an infected partner. The rate of transmission is the product of the rate of sex within the pair and the probability that each sexual contact transfers infection. We currently assume that long-term partners do not use condoms for disease prevention. When polygamy occurs, men are considered paired to infected if any spouse is infected. Contact rates are matched between the partners (men of age a and women of age a') using a harmonic mean between the age-dependent rates for men and for women, prorated by the number of wives per man. Circularly migrating spouses are assumed to have no sex with their spouses while out of the region.

For *uninfected women paired to infected men*, we sum over the possible pairing states of their male partners, who can have either no infected wives ($P_k = \text{PU}$) or some infected wives ($P_k = \text{PI}$):

$$\lambda_{\text{partner}}(t, a, F, r) = [1 - d(F, r) c(t, a, F, \text{Unif}, \text{PI}, r)] \bar{T}_P(t, a, F, \text{Unif}, \text{PI}, r) \\ \times \sum_k \int_{a'_{\min}}^{a'_{\max}} S_P(t, a, F, r; a') F_P(t, a, F, \text{Unif}, \text{PI}, r; a', P_k) [1 - d(M, r) c(t, a', M, \text{Inf}, P_k, r)] \bar{T}_P(t, a', M, \text{Inf}, P_k, r) da'$$

where we split the standard probability of transmission into the product of the susceptibility and the infectiousness

$\bar{T}_P(t, a, s, \text{Unif}, \text{PI}, r)$	mean susceptibility of the uninfected person to unprotected contacts;
---	---

$\bar{T}_P(t, a, s, \text{Inf}, P_k, r)$ mean infectiousness during unprotected intercourse of the infected spouse;
 $S_P(t, a, s, r; a')$ rate of sex within a pair in region r where the person of sex s has age a and the other person has age a'

and

$F_P(t, a, F, \text{Unif}, P_j, r, a', P_k)$ fraction of the partners (for uninfected women age a , pairing status P_j , from region r) that are age a' and pairing status P_k .

For *uninfected men paired to infected women* we need to account for the fact that some fraction $k_U(t, a, r, a')$ of their age a' partners are not infected

$$\lambda_{\text{Partner}}(t, a, M, r) = [1 - d(M, r) c(t, a, M, \text{Unif}, P_l, r)] \bar{T}_P(t, a, M, \text{Unif}, P_l, r) \\ \times \int_{\text{amin}}^{\text{amax}} S_P(t, a, M, r; a') F_P(t, a, M, \text{Unif}, P_l, r, a') \\ \times [1 - k_U(t, a, r, a')] [1 - d(F, r) c(t, a', F, \text{Inf}, P_U, r)] \bar{T}_P(t, a', F, \text{Inf}, P_U, r) da'$$

Here

$F_P(t, a, F, \text{Unif}, P_j, r, a')$ fraction of the partners (for uninfected men age a , pairing status P_j , from region r) that are age a' .

We do not track partners in the manner used by Dietz (1988). Rather, we recalculate the distribution F_P for spouse ages at each time for the paired populations in a manner that can be age-biased (currently we take this age-biasing to be the same as for casual sexual contacts).

Infection by casual sex. We treat sexual relationships that last less than a few months as single contacts between individuals (with the rate of transmission per contact the same as for a single sex act in a long-term pairing). The rate of infection by casual sexual contacts depends on the partner-acquisition rate, the probability that the person with whom a contact occurs is infected, and the probability that a single contact with an infected person transfers infection. Casual sex contacts between regions, which may be a major route of urban/rural spread, are a user specified fraction of all contacts.

The casual sex infection rate for the uninfected status group of age a , sex s , pairing p is calculated according to home region, r . It is the fraction of those at home times the infection rate for those at home plus the fraction residing away from home times the infection rate for those currently residing in the opposite region, R . Circular migrants are assumed to behave like the single population in their residence region when they are away from home, giving

$$\lambda_{\text{Cas}}(t, a, s, p, r) = [1 - d(s, r) c(t, a, s, \text{Unif}, p, r)] \tilde{\lambda}_{\text{Cas}}(t, a, s, p, r) \\ + d(s, r) c(t, a, s, \text{Unif}, p, r) \tilde{\lambda}_{\text{Cas}}(t, a, s, \text{Sing}, R)$$

For those residing in region r , the infection by casual sex can occur because of contacts with someone residing in region r or someone residing in the other region.

$$\tilde{\lambda}_{Cas}(t, a, s, p, r) = \sum_{r'=r, R} S_{Cas}(t, a, s, \text{Unif}, p, r; r') i_C(t, a, s, p, r, r')$$

where

$S_{Cas}(t, a, s, i, p, r; r')$ mean rate at which people age a , sex s , infection class i , who are currently residing in region r and behaving like they have pairing status p have casual sex with members of the opposite sex who are residing in region r' ;

and

$i_C(t, a, s, p, r, r')$ probability that an uninfected person becomes infected through a single casual sex contact with someone living in region r' .

$S_{Cas}(t, a, s, i, p, r; r')$ is obtained from the population distribution over casual sex frequencies and a balancing algorithm that ensures contact rates are consistent between the two sexes.

The probability of infection from a single casual sex partner residing in region r' is determined by the prevalence of infection in the group that the contact comes from and the transmissibilities of the uninfected and infected partners. A casual sex partner residing in region r' can have either of the two regions as a home region:

$$i_C(t, a, s, p, r; r) = \bar{T}(t, a, s, \text{Unif}, p, r)$$

$$\times \left[\int da' \left[\sum_{p'} \bar{T}(t, a', s', \text{Inf}, p', r) \frac{P(t, a', s', \text{Inf}, p', r)}{P(t, a', s', r)} \omega(t, a', s', r) \rho_{Cas}(t, a, s, r; a', r') \right. \right. \\ \left. \left. + \bar{T}(t, a', s', \text{Inf}, \text{Sing}, r) \frac{c(t, a', s', \text{Inf}, R) P(t, a', s', \text{Inf}, R)}{c(t, a', s', R) P(t, a', s', R)} (1 - \omega(t, a', s', r)) \rho_{Cas}(t, a, s, r; a', r') \right] \right]$$

where

$$\omega(t, a, s, r) = \frac{1 - d(s, r) c(t, a, s, r)}{1 - d(s, r) c(t, a, s, r) + d(s, R) c(t, a, s, R)}$$

is the fraction of people of age a , sex s , living in region r who are at home; $\rho_{Cas}(t, a, s, r; a', r')$ is the density distribution over age a' of the casual sex contacts of people of age a , sex s , residence r that are with people of the opposite sex, residence r' ; and $\bar{T}(t, a, s, i, p, r)$ is the mean transmissibility (susceptibility for uninfected, infectiousness for infected people) for casual sexual contacts of a person residing in region r , age a , sex s , infection stage i , and pairing p . Note that we currently assume that casual sexual contacts are chosen in a manner that is biased by age. Since activity levels are also strongly age-biased, this creates a defacto biasing in risk. Mean transmissibility depends on the amount of condom use and the prevalence of other STDs in the population and the product of the mean susceptibility and the mean infectiousness is the probability of transmission.

Appendix 2: Testing and Validation

Lt Col. Thomas F. Curry, Capt Barbara A. Yost,
Capt John J Tomick, Lt Col Max A Stafford,
LCDR Christopher C Whitehead, Capt Ronnie E King

Department of Mathematical Sciences
United States Air Force Academy

Because the AIDS data from Africa and the rest of the world are uncertain, the question arises as to the value of forecasts that are based on input parameters estimated from such data. The answer depends on the question being asked with the model; the uncertainty in the result is determined by the answer's sensitivity to uncertainties in the input data. It may also depend on uncertainties in the model assumptions and formulation. In this appendix we discuss techniques that can be used to analyze the sensitivity of a model to the input data, and how this sensitivity analysis can be used to help inform decisions based on model results. The second, broader, uncertainty we leave to future studies.

As an example, suppose we want to anticipate the load on the health care system of a country for the next decade. To do this, we would like confidence intervals for AIDS cases over this time period. We can estimate these by repetitively using the model to obtain an average number of AIDS cases and its standard deviation. Each of a large number of code runs would use parameter values that are randomly selected but constrained within their known possible ranges.

Suppose, on the other hand, that we want to determine the sensitivity of the epidemic to small parameter variations. For example, do small differences in health and behavior among Pattern II countries help explain why the epidemic patterns appear to vary so much from one country to the next? By ordering the parameters in terms of the amount of effect they have on the epidemic, we also gain some understanding of the most critical parameters, both in terms of data needs and for slowing the epidemic's spread.

Sensitivity analysis

To obtain a preliminary estimate of the impact of uncertainties in input parameters, we set up triangular distribution functions for most of the values in our input file. These distribution functions had endpoints that were the known possible ranges for the input parameter, and a peak at some "likely" value. We then used the Method of Inverse Transformations to randomly draw 50 input files from this parameter space. The code results from these input files were used to assess the effect of our data uncertainties on the number infected, the cumulative number of AIDS deaths, the total population, and several other quantities. These effects can be expressed in terms of the coefficient of variation (CV = the ratio of the standard deviation between runs to the mean value from the runs). The higher the value of the CV, the more sensitive the result is to uncertainties in the parameters. We found that the size of the epidemic is fairly sensitive (CVs for different measures of the epidemic vary from 0.6 - 0.8), even over a period of only eight years, to our parameter ranges, but that the total population is not (CV = 0.07).

This study was done for an earlier version of our model, and has not yet been repeated for the most current version, but it illustrates the procedure as well as the difficulty involved in trying to make predictions about the future of the epidemic.

Experimental design

To study the relative sensitivity of the results to each parameter in the input file, we investigated the "local response surface" by setting up a Plackett-Burman design to vary the parameters by small amounts around some baseline set of values. Using a Hadamard matrix to set up $n+1$ runs to study n parameters, all parameters can be varied simultaneously, and the effects of each can be ranked (since the parameters each take on their high and low values for half the runs, there is enough information to do t-tests to statistically compare the means of the outputs from the high value and low value runs and even to test for interactions with other parameters).

In our study of a recent version of this model, most inputs were varied about 10% of their range of uncertainty. The total population was most sensitive to the initial female population, but all infected populations were most sensitive to the infectivity of men, which was varied only from 0.001 to 0.005 per contact (while its value has been estimated at 0.0001 to 0.05). Infected populations were also very sensitive to the infectivity of women. Thus differences in health and other factors that change transmission probabilities by only small amounts can have a large impact on the spread of the epidemic. Obtaining better estimates of these parameters are key to narrowing confidence bounds on the model results.

In all, six epidemic response surfaces were examined. The five most important input variables driving each response are listed in the following table:

Frequently appearing variables key:

Risk/cntct M = Risk of infection per sexual contact, for males

Risk/cntct F = Risk of infection per sexual contact, for females

Cntct/mth HR F 15-44 = # of sexual contactsd per month by high risk females
age 15-44

Cntct/mth HR M 15-44 = # of sexual contacts per month by high risk males
age 15-44

STD's female = relative increase of risk of infection due to cocirculating sexually
transmitted diseases, females

Response variable key:

1 = Total number of people infected

2 = Cumulative AIDS deaths

3 = Number of population infected

4 = Percent of the population infected

5 = Percent of females age 15-44 infected

6 = Percent of deaths due to AIDS

Frequently appearing variables	Response Variable Ranking						Avg
	1	2	3	4	5	6	
Risk/cntct M	1	1	1	1	1	1	1.0
Risk/cntct F	2	2	2	2	2	2	2.0
Cntct/mth HR F 15-44	3	3	3	3	3	3	3.0
STD's female	6	4	5	6	4	4	5.0
Cntct/mth HR M 15-44	4	7	7	4	8	6	6.0

References

Hadamard, J., "resolution d'une Question Relative aux Determinants," Bulletin of Scientific Mathematics, Vol. 17, Part 1 (1893).

Plackett, R.L., and Burman, J.P., "The Design of Optimum Multifactorial Experiments," Biometrika, Vol. 33, pp. 305-325 (1946).

Hedayat, A., and Wallis, W.D., "Hadamard Matrices and Their Applications," The Annals of Statistics, Vol. 6, No. 6, pp. 1184-1238 (1978).

Schmidt, S.R., and Launsby, R.G., Understanding Industrial Designed Experiments, CQG Ltd Printing, Longmont, CO (1989).

Figure 3a

--- LEGEND ---

○ Inf, Population, UNBES.THF
 □ Aids, Population, UNBES.THF
 ▼ Aids, Aids Deaths, UNBES.THF

Figure 3b

--- LEGEND ---

X Axis: Time (in years)

X Min: 1985

X Max: 2010

Y Axis: Age (in years)

Y Min: 0

Y Max: 80

Z Axis: Population, UNBES.THF

Z Min: 0

Z Max: 13651

Figure 3c

--- LEGEND ---

X Axis: Time (in years)
X Min: 1985 X Max: 2010
Y Axis: Age (in years)
Y Min: 0 Y Max: 80
Z Axis: Inf. Population, UNBES.THF
Z Min: 0 Z Max: 274.119

Figure 4a

Figure 4b

--- LEGEND ---

X Axis: Time (in years)
X Min: 1985 X Max: 2010
Y Axis: Age (in years)
Y Min: 0 Y Max: 80
Axis: Population, UNWORS.THF
Z Min: 0 Z Max: 14261.1

Figure 4c

--- LEGEND ---

X Axis: Time (in years)
 X Min: 1985 X Max: 2010
 Y Axis: Age (in years)
 Y Min: 0 Y Max: 80
 Z Axis: Inf. Population, UNWORS.THF
 Z Min: 0 Z Max: 4351.21

Figure 5a

Figure 5b

--- LEGEND ---

X Axis: Time (in years)
X Min: 1985 X Max: 2010
Y Axis: Age (in years)
Min: 0 Y Max: 80
Z Axis: Population, UNMID.THF
Z Min: 0 Z Max: 14290.6

Figure 5c

--- LEGEND ---

X Axis: Time (in years)	X Min: 1985	X Max: 2010
Y Axis: Age (in years)	Y Min: 0	Y Max: 80
Z Axis: Inf. Population, UNMID.THF	Z Min: 0	Z Max: 1821.11

Figure 6a

Figure 6b

Figure 6c

--- LEGEND ---

▲ Inf, No Partners, UNBES.THF
 ■ Inf, No Partners, UNMID.THF
 ● Inf, No Partners, UNWORS.THF

Figure 7a

OUTCOMES

	Best case			Intermediate			Worst case		
	1985	2000	2010	1985	2000	2010	1985	2000	2010
HIV prevalence (%)	3.62	3.29	0.767	3.62	2.35	3.50	3.62	37.8	42.4
AIDS prevalence (%)	.299	.0988	.00779	.299	.546	.828	.299	10.1	11.95
Cumulative AIDS cases	938	6119	6570	938	12700	21215	938	128000	213000
Cumulative AIDS deaths	239	6610	7020	239	12100	19800	239	114000	213000
e ₀	50.3	58.9	64.7	50.3	54.7	58.2	50.3	27.5	27.7
IMR	92.3	106.5	73.5	92.3	64.2	53.2	92.3	112	105
q(5)	160	63.6	47.5	160	108	87.8	160	232	223
Growth rate	3.03	2.81	2.97	3.03	2.64	2.61	3.03	-665	-2.47
CBR	49.2	39.7	38.2	49.2	39.5	37.7	49.2	44.5	31.2
CDR	18.9	11.5	8.51	18.9	13.1	11.7	18.9	51.4	56.5
Population size	196	305	392	196	305	378	196	257	245

RUNS WITHOUT DIVORCE, MIGRATION, + COFACTORS

OUTCOMES

	Best case			Intermediate			Worst case		
	1985	2000	2010	1985	2000	2010	1985	2000	2010
HIV prevalence (%)	3.62	3.74	1.00	3.62	6.27	4.34	3.62	46.8	57.5
AIDS prevalence (%)	.299	.111	.0132	.299	1.69	1.19	.299	11.7	16.7
Cumulative AIDS cases	938	5900	6500	938	20300	35300	938	130000	242000
Cumulative AIDS deaths	239	6400	69600	239	17800	34900	239	114000	238000
e ₀	50.3	58.0	65.5	50.3	48.0	56.8	50.3	27.2	26.5
IMR	92.3	65.0	49.4	92.3	75.4	58.8	92.3	114	137
q(5)	160	110	76.3	160	137	100	160	245	290
Growth rate	3.03	2.59	2.59	3.03	1.93	2.19	3.03	-2.47	-4.78
CBR	49.2	37.3	33.8	49.2	36.3	34.1	49.2	32.1	25.8
CDR	18.9	11.4	7.90	18.9	17.2	12.3	18.9	57.2	73.9
Population size	196	292	354	196	283	318	196	236	203