teen-SPARC

STI Prevention and Risk Calculator

Version 1.0

User's Manual

Steven M. Goodreau, PhD[1]; Emily D. Pollock, MA[1]; Eli S. Rosenberg, PhD[2];

Elizabeth M. Rosenthal, MPH[2]; Meredith A. Barranco, BS[2]; Richard L. Dunville, MPH[3];

Lisa C. Barrios, DrPH[3]; Maria V. Aslam, PhD[3]; Li Yan Wang, MBA, MA[3]

[1]Dept. of Anthropology and Center for Studies in Demography and Ecology, Univ. of Washington, Seattle WA. [2]Dept. of Epidemiology and Biostatistics, Univ. at Albany School of Public Health, State Univ. of New York, Rensselaer NY. [3]Division of Adolescent and School Health, Centers for Disease Control and Prevention, Atlanta GA.

This work was funded by the US Centers for Disease Control and Prevention National Center for HIV, Viral Hepatitis, STD, and TB Prevention (Epidemiologic and Economic Modeling Agreement number U38PS004646). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Contents

i. Welcome! 4

ii. Acknowledgments 6

iii. Disclaimers 7

A. Definitions 8

B. Basic model structure and systems thinking 10

C. Accessing YRBS data 14

D. Guided instructions 16

- **D0. General instructions** 16
- **D1. *Welcome* worksheet and *Table of Contents*** 18
- **D2. *Population sizes*** 19
- **D3. *Sexual Behavior*** 24
- **D4: *Diagnoses*** 27
- **D5: *Interventions*** 28
- **D6: *Results*** 29
- **D7: *Export*** 30
- **D8: *Advanced options*: alternative population size calculator** 31
- **D9: *Advanced Options*: New partners per year calculator** 32
- **D10: *Advanced options*: Transmissibility** 40
- **D11: *Advanced options*: infection prevalence among partners** 41
- **D12: Additional worksheets** 42

E: Example Scenario 1: User with jurisdiction-specific YRBS data using SAS 43

F: Example Scenario 2: User without jurisdiction-specific YRBS data 49

G. Default parameter documentation 53

- **G1. *Population* worksheet** 53
- **G2. *Sexual behavior* worksheet** 53
- **G3. *Diagnoses* worksheet** 54
- **G4. *Advanced options*: Alternative population size calculator** 56
- **G5. *Advanced options*: New partners per year calculator** 56
- **G6. *Advanced options*: transmissibility** 56
- ***G7. Advanced options*: Infection prevalence among partners** 57

H. Technical note: Model calibration 60

I. Technical note: Final equations 62
J. Citations 64

i. Welcome!

The teen-SPARC tool is designed to help public health jurisdictions and other entities explore the potential impact of behavior change on reducing sexually transmitted infections (STIs) among the sexually active high-school-attending adolescent population.

The tool focuses on three STIs: gonorrhea (*Neisseria gonorrhea*), chlamydia (*Chlamydia trachomatis*), and HIV (Human immunodeficiency virus I).

It includes three age groups (13-15 year olds, 16-17 year olds, and 18 year olds).

Likewise, the tool focuses on three "sexual partnering groups" (SPGs): males who have sex with males (MSM), males who have sex with females only (MSF), and females who have sex with males (FSM). These are defined in more detail below.

NOTE: use of teen-SPARC requires Microsoft Windows.

The tool is designed to be used with data from the Youth Risk Behavior Surveillance System (YRBSS), and includes instructions specific to building parameters from those data. It also includes default parameters derived from the 2015 national youth risk behavior survey (YRBS) and other sources. The YRBSS includes national, state, territorial, tribal government, and local school-based surveys of representative samples of 9th through 12th grade students. These surveys are conducted every two years, usually during the spring semester. The national survey, conducted by CDC, provides data representative of 9th through 12th grade students in public and private schools in the United States. The state, territorial, tribal government, and local surveys, conducted by departments of health and education, provide data representative of mostly public high school students in each jurisdiction.

For jurisdictions that do not have local YRBS data, the tool is flexible and can use other data sources or explore what-if scenarios. Users can also combine data they have for some parameters with default values for others.

Key outputs include the expected number of new infections and diagnoses in the next year (for base scenarios), and the change in these relative to a base scenario (for intervention scenarios). These output numbers are for adolescents attending high school, due to the nature of the YRBS data underlying many parameters. Thus, they do not reflect all incident infections and diagnoses among adolescents in the jurisdiction.

Some important caveats to keep in mind:

- Due to data limitations, this tool only tracks penile-vaginal sex between males and females, and penile-anal sex between males. It does not consider oral or other forms of sex. Thus, it does not explicitly consider pharyngeal gonorrhea or chlamydia, nor does it consider rectal gonorrhea or chlamydia in females. It also does not include cases among females who have sex with females only, given the relatively low levels of transmission risk for this group.

- Due to data limitations, this tool does not separate out transgender adolescents, although transgender adolescents are included with the sex with which they identify, given the options included in the source surveys.
- The age distribution of the population of high school-attending students in the United States varies dramatically over the course of the year, since students experience birthdays evenly throughout the year, but graduation is highly concentrated around June. Thus, in autumn, there are relatively few 18 year-olds-in high school (approximately 1.5 million) but by May there are relatively many (approximately 2.5 million). This is despite the fact that the overall population of high-school-attending adolescents differs only slightly between these two time periods. Given the nature of our data sources, this tool should be interpreted as reflecting the approximate population during the time of year when most states and local jurisdictions conduct the YRBS, i.e., February-May.
- The highest age category in the tool is meant to represent 18-year-olds, and those parameters derived from non-YRBS sources will generally be limited to this specific age. However, YRBS combines high school students over 18 with those who are 18 into a category labeled 18+. Since relatively few persons 19 and older are still in high school, and these individuals are impossible to distinguish in the public data, we include them, knowing that they will only impact the overall results by a small percentage. For cells in which most users will be inputting data from YRBS specifically, we write "18+." Users with other data sources should feel free to limit these entries to 18-year-olds only.
- This tool can be used to evaluate the impact of sexual behavior changes on the burden of gonorrhea, chlamydia and HIV, but it cannot be used to determine what intervention has caused the sexual behavior changes.
- Use of teen-SPARC requires Microsoft Windows.

As you use the tool, any questions, comments, or requests you have can be directed to Steven M. Goodreau, at goodreau@uw.edu.

ii. Acknowledgments

We would like to thank the staff of Emory-CAMP (Coalition for Applied Modeling Project), most especially Monica Trigg, Taylor Wimbly, and Megan Fields; the members of CAMP's Public Health Advisory Group (Jane Kelly, Michelle Allen, Nanette Benbow, Thomas Bertrand, Susan Blank, Mary Ann Chiasson, John Douglas, Gregory Felzien, David Harvey, Jonathan Poe, and Dan Wohlfeiler), and other CAMP researchers, including Deven Hamilton, Samuel Jenness and David Katz. We also thank Rachel Malloy of the New York State Department of Health, Katharine Howe of the Rhode Island Department of Health, and members of the Network Modeling Group at UW, especially Martina Morris.

iii. Disclaimers

This tool requires Microsoft Windows to run.

This tool may not be compatible with Microsoft Excel 2013 or earlier versions.

Users must enable macros. Please see section D0 for instructions.

Estimating the impact of behavior change is subject to uncertainty. As specified in the *Welcome* section, this tool is designed to estimate the potential impact of different behavior change interventions. The results are subject to uncertainty. The actual impact of behavior change could differ from those estimated by this tool.

This tool, like all software, may be subject to programming errors. The authors have taken reasonable efforts to identify and eliminate as many as they can, but others may remain, impacting results. Any users who believe they have identified an error in the program should contact Steven M. Goodreau (goodreau@uw.edu).

A. Definitions

In the tool, and throughout this manual, we will use the following standardized terms. Note that the details in these definitions are important in many cases for specifying who should be included in the model. This impacts decisions to be made about calculating population sizes and model parameters relevant to the user. Please note them carefully.

- *MSM*: male(s) who has sex with males. This includes males who have sex with males only and those who have sex with both males and females.
- *MSF*: male(s) who have sex with females only.
- *FSM:* female(s) who have sex with males. This includes females who have sex with males only and those who have sex with males and females both.
- *FSF:* female(s) who have sex with females only. This group is not modeled in this tool given their relatively low risk of the three STIs of interest. They do appear in one of the methods for estimating population sizes, however.
- *sexually experienced*: having had sexual intercourse at least once. Although we write out this full definition in most cases, we at times rely on the shorter phrase "sexually experienced" to avoid repetition.
- *CDC:* The United States Centers for Disease Control and Prevention.
- *DASH:* The Division of Adolescent and School Health; a division of CDC's National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention.
- *YRBS*: Youth Risk Behavior Survey, a component of the Youth Risk Behavior Surveillance System (YRBSS). The YRBSS monitors health behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and young adults in the United States. The national YRBS is a survey, conducted every two years by the CDC, which provides data representative of 9th through 12th grade students in public and private schools in the US. The state, territorial, tribal government, and local surveys, conducted by departments of health and education, provide data representative of mostly public high school students in each jurisdiction.
- *NSFG:* National Survey of Family Growth. A household survey conducted among 15-44 year olds in the United States designed to collect information on family life, marriage and divorce, and general and reproductive health. Developed and administered by CDC's National Center for Health Statistics (NCHS). For this tool, NSFG data are used to develop a few of the default parameters; for these, we combine data from 2006-2015 in order to have a sizeable sample of adolescent MSM.
- *Population:* in the context of this tool, *population* means the set of all people who meet the inclusion criteria for the tool. In most cases, it refers to 13-18 year olds attending high school within the user's jurisdiction who have ever had sexual intercourse. For some parameters or users, it can include those not attending high school as well. (The distinction is explained

throughout this manual). Population is thus a separate concept from a *sample* from this population used to obtain information about the population (e.g., school- or clinic-based surveys).

- *Subgroups*: in the context of this tool, the word *subgroups* refers to the nine groups modeled, defined by three sexual partnering groups (MSM, MSF, FSM) crossed with three age groups (13-15, 16-17, 18).

B. Basic model structure and systems thinking

The main aim of this tool is to help health departments predict the impact of behavior change on the expected number of STI diagnoses among adolescents in high school in their jurisdictions. This allows them to conduct thought experiments about the relative impact of different interventions, and assess how it might differ by STI, age, and sexual partnering group.

The basic logic of the calculations follows the principle that, for any given subgroup, the expected number of new infections in the coming year should be a function of:

- number of susceptible people (those not currently infected)
- mean contact rate for those susceptible people (including number of partners, frequency of sexual activity with these partners)
- probability that a partner is infected
- probability of transmission given sexual contact between susceptible and infected persons

Although this core logic is simple, each of the four components within it involves a variety of forms of complexity, explained in the Guided Instructions (Section D) as they appear. The entire logic is also complicated by a number of factors, also addressed throughout Section D:

(1) Calculations are subdivided by 3 sexual partnering groups (MSM, MSF, FSM) and three age groups (13-15, 16-17, 18), and conducted for multiple STIs (gonorrhea, chlamydia and HIV for MSM, gonorrhea and chlamydia for MSF and FSM).

(2) Data are generally only available for diagnoses, which is an incomplete picture of actual infections.

(3) Some data sources limit the frame of reference to those attending high school (our interest), while others do not.

(4) The age distribution of high school students changes substantially over the course of the calendar year.

(5) In many cases, we are only interested in those who have ever had sexual intercourse, as they are the only ones at risk of acquisition, but some parameter sources include everyone of any age.

(6) Gonorrhea and chlamydia can infect multiple sites in the body, and can be transmitted via different types of contact, with different probabilities.

Despite these complexities, the basic logic of the mathematical relationship above helps us to predict upcoming incident infections. In a traditional “dynamic” epidemic model, these incident infections then feed back to change prevalence, which changes both the number of susceptible people in the population and the probability that their partners are infected, which then affects incidence further. This feedback loop is the defining feature of dynamic epidemic models. This tool, on the other hand, is a short-term or "static" model: we present expected calculations for one upcoming time period (in this case, the next year), without compounding outcomes any further.

This general approach, in which future population-level outcomes like incidence are calculated explicitly in terms of current conditions like prevalence, transmission probabilities, sexual contact patterns, and the like is often called "systems thinking", because it recognizes that infectious disease transmission is a system of interacting individuals. In that system, there are important mathematical links between behavioral, biological, clinical, and demographic inputs.

We now turn to the four components listed in the expression above, and provide an overview of the logic this tool uses for each. Here we walk through the conceptual logic so it is all in one place; in later sections we walk through each worksheet of the tool in greater detail, explaining how users can include their own data (Section C), how we derived our default values for all parameters (Section G), and how we conducted *model calibration* using the transmission probabilities (Section H). We provide this information to fully document all of the assumptions and resulting mechanics for the model. Users who learn more through hands-on application may wish to skip this section and proceed to the Guided Instructions and Examples.

Number of susceptibles. The tool calculates this number by first determining (a) the number of people in each subgroup, and then subtracting (b) the estimated number of those who are currently living with an infection and thus unavailable to become infected.

For (a), the tool offers two alternatives, a main one that can be followed for those with YRBS data, and an alternative that will be easier for most users without YRBS data. The logic for the main option (on the *Population* worksheet) is to first enter the number of enrolled high school students in the jurisdiction, then the proportion of high school students by sex and age, then the proportion of each sex/age group that has ever had sexual intercourse, by the sex(es) of their partners. The logic for the alternative option (on the *Advanced Options* worksheet) is to begin with the population of the jurisdiction by age and sex (regardless of whether in school); then multiply by the proportion of each age/sex group who attend high school; and then multiply by the proportion who have ever had sexual intercourse, by the sex(es) of their partners.

For (b), the calculations rely on the fact that jurisdictions typically only have data on diagnosed cases, not on prevalence directly. We begin with new diagnoses by sex partnering group (per 12-month period for gonorrhea and chlamydia, and cumulative for HIV). We then divide by estimates for the proportion of incident cases that get diagnosed (*Advanced Options* worksheet; national default values filled in). This gives us an estimate for incidence. For gonorrhea and chlamydia, we rely on the approximation that holds in stable epidemics, where prevalence equals incidence times duration of infection. Thus, we multiply the incidence value by the average duration of infection (*Advanced Options* worksheet) to get prevalence. HIV is estimated differently, since it does not have a short duration. Instead, we use percent of the adolescent MSM population diagnosed with an HIV infection (overall, not per 12-month period). We then divide by the estimated proportion of MSM adolescents living with HIV who are aware of their status. We then multiply by a value that represents the ratio between prevalence and incidence estimated from national data. This yields an estimate for the number of adolescent MSM currently living with HIV.

For all diseases, these numbers represent prevalent cases in the relevant age group, whether in or out of school. In order to limit these to current students (in order to make the number comparable to those calculated in (a), we multiply these by the proportion of students in school in the relevant age and sex group.

Throughout this logic, we are assuming that adolescent who are not in school are equally likely to have an STI as those who are; we recognize that this assumption is likely false, but solid data to quantify the actual disparity for each subgroup for each STI are lacking.

Mean contact rate. Here we define contact as an act of penile-vaginal or penile-oral sex. Simple models, like this one, typically rely on a single (mean) contact rate. Reality is of course much more complex; individuals vary in their rates of sexual activity across many different dimensions. They also vary in whom they have sex with; partners are typically not chosen randomly from the population. Many (but not all) acts occur within the contexts of relationships, and the mathematics behind dealing fully with repeated sexual acts structured in relationships which are themselves structured into complex sexual networks can get quite complex. It also requires rich forms of data that are generally lacking for adolescent populations across many jurisdictions. Finally, in reality there are also additional types of sexual acts that carry non-zero risk of transmission, which we exclude here for simplicity.

Given that our population was defined above as only including those who have ever had sexual intercourse, our contact rate is also conditional on sexual experience. We are thus implicitly recognizing one form of contact heterogeneity in the population; some individuals have never had sex and are at no risk; while others have, and all calculations focus only on the latter group.

Our approach draws on data types we found to be relatively available: cumulative number of lifetime partners reported by adolescents of different ages, and numbers of acts per partner. We use back-calculation to determine the number of new partners per year adolescents likely have in order to obtain the age pattern in cumulative lifetime partners reported in data. For users with lifetime partner numbers, from YRBS or other source, the tool provides a place to enter the information (the New Partners per Year section of the *Advanced Options* worksheet). The back-calculation itself is then performed on the *ppy_calc* worksheet, although users should not ever need to visit this worksheet directly. We then multiply partners per year by sex acts per partner to obtain an estimate of the number of sex acts per individual per year. Again, this is conditional on having ever had sexual intercourse. Because of this, our estimates for sex frequency do not increase across the age span of adolescents as one might otherwise expect; rather, it is the size of the sexually-experienced population, to whom these numbers are applied, that increases dramatically.

Finally, we reduce the effective contact rate by including a parameter for the percentage of sex acts that involve the use of a condom. For the sake of simplicity and given available parameter estimates, the model only focuses on transmission events when condoms are not used (i.e., it assumes an effectiveness rate of 100% for condoms).

Probability partner is infected. This is another case where reality is very complex. The probability that a person's partner is infected depends on the pool of people who serve as likely partners for that person, and the fact that those with greater sexual activity are more likely to have STIs and are also more likely to be someone's partner (by definition). Here again, we use a relatively simple model, given the lack of rich data on sexual networks for adolescents. We assume that the prevalence of infection in the sexually-experienced adolescent population, both in and out of school, provides an estimate for the probability that one's partner is infected. In doing so, we match up the sexual partnering groups appropriately; i.e., FSM represent the partner pool for MSF, and vice versa; while MSM represent their own main partner pool. MSM also have contacts with FSM at considerable rates, which is included in the calculations. However, since the FSM population is so much larger than the MSM population, the same number of contact represents a much smaller fraction of FSM total contacts than it does for MSM; so small, indeed, that it can be ignored in terms of the general risk for FSM. (This fact is much less true for HIV than for gonorrhea and chlamydia, given the high disparity in HIV prevalence between MSM and other populations; however, we do not model HIV for FSM in this tool).

Given this, the methods for calculating partner prevalence parallel those for calculating prevalence in each subgroup, described above under *Number of susceptibles*. The only difference is that there we adjust to make the count among school-attending adolescents only (since they reflect the population of interest) and here we are interest in all adolescents (the partner pool). Moreover, since here we are interested in the probability the partner is positive, we convert the absolute number into a probability by dividing by the total number of adolescents in each relevant subgroup.

Probability of transmission

The probability of transmission per condomless sex act between a susceptible and an infected individual differs by STI and by act type. Act types are listed with the role of the individual with infection first, and the susceptible individual second. Thus, the transmission probability for penile-vaginal sex refers to the risk to an uninfected female from an infected male, while vaginal-penile sex is the reverse.

Transmission probabilities are the set of parameters we used to calibrate our model; those interested in understanding the values here or perhaps changing them should first read *Section H: Technical Note: Model calibration*.

C. Accessing YRBS data

For users who will be using YRBS as their main data source and who do not already have access to the data locally, they can be accessed at https://www.cdc.gov/healthyyouth/data/yrbs/data.htm. Users should choose the appropriate dataset for their jurisdiction. Note that this may be part of the "Combined YRBS Datasets and Documentation" section of the webpage, or may require a request, as explained in the "Requesting Data Files" section of the webpage. Many default values in the tool are based on the 2015 National YRBS Dataset.

Users can use their preferred statistical software. Data analysis guidelines as well as SAS and SPSS program code are found at the above link under the "Programs" tab of the "Combined YRBS Datasets and Documentation" section and the "National YRBS Datasets and Documentation" section. Users will need to edit the program code to provide correct file locations on their computer; instructions for this can be found at the top of the file containing the code.

In addition to reviewing this tool manual, we strongly recommend reviewing the YRBS Data User's Guide prior to conducting YRBS analyses. More detailed guidance on analytic software entitled "Software for Analysis of YRBS Data" is found under the "Additional Resources" section of the webpage.

We have created additional SAS code to generate tool inputs from jurisdiction-specific YRBS data. All of the SAS code (including code to export it into an Excel spreadsheet for easy copying and pasting into the tool) is compiled in a single file for easy extraction, found on the tool website. The code is designed so users will simply need to edit the program code to define the file location of their YRBS formats and import programs and to edit question numbers as needed (shown below). YRBS question numbers may vary by jurisdiction or year. As a result, the relevant question is also provided in the SAS code.

```
*Define location for formats and import programs;
      %let filepath = C:\YRBSTool;
*Define variable for "How old are you?";
      %let age_question = Q1;
*Define variable for "What is your sex?";
      %let sex_question = Q2;
*Define variable for "Have you ever had sexual intercourse?";
      %let intercourse_ever = Q60;
*Define variable for "During your life, with whom have you had sexual contact?";
      %let intercourse_who = Q67;
*Define a variable for "The last time you had sexual intercourse, did you or your
partner use a condom?";
      %let condom_question = Q65;
*Define "How old were you when you had sexual intercourse for the first time?";
      %let firstsex_question = Q61;
*"During your life, with how many people have you had sexual intercourse?";
      %let partners_question = Q62;
*********************************************************************************;
```

```
*Once the YRBS format and input programs are set up with the user's file locations,
they can be easily run from this program in the following step;
%include "&filepath.\YRBS-2015-SAS_Formats_Program.sas";
%include "&filepath.\YRBS-2015-SAS_Input_Program.sas";
```

Sections of this code are provided in this manual, in the relevant sections and described in more detail. For those using the SAS-to-Excel code, the following code is useful to run for all sections, since it makes the output match the formatting for the tool:

```
*Change fonts for excel output;
proc template;
	define style excel;
	class data / fontfamily="Arial" fontsize=11pt;
	class header / fontfamily="Arial" fontsize=11pt fontweight=bold;
	class byline / fontfamily="Arial" fontsize=11pt fontweight=bold
	textalign=center;
	class systemtitle / fontfamily="Arial" fontsize=12pt fontweight=bold;
	class systemfooter / fontfamily="Arial" fontsize=10pt textalign=left;
	class table / backgroundcolor=#f0f0f0 bordercolor=black borderstyle=solid
	borderwidth=1pt cellpadding=5pt cellspacing=0pt rules=groups;
	end;
run;
```

D. Guided instructions

D0. General instructions

Begin by opening up the tool file (teen-SPARC_1.0.xlsm).

The instructions below will walk you through the tool worksheet by worksheet. Here we note some general features.

Macros:

This file utilizes macros as part of the navigation and model programming, and these macros need to be enabled in order for the model to run properly. Depending on the user's comfort level, there are a few options for allowing this to happen.

- When first opening the document, the user should see either a pop-up or a message bar below the toolbar asking whether or not this file should be classified as a "Trusted Document." Click "yes, make this file a trusted document" (or "enable macros" if the message is in the tool bar, as in the example in the website below). Either option will enable the macros to run *in this document only*. This option will not be present in the rare case where users have already disabled all macros for all documents.
- Go to File > Info > Security Warnings. Here there are 2 options: 1) make this file a trusted document and enable macros (same options as above), or 2) enable macros for this session only. Choose this option if you do not wish to make this file a trusted document, but want the macros to run in your current session. If you choose this option, every time you open the document, you will be re-asked verify your security preferences.
- We do NOT recommend changing your advanced settings in the "Trust Center", as this will change your preferences for ALL excel documents, which can make the user vulnerable to untrustworthy files with mal-intentioned macros.
- For more information, visit: https://support.office.com/en-us/article/trusted-documents-cf872bd8-47ec-4c02-baa5-1fdba1a11b53.

Cell colors:

- White indicates cells that are designed for users to enter information.
- Green and dark gray (on the *Results* worksheet) indicate cells that yield outcomes of key interest to users
- Light gray buttons indicate places to press to move through the tool, or update or revert cell values
- Yellow cells and blue boxes indicates calculations made automatically that are steps along the way to the final outcomes of interest.
- Red indicates check boxes; whether these are checked or unchecked determines which option is run, in cases where there is more than one option available

Defaults:

- Nearly all white cells contain default numbers in them. The sources for these numbers are described in the Section G: *Parameter Documentation* of this document. Users are free to leave default parameters in place and change those they have data for.
- Most worksheets have buttons on them. These include:
  - *Reset values*: this will remove the data from the adjacent set of cells and leave them blank.
  - *Set to defaults*: this will change the values of the adjacent set of cells to the default values provided with the tool
  - *Continue to next step*: this advances the user to the next worksheet in the tool.
  - *Back to Table of Contents*: this returns the user to the *Table of Contents* worksheet.
- Other buttons that appear only on one or two worksheets are described below with those specific worksheets.
- **VERY IMPORTANT**: although in general users have the choice to input their own data or accept defaults, **any inputs that represent numbers of adolescents must either all be updated together, or all set to defaults.** Since they represent different aspects of absolute population size for the same jurisdiction, **changing some but not others would lead to values that do not represent any reasonable population scenario**. These types of inputs are:

  - For those using the *Population* worksheet:
    - *Population* worksheet Step 1: total population of high-school-attending adolescents
    - *Population* worksheet step 4: total adolescent population
    - *Diagnoses* worksheet: number of diagnoses

  - For those using the *Alternative Population Size Calculator* on the *Advanced Options* worksheet:
    - *Alternative Population Size Calculator* Step 1: total adolescent population
    - *Diagnoses* worksheet: number of diagnoses

D1. *Welcome* worksheet and *Table of Contents*

If this is your first time using the tool, we suggest reading the *Welcome* Worksheet, and then proceeding through the Table of Contents on to the *Population Size* worksheet, using the button at the top of the *Table of Contents* worksheet. If you are returning, you may use the *Table of Contents* worksheet to jump straight to a worksheet you want to change, or use the worksheet tabs at the bottom to select that worksheet directly.

D2. ***Population sizes***

This worksheet provides the user an option to enter the population sizes for each of their nine subgroups, using data types that are most commonly available for YRBS jurisdictions. Users may also choose to use the alternative population size calculator found on the *Advanced Options* worksheet.

For users with YRBS data, we recommend the general process described below. SAS code using national YRBS variable names and numbers is also provided. Variable names and question numbers may vary in jurisdiction-specific YRBS datasets. As a result, the relevant YRBS questions are also listed in the SAS code with instructions on how to customize the code for jurisdiction-specific YRBS datasets.

Step 1. Enter the total number of high-school-attending adolescents (grades 9-12) in your jurisdiction. Instructions for obtaining this number are:

a. Go to American Factfinder (factfinder.census.gov)
b. Click on Guided Search, then “Get me Started”.
c. Select “I'm looking for information about people”, then click “Next.”
d. Click the + next to “Education” to expand options, then select “School Enrollment”, then click “Next”.
e. Select your geographic *type*, e.g., United States, State, County, Place (meaning city/town), Metropolitan Statistical Area, American Indian Area, School District.
f. Select the specific geographic area from the list that appears (for smaller areas like counties or cities, this may require two steps—selecting the state first, then the specific county or city).
g. Click “Add to your selections”, then click “Next”.
h. If you see a question asking if you are looking for data for a race or ethnic group, click “Skip this step” or “Next”.
i. Click on one of the entries labeled “School Enrollment by Level of School for the Population 3 Years and Over” (code B14001). If this label appears more than once, it represents different years; select the year closest to your year of interest.
j. From the resulting table, read off the number next to “Enrolled in grade 9 to grade 12”.

Step 2. Enter the weighted proportion of males and females by age group in your YRBS dataset. To obtain this:

a. Define a new variable that determines whether someone is in age group 13-15, age group 16-17, age group 18, or none of the above. Call this variable *agegroup*.

```
data yrbs_tool;
   set dataout.yrbs2015;
   if &age_question in (2,3,4) then agegroup=1;
   else if &age_question=5 | &age_question=6 then agegroup=2;
   else if &age_question=7 then agegroup=3;
   else agegroup=4;
```

b. Define a second age variable that does not include none of the above. Call this variable *age2.*

```
if agegroup~=4 then age2=agegroup;
```

c. Define a new variable that indicates whether someone is male or female. Call this variable *sex*.

```
if &sex_question=1 then sex=1;
else if &sex_question=2 then sex=0;
```

d. Conduct a crosstab on *age2* by *sex* to determine the proportion of males and females by age group. This will be shown as the percent in the *proc surveyfreq* output using the code below.

```
proc surveyfreq data=yrbs_tool nomcar varheader=label;
   strata stratum; cluster psu; weight weight;
   tables age2*sex/cl row;
   format age2 age3cat. sex sex.;
   ods select CrossTabs;
   ods output crosstabs= yrbs_totalpop;
run;
```

e. The following steps clean the resulting table, rearrange the table to remove unnecessary columns, and prepare the data in an easier to read format.

```
data yrbs_total;
   set yrbs_totalpop;
   *Calculate proportion from percent and round;
   sexdist=round(percent/100,0.0001);
   *Remove total rows generated by crosstab;
   if F_age2="Total" or sex=. or sexdist=. then delete;
   keep  age2 sex sexdist;
run;
proc sort data = yrbs_total; by descending   age2; run;
data yrbs_total_final;
   merge
   yrbs_total (where = (age2 = 1) rename = (sexdist = sexdist1) )
   yrbs_total (where = (age2 = 2) rename = (sexdist = sexdist2) )
   yrbs_total (where = (age2 = 3) rename = (sexdist = sexdist3) );
   by  sex;
   drop age2;
   label        sexdist1="13-15"
                sexdist2="16-17"
                sexdist3="18+";
run;
```

f. Export your results into Excel for easy copying-and pasting (replacing USERFILENAME with the full file name of your choice):

```
ODS tagsets.ExcelXP FILE="&filepath\USERFILENAME.xls" style=excel
      options(sheet_name='Adolescent population size'
      embedded_titles='yes' embedded_footnotes='yes'
      merge_titles_footnotes='yes');
   proc print data=yrbs_total_final label noobs;
```

```
        title 'YRBS weighted proportion males and females by age group';
           footnote 'On the pop tab of the tool, copy the weighted
           proportions shown here into step 2';
   run;
ODS tagsets.ExcelXP CLOSE;
```

Step 3. Enter the proportion of adolescents who have ever had sexual intercourse into the three sexual partnering groups in your YRBS dataset. Note that these proportions are out of the sex-specific population—e.g., the value for MSF 16-17 is the proportion of *all 16-17 year-old males* who report having had sex with females, and only females.

a. Define a new variable that indicates whether someone is a sexually active MSM, MSF, FSM, or none of the above. To ensure there are no logical inconsistencies in the data, this variable should be defined based on the individual's sex, whether the individual has had sexual intercourse, and the sex of partners the individual has had sex with. Call this variable *sexpartgp*, for sexual partner group (SPG).

```
if (&sex_question=2 & &intercourse_ever=1 & (&intercourse_who=3 |
&intercourse_who=4)) then sexpartgp = 1;
else if (&sex_question=2 & &intercourse_ever=1 & &intercourse_who=2) then
sexpartgp=2;
else if (&sex_question=1 & &intercourse_ever=1 & (&intercourse_who=3 |
&intercourse_who=4)) then sexpartgp=3;
else sexpartgp=4;
```

b. Define a second SPG variable that does not include "none of the above". Call this variable spg2.

```
if sexpartgp ~=4 then spg2 = sexpartgp;
```

c. Using the crosstab of *agegroup* by *sex* by *sexpartgp*, calculate the percent of students with valid responses who appear in each of the nine subgroups of interest to the model. Note that these do not need to add to 100%, because some students will be not sexual active, or females who are only sexually active with other females, or will be younger than 13 or older than 18. The SAS code provided outputs a table of sex by SPG for each age group. The row percents for females with SPG FSM and males with SPGs MSM and MSF partners provide the relevant proportions of sexually experienced adolescents in each age group.

```
proc surveyfreq data=yrbs_tool nomcar varheader=label ;
   strata stratum; cluster psu; weight weight;
   tables agegroup*sex*sexpartgp / cl row ;
   format agegroup age4cat. sexpartgp spg4cat. sex sex.;
   ods select CrossTabs;
   ods output crosstabs = yrbs_eversex_out;
run;
```

d. The following steps clean the resulting table, rearrange the table to remove unnecessary columns, and prepare the data in an easier to read format.

```
data yrbs_eversex;
   set yrbs_eversex_out;
   *Create proportion sexually experienced from row percent and round;
   eversex=round(rowpercent/100,0.0001);
   *Remove total rows generated by crosstab;
   if agegroup=4 or sexpartgp=4 or (sex=1 and sexpartgp=1) or (sex=1 and
   sexpartgp=2) or (sex=2 and sexpartgp=3)
   or F_sexpartgp="Total" or eversex=. then delete;
   keep  agegroup eversex sexpartgp;
run;
proc sort data = yrbs_eversex; by sexpartgp agegroup; run;
data yrbs_eversex_final;
   merge
   yrbs_eversex (where = (agegroup = 1) rename = (eversex = eversex_age1))
   yrbs_eversex (where = (agegroup = 2) rename = (eversex = eversex_age2))
   yrbs_eversex (where = (agegroup = 3) rename = (eversex = eversex_age3))
   ;
   by sexpartgp;
   drop agegroup;
   label       eversex_age1="13-15"
               eversex_age2="16-17"
               eversex_age3="18+"
               sexpartgp="Sexual partner groups";
run;
```

e. Export your results into Excel for easy copying-and pasting (replacing USERFILENAME with the file name of your choice):

```
ods tagsets.Excelxp FILE="&filepath\USERFILENAME.xls" style=excel
      options(sheet_name='Proportion sexually experienced');
   proc print data = yrbs_eversex_final label noobs;
            title 'YRBS proportion of sexually experienced adolescents by
            age group and SPG';
            footnote 'On the pop tab of the tool, copy the proportion of
            adolescents who are sexually experienced into the three sexual
            activity groups in step 3';
            footnote2 'If using the advanced options, copy the proportion
            of adolescents who are sexually experienced into the three
            sexual activity groups in step 3';
   run;
ODS tagsets.ExcelXP CLOSE;
```

Step 4. Enter the estimates for the total number of individuals in the user's population by sex and age group. Although this step may seem unnecessary after having completed Step 1, it is useful for interpreting later data on diagnoses that are not limited to school-attending adolescents only. To use US Census Bureau numbers, follow these steps:

a. Go to American Factfinder (factfinder.census.gov)
b. Click on Advanced Search, then "Show me All"
c. Enter "age and sex" into the "Topic or table name" box, and enter your jurisdiction's geographical name (state, county, city, etc.) in the "state, county or place" box. Press go.

d. You may be asked to clarify the geographical entity you are looking for.
e. Look through the results for a table reporting data by "single *year* of age and sex" or "single *years* of age and sex" (the Census Bureau's wording is inconsistent). Select the table for the most recently available data. If your jurisdiction is a state, this should be within the last few years. If it is anything else, it will likely be the most recent census year.
f. From this table, extract the number of females and males, in each age 13-18, and add as needed to obtain the tool's three age categories (13-15, 16-17, 18). (Tip: if the data for your jurisdiction are old, you might choose to increase them by a fixed percentage based on how much your jurisdiction's overall population has grown between then and now. This is especially important for jurisdictions that are growing very rapidly. The Census Bureau table on overall population size by year can help you to determine how much that growth has been).
g. Enter these numbers in the white cells in Step 1. You may also choose to use the defaults, which represent the relevant group sizes for the entire US.

After you have filled in all cells, click *Proceed to Next Step*.

Users who do not have YRBS data may choose to enter estimates of population sizes here derived from other means, or use American Factfinder to obtain an overall school population size and accept national defaults for the remaining steps. Alternatively, they may choose to adopt the alternative approach, implemented on the *Advanced Options* worksheet. See section D8 if you chose to explore this option.

D3. *Sexual Behavior*

In the context of this model, behavior comprises three sets of metrics. We take each in turn below. As a reminder, this tool only tracks penile-vaginal sex between a male and a female, and penile-anal sex between two males.

1. *Average number of new partners per year, among those who have ever had sexual intercourse, by sup-population.* Note that these values condition on sexual experience; those who have never had sexual intercourse are not included in the denominator. This must be taken into account when conducting calculations or interpreting numbers. For example, in the default values for MSF, the quantity is the same across all age groups (1.32). This does not mean, however, that the total amount of sexual activity is the same across age groups, since the proportion of the population that has ever had sexual intercourse increases with age.

   The average number of new partners per year is a necessary part of calculating expected incident cases; however, it is not a question that many surveys ask directly. If users do have access to data on this metric from their jurisdiction, they may enter it directly. They might also choose to accept the default national numbers, which have been back-calculated from YRBS questions about lifetime partner counts. Users from YRBS jurisdictions may choose to use their local data on lifetime partner counts to generate new estimates of annual numbers of new partners. This can be done on the *Advanced Options* worksheet, in the "New Partners per Year Calculator" section; please see section D9 for instructions, with SAS code.

   Note that the national default numbers for 16-7 year olds and 18-year olds are the same; this is a coincidence, and does not reflect some assumption of the model that requires them to be the same.

   Users may be wondering whether, by focusing only on new partners, we are ignoring the possibility that someone could have a partnership that lasts longer than one year, and is thus carried over from a previous year. Doing so would bias down the amount of sexual contacts individuals are having. We are indeed unable to distinguish partners lasting less than one year from those lasting more than one year in the data. However, our method does not bias the overall sexual activity downward because it is based on reported sexual frequencies per time unit among those in partnerships. Instead, it simply reallocates them to the new partners at the relevant age.

2. *Average number of acts of intercourse per partner that each subgroup averages*

   This measure is also a necessary step in calculating expected incident cases, but is also something not commonly asked in surveys. YRBS does not include a metric for this. The defaults included are calculated from NSFG national data. Users may accept these defaults; change to other values if they have their own local data; or try out different numbers as a sensitivity analysis.

3. *Probability of condom use per sex act, by subgroup.* Condom use is asked about directly in YRBS. Instructions for extracting this for a local jurisdiction are:

   a. Make sure that you have the *spg2* and *age2* variables created from the Population size section (see Section D2).

   b. Define a new variable that indicates whether someone had sex with a condom the last time they had sexual intercourse. Call this variable *condomuse.*

```
if &condom_question=2 then condomuse=1;
if &condom_question=3 then condomuse=0;
```

   c. Perform logistic regression to determine the predicted condom use for each SPG and age group. The resulting mean is the predicted probability for each group.

```
proc surveylogistic data=yrbs_tool  varmethod=taylor;
   format age2 age3cat. spg2 spg3cat. condomuse condomuse.;
   strata stratum; cluster psu; weight weight;
   class condomuse (ref="No") age2 (ref="18+") spg2 (ref="FSM")/param=ref;
   model condomuse= age2 spg2;
   ods select Estimates;
   ods output estimates = yrbs_condom_out;

   *Create estimate statements to get predicted probability for SPG by age
   group combination;
   estimate '13-15 MSM' intercept 1 age2 1 0 spg2 0 1 /ilink;
   estimate '16-17 MSM' intercept 1 age2 0 1 spg2 0 1 /ilink;
   estimate '18+ MSM'   intercept 1 age2 0 0 spg2 0 1 /ilink;
   estimate '13-15 MSF' intercept 1 age2 1 0 spg2 1 0 /ilink;
   estimate '16-17 MSF' intercept 1 age2 0 1 spg2 1 0 /ilink;
   estimate '18+ MSF'   intercept 1 age2 0 0 spg2 1 0 /ilink;
   estimate '13-15 FSM' intercept 1 age2 1 0 spg2 0 0 /ilink;
   estimate '16-17 FSM' intercept 1 age2 0 1 spg2 0 0 /ilink;
   estimate '18+ FSM'   intercept 1 age2 0 0 spg2 0 0 /ilink;

run;
```

   d. The following steps clean the resulting table, rearrange the table to remove unnecessary columns, and prepare the data in an easier to read format.

```
data yrbs_condom;
   set yrbs_condom_out;
   if label = '13-15 MSM' then do; age2=1; spg_condom=1; end;
   if label = '16-17 MSM' then do; age2=2; spg_condom=1; end;
   if label = '18+ MSM' then do;    age2=3; spg_condom=1; end;
   if label = '13-15 MSF' then do; age2=1; spg_condom=2; end;
   if label = '16-17 MSF' then do; age2=2; spg_condom=2; end;
   if label = '18+ MSF' then do;    age2=3; spg_condom=2; end;
   if label = '13-15 FSM' then do; age2=1; spg_condom=3; end;
   if label = '16-17 FSM' then do; age2=2; spg_condom=3; end;
   if label = '18+ FSM' then do;    age2=3; spg_condom=3; end;
   *Round predicted condom use;
   mean=round(mu,.0001);
   keep age2 spg_condom mean;
```

```
      label spg_condom = "Sexual partner groups";
   run;

   proc sort data = yrbs_condom; by spg_condom age2; run;

   data condomout_final;
      merge
      yrbs_condom (where = (age2 = 1) rename = (mean = condom1) )
      yrbs_condom (where = (age2 = 2) rename = (mean = condom2) )
      yrbs_condom (where = (age2 = 3) rename = (mean = condom3) );
      by spg_condom;
      drop age2;
      label    condom1="13-15"
               condom2="16-17"
               condom3="18+";
   run;
```

e. Export your results into Excel for easy copying-and pasting (replacing USERFILENAME with the file name of your choice):

```
ods tagsets.Excelxp FILE="&filepath\USERFILENAME.xls" style=excel
      options(sheet_name='Predicted condom use');
      proc print data=condomout_final label noobs;
            title 'YRBS predicted condom use for each SPG and age group';
            footnote 'On the behavior tab of the tool, copy the predicted
            condom use for each SPG and age group under "Probability of
            Condom Use"';
            footnote2   'Copy the predicted condom use table onto the MSM
            anal-penile row down through the FSM row';
            footnote3   'Copy the predicted condom use for MSF onto the Group
            1 Vaginal-Penile row';
            format spg_condom spg3cat.;
      run;
ODS tagsets.ExcelXP CLOSE;
```

Alternatively, users may accept the national default value (calculated this way from YRBS data) or use other local data sources, since many surveys ask respondents if they used a condom the last time they had intercourse.

D4: *Diagnoses*

This worksheet allows users to enter the number of diagnosed cases in the most recent 12-month period for which data are available for each of the three pathogens, and for each of the nine subgroups. These will typically be available from surveillance data, and should be scaled to reflect diagnoses in one year. If users have data from another time frame, they can rescale their numbers to reflect comparable sizes for a year (e.g., if a user has diagnoses in a 3-month window, they can multiply those numbers by 4).

A key point is that these data should represent diagnoses among adolescents *regardless of whether they are in school or not*. This is despite the fact that the outcome metrics are in terms of school-attending adolescents. While this may seem contradictory, it is done because we expect almost all jurisdictions to have existing surveillance data in this form. The methods underlying this tool automatically convert from the input (diagnoses among adolescents regardless of school status) to the relevant outputs (incidence and diagnoses among adolescents in school), so the users should not attempt to do this themselves.

Users will also need to remember aspects of the definitions specific to this tool; when the exact definitions do not match between the tool and surveillance data, users may need to make assumptions or educated guesses to approximate the numbers. These aspects include:

- Gonorrhea and chlamydia cases in the tool do not include pharyngeal cases in either sex, or rectal cases for females. However, these tend to be diagnosed infrequently among adolescents, especially as an exclusive site of infection.[1, 2]
- Ideally, females who have sex with females only are not to be included in diagnosis totals; however, rates of these three STIs among females engaging exclusively in sex with other females are low,[3] so even if jurisdictional data do not allow them to be distinguished, the actual impact on estimates is likely to be minimal.

Diagnoses data are used in the tool to estimate the existing prevalence of each infection in each subgroup in the modeled population. This is used to then determine the probability that an uninfected individual's partner is infected, a necessary condition for generating an incident case (see Section D11 for full details). The estimation of prevalence from diagnoses involves additional parameters found on the *Advanced Options* worksheet. Most users might wish to leave these defaults in place, but they can choose to alter them if desired; see the *Advanced Options* sections below.

For users proceeding through the tool for the first time, we recommend using the *Skip to Results* button after this worksheet. This will allow you to see the results for the set of numbers you have already entered; we call these the "baseline" results. For users who have already generated baseline results, and are ready to explore interventions, we recommend using the *Proceed to Interventions* button.

D5: *Interventions*

Users on their first time through the tool may wish to skip this worksheet for now, and proceed to the *Results*. To do so, make sure the red boxes are **unchecked**, and click *Continue to Results*.

Users who have already generated and viewed their baseline results, may now wish to consider the impacts of different forms of behavior change interventions. This worksheet allows the user to implement changes to their baseline parameters in order to compare scenarios. Interventions can take many forms in terms of implementation, and we do not lay out all of these possible details, since it is impossible to anticipate all of the types of interventions different jurisdictions might consider. Instead, this tool allows the user to input the expected impact of an intervention that either decreases the frequency of sex, increases the probability of condom use, or both. Indeed, despite the name of the worksheet, one may just as easily model the impact of behavior change that arises from causes other than explicit public health interventions.

For either form of behavior change, users may model a change of the same impact across all nine subgroups. This is done in the "across all groups" cell, which is found on the left side of the *Interventions* worksheet. Users may also model a change of different magnitude across any or all of the nine subgroups. This is done in the "within each group" table, which is found on the right side of the *Interventions* worksheet.

Percentages should be entered using decimal format for both the percent by which the frequency of sex acts should decrease and for the percent by which condom use should increase.

There are three very important things to note in completing these interventions:

1. The red boxes on the left must be used to indicate which of the interventions is to be considered. Users have the option of using neither, one or the other, or both.
2. For either intervention, if both the "Across all groups" and "Within each group" numbers are filled in, the "Across all groups" value will be used and the others ignored.
3. For either intervention, if the "Within each group" values are being used, and some cells are left blank, the tool will assume a value of 0, i.e., that there is no behavior change in that group.

Once the intervention is selected, users should return to the *Results* worksheet to examine their new scenario as a "comparison model."

Note that one also has the option of considering reductions in condom use or increases in sex frequency, simply by entering negative numbers instead of positive. (For example, -0.1 would represent a 10% decrease in condom use).

D6: *Results*

Baseline model

The first time through the tool, users will generally wish to click *Calculate Baseline Model.* This will calculate results for the model ***without*** applying any interventions on the *Intervention* worksheet.

The button to calculate the baseline model is located on the upper left, while results for that model will appear to the right. These include both expected incident cases (whether or not diagnosed) and expected diagnosed cases, for each of the three infections being modeled, for each relevant subgroup. Note cases are shown rounded to the nearest whole number, while % averted is calculated on the unrounded numbers.

If at any point a user wishes to change their baseline model, they may click *Reset Model* here to empty out the results cells. They can then go and change various parameters, return here and click *Calculate Baseline Model* again. Note that prior results will be lost in this case. If the user wishes to save those prior results before overriding them, they should use the *Export* worksheet.

On the far right is a small set of outputs labeled "Calibration"— this output is useful to guide developers and advanced users only; most users can choose to ignore these numbers. For more information, see *Section H: Technical note: Model calibration*.

Comparison scenarios

A user may then wish to consider one or more other models, and compare the results from these directly to the baseline model. These comparison scenarios might involve one or both of the interventions on the *Interventions* worksheet.

After setting up the comparison model, users should return here to the *Results* worksheet, and click one of the buttons saying *Calculate Comparison Model* (on the left side of the worksheet). Users will be prompted to enter a name to identify the comparison model. The comparison model will provide outcome numbers for the behavioral scenario that includes either the marked interventions effects and/or the changes made to any other parameters since the baseline model was calculated. These results include the absolute numbers of incident and diagnosed cases per subgroup, as well as the number and percent of cases averted relative to the baseline model. The number of cases averted is expressed in terms of all incident cases (not just diagnosed). Given the model assumptions and allowed interventions, the percent of cases averted is the same, regardless of whether one is considering all incident cases or only diagnosed cases.

Users can explore up to four different intervention scenarios, using a different *Calculate Comparison Model* button each time. Each time a button is pressed, the model will use whatever interventions are checked and filled in on the *Intervention* worksheet. Thus, the user who wants to explore multiple scenarios will typically find themselves alternating between *the* Interventions and *Results* worksheets. If users wish to explore additional scenarios, they should visit the *Export* worksheet in order to save any of the first four they wish to keep.

D7: *Export*

Users may wish to export model results in order to save scenarios or explore additional scenarios for further comparison. To do this, users should click on the *Export Results* button on the *Results* worksheet.

On the *Export* worksheet, users can choose which results they would like to export. Exports will ***always*** include baseline model and parameters so these are not listed as options, but users may select up to four comparison models to export in addition. Once all desired exports are checked, users should click the *Export to PDF* button. Exports will open in the user's default PDF viewer and automatically save on the user's desktop as "Model_Output." Users then should change the name of the file in order to prevent it from being written over by exported results at a later time.

D8: *Advanced options*: alternative population size calculator

This alternative option for specifying population sizes makes most sense for users who do not have access to YRBS data for their jurisdiction. Given that the model focuses on subgroups of sexually experienced adolescents in high school, this method breaks down the derivation of final population sizes into multiple steps.

For those using this method, turn to the *Advanced Options* worksheet.

First, it is **crucial** that you check the box in the upper-left corner of the worksheet to let the model know that you will be using this method instead of the numbers found on the *Population* worksheet.

Step 1 entails obtaining estimates for the total number of individuals in the user's population by sex and age group. This process is identical to Step 4 on the *Population* worksheet. To use US Census Bureau numbers, follow these steps:

a. Go to American Factfinder (factfinder.census.gov)
b. Click on Advanced Search, then "Show me All"
c. Enter "age and sex" into the "Topic or table name" box, and enter your jurisdiction's geographical name (state, county, city, etc.) in the "state, county or place" box. Press go.
d. You may be asked to clarify the geographical entity you are looking for.
e. Look through the results for a table reporting data by "single *year* of age and sex" or "single *years* of age and sex" (the Census Bureau's wording is inconsistent). Select the table for the most recently available data. If your jurisdiction is a state, this should be within the last few years. If it is anything else, it will likely be the most recent census year.
f. From this table, extract the number of females and males, in each age 13-18, and add as needed to obtain the tool's three age categories (13-15, 16-17, 18). (Tip: if the data for your jurisdiction are old, you might choose to increase them by a fixed percentage based on how much your jurisdiction's overall population has grown between then and now. This is especially important for jurisdictions that are growing very rapidly. The Census Bureau table on overall population size by year can help you to determine how much that growth has been).
g. Enter these numbers in the white cells in Step 1. You may also choose to use the defaults, which represent the relevant group sizes for the entire US.

Step 2 entails identifying the proportion of adolescents in each of the six age-sex groups who attend high school. (Note this does *not* mean the proportion of high school students who are within each group—that is, the numbers do not need to add to 1). These value may be difficult to obtain for many jurisdictions; users may thus choose to use the default values from national data, or scale them up or down based on overall school enrollment rates, if known.

Step 3 entails entering the percentage of students in each age group who have experienced sexual intercourse, according to the sex of their partners. Note that these proportions are out of the sex-specific population—e.g., the value for MSF 16-17 is the proportion of *all 16-17 year-old males* who report having had sex with females, and only females. These numbers may also be difficult to obtain for non-YRBS jurisdictions, so national defaults can again be accepted, or the values scaled up and down based on knowledge of overall adolescent sexual activity relative to national numbers.

D9: *Advanced Options*: New partners per year calculator

This method for calculating the number of partners per year is set up specifically for users with YRBS data. From the *Table of Contents* worksheet, users can access this method by clicking on *Advanced Options: Frequency of Sex Acts* or you can access it directly from the *Advanced Options* worksheet by scrolling down to the section called *New Partners per Year Calculator*.

First off, it is crucial that you check the box in the upper-left corner of the section to let the model know you will be using this method.

The purpose of this method is to convert from the specific type of data collected in YRBS (number of partners in the lifetime) to that needed for this tool (number of new partners in the last 12 months). We lay out the mechanics for doing so here; for interested readers, we lay out the conceptual logic behind this process in the Technical Note at the end of this section below.

Step 1. Calculate the weighted number of respondents by current age and age of first sexual intercourse.

a. Define a new variable of age that is continuous ages 13-18. Call this variable *age*.

```
if &age_question = "2" then age = 13;
else if &age_question = "3" then age = 14;
else if &age_question = "4" then age = 15;
else if &age_question = "5" then age = 16;
else if &age_question = "6" then age = 17;
else if &age_question = "7" then age = 18;
else if &age_question = "1" then age = .;
```

b. Define a new variable for age at first sex, ignoring those who have never had sex. Call this variable *firstsexage*.

```
if &firstsex_question = "2" then firstsexage=11;
else if &firstsex_question = "3" then firstsexage=12;
else if &firstsex_question = "4" then firstsexage=13;
else if &firstsex_question = "5" then firstsexage=14;
else if &firstsex_question = "6" then firstsexage=15;
else if &firstsex_question = "7" then firstsexage=16;
else if &firstsex_question = "8" then firstsexage=17;
```

c. Define a new variable for lifetime number of partners, ignoring those who have never had sex. Call this variable *lifesex*.

```
if &partners_question="2" then lifesex=1;
else if &partners_question="3" then lifesex=2;
```

```
else if &partners_question="4" then lifesex=3;
else if &partners_question="5" then lifesex=4;
else if &partners_question="6" then lifesex=5;
else if &partners_question="7" then lifesex=6;
```

d. Conduct a crosstab on *spg2* (created in the Population size section; see Section D2) by *age* by *firstsexage* and output to a new dataset.

```
proc surveyfreq data=yrbs_tool;
      strata stratum; cluster psu; weight weight;
      tables spg2*age*firstsexage;
      ods select CrossTabs;
      ods output crosstabs = yrbs_matrix1_out;
run;
```

e. The following steps clean the resulting table, rearrange the table to remove unnecessary columns, and prepare the matrix in an easier to read format.

```
data yrbs_matrix1;
      set yrbs_matrix1_out;
      if firstsexage=. then delete;
      if age=. then delete;
      if age=firstsexage then delete;
      freq=round(wgtfreq,.01);
      keep spg2 age firstsexage freq;
run;

proc sort data = yrbs_matrix1; by spg2 age firstsexage; run;

data matrix1;
merge yrbs_matrix1 (where = (firstsexage = 11) rename = (freq = freq1) )
            yrbs_matrix1 (where = (firstsexage = 12) rename = (freq = freq2))
            yrbs_matrix1 (where = (firstsexage = 13) rename = (freq = freq3))
            yrbs_matrix1 (where = (firstsexage = 14) rename = (freq = freq4))
`           yrbs_matrix1 (where = (firstsexage = 15) rename = (freq = freq5))
            yrbs_matrix1 (where = (firstsexage = 16) rename = (freq = freq6))
            yrbs_matrix1 (where = (firstsexage = 17) rename = (freq = freq7))
            ;
      by spg2 age;
      drop firstsexage;
      label freq1="11"
            freq2="12"
            freq3="13"
            freq4="14"
            freq5="15"
            freq6="16"
            freq7="17";
run;
proc sort data=  matrix1; by spg2 age; run;
data matrix1_table;
      set matrix1;
      by spg2 ;
      retain expected_age;

      *Reset expected age counter for each risk group;
      if first.spg2 then expected_age = 13;
else expected_age++1;
```

```
*Check if expected age = the age on the row. If they don't match, make a
blank version of the row;
      if expected_age ~= age then do;
output;     * Copy over the current row WITH data before we make a new one;

      *Begin to make a new blank row for the missing age year;
      age = expected_age;
      if expected_age = 13 then do;
freq1 = 0; freq2 = 0; freq3 = .; freq4 = .; freq5 = .; freq6 = .; freq7 = .;
end;
      else if expected_age = 14 then do;
freq1 = 0; freq2 = 0; freq3 = 0; freq4 = .; freq5 = .; freq6 = .; freq7 = .;
end;
      else if expected_age = 15 then do;
freq1 = 0; freq2 = 0; freq3 = 0; freq4 = 0; freq5 = .; freq6 = .; freq7 = .;
end;
      else if expected_age = 16 then do;
freq1 = 0; freq2 = 0; freq3 = 0; freq4 = 0; freq5 = 0; freq6 = .; freq7 = .;
end;
      else if expected_age = 17 then do;
freq1 = 0; freq2 = 0; freq3 = 0; freq4 = 0; freq5 = 0; freq6 = 0; freq7 = .;
end;

      *Because we added a row we need to increment expected age by 1 year;
      expected_age++1; end;
output;
run;

*Add zeros where numbers in matrix are missing;
proc sort data=  matrix1_table; by spg2 age; run;
data clean_matrix1;
      set matrix1_table;
      if age=13 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0; end;
      if age=14 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0;
            if freq3=. then freq3=0; end;
      if age=15 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0;
            if freq3=. then freq3=0;
            if freq4=. then freq4=0; end;
      if age=16 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0;
            if freq3=. then freq3=0;
            if freq4=. then freq4=0;
            if freq5=. then freq5=0; end;
      if age=17 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0;
            if freq3=. then freq3=0;
            if freq4=. then freq4=0;
            if freq5=. then freq5=0;
```

```
            if freq6=. then freq6=0; end;
        if age=18 then do;
            if freq1=. then freq1=0;
            if freq2=. then freq2=0;
            if freq3=. then freq3=0;
            if freq4=. then freq4=0;
            if freq5=. then freq5=0;
            if freq6=. then freq6=0;
            if freq7=. then freq7=0; end;
        drop expected_age;
run;
```

f. Export your results into Excel for easy copying-and pasting (replacing USERFILENAME with the file name of your choice):

```
ods tagsets.Excelxp FILE="&filepath\USERFILENAME.xls" style=excel
        options(sheet_name='Partners per year matrix 1' sheet_interval='proc');
        proc print data=clean_matrix1 label noobs;
            by spg2;
            format spg2 spg3cat. freq1 blank. freq2 blank. freq3 blank. freq4
            blank. freq5 blank. freq6 blank. freq7 blank.;
            title 'Matrix 1: Age at first sex by current age, cells=counts of
            persons';
            footnote 'Copy into Matrix 1 on the "Advanced Options Page" under
            "New Partners per Year Calculator"';
        run;
        ODS tagsets.ExcelXP CLOSE;
```

g. Enter the resulting weighted frequencies into *Matrix 1: Age of First Intercourse by Current Age, Cells = Weighted Counts of Persons*.

Step 2. Calculate the average number of lifetime partners among each set of respondents of a given current age and age of fist intercourse.

a. Using the variables created in step 1, run a proc surveymeans of spg2 by age by age of first intercourse in the domain line and lifesex in the var line. Output to a new dataset.

```
proc surveymeans data=yrbs_tool ;
        strata stratum; cluster psu; weight weight;
        domain spg2*age*firstsexage;
var lifesex;
ods output domain=matrix2;
run;
```

b. The following steps clean the resulting table, rearrange the table to remove unnecessary columns, and prepare the matrix in an easier to read format.

```
data partner;
        set matrix2;
        if age=firstsexage then delete;
        meanpartner=round(mean,.01);
        keep spg2 age firstsexage  meanpartner;
```

```
run;

proc sort data = partner; by spg2 age firstsexage; run;

data yrbs_partner;
        merge partner (where = (firstsexage = 11) rename = (meanpartner =
        mean1) )
        partner (where = (firstsexage = 12) rename = (meanpartner =
        mean2) )
        partner (where = (firstsexage = 13) rename = (meanpartner =
        mean3) )
        partner (where = (firstsexage = 14) rename = (meanpartner =
        mean4) )
        partner (where = (firstsexage = 15) rename = (meanpartner =
        mean5) )
        partner (where = (firstsexage = 16) rename = (meanpartner =
        mean6) )
        partner (where = (firstsexage = 17) rename = (meanpartner =
        mean7) );
    by spg2 age;
    drop firstsexage;
    label mean1="11"
        mean2="12"
        mean3="13"
        mean4="14"
        mean5="15"
        mean6="16"
        mean7="17";
run;
proc sort data=  Yrbs_partner; by spg2 age; run;

data matrix2_table;
    set Yrbs_partner;
    by spg2 ;
    retain expected_age;

    * reset expected age counter for each risk group;
    if first.spg2 then expected_age = 13;
    else expected_age++1;

    * check if expected age = the age on the row.
    if expected_age ~= age then do;
    output;* copy over the current row WITH data before we make a new one ;

    * begin to make a new blank row for the missing age year;
    age = expected_age;
    if expected_age = 13 then do;
    mean1 = 0; mean2 = 0; mean3 = .; mean4 = .; mean5 = .; mean6 = .; mean7
    = .; end;
    else if expected_age = 14 then do;
    mean1 = 0; mean2 = 0; mean3 = 0; mean4 = .; mean5 = .; mean6 = .; mean7
    = .; end;
    else if expected_age = 15 then do;
    mean1 = 0; mean2 = 0; mean3 = 0; mean4 = 0; mean5 = .; mean6 = .; mean7
    = .; end;
    else if expected_age = 16 then do;
```

```
        mean1 = 0; mean2 = 0; mean3 = 0; mean4 = 0; mean5 = 0; mean6 = .; mean7
        = .; end;
        else if expected_age = 17 then do;
        mean1 = 0; mean2 = 0; mean3 = 0; mean4 = 0; mean5 = 0; mean6 = 0; mean7
        = .; end;

        * because we added a row we need to increment expected age by 1 year;
        expected_age++1; end;
        output;
run;

*Add zeros where numbers in matrix are missing;
proc sort data=  matrix2_table; by spg2 age; run;
data clean_matrix2;
        set matrix2_table;
        if age=13 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0; end;
        if age=14 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0;
                if mean3=. then mean3=0; end;
        if age=15 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0;
                if mean3=. then mean3=0;
                if mean4=. then mean4=0; end;
        if age=16 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0;
                if mean3=. then mean3=0;
                if mean4=. then mean4=0;
                if mean5=. then mean5=0; end;
        if age=17 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0;
                if mean3=. then mean3=0;
                if mean4=. then mean4=0;
                if mean5=. then mean5=0;
                if mean6=. then mean6=0; end;
        if age=18 then do;
                if mean1=. then mean1=0;
                if mean2=. then mean2=0;
                if mean3=. then mean3=0;
                if mean4=. then mean4=0;
                if mean5=. then mean5=0;
                if mean6=. then mean6=0;
                if mean7=. then mean7=0; end;
        drop expected_age;
run;
```

c. Export your results into Excel for easy copying-and pasting (replacing USERFILENAME with the file name of your choice):

```
ods tagsets.Excelxp FILE="&filepath\USERFILENAME.xls" style=excel
```

```
    options(sheet_name='Partners per year matrix 2' sheet_interval='proc');
    proc print data=clean_matrix2 label noobs;
          by spg2;
          format spg2 spg3cat. mean1 blank. mean2 blank. mean3 blank. mean4
          blank. mean5 blank. mean6 blank. mean7 blank.;
          title 'Matrix 2: Age at first sex by current age, average number
          of lifetime partners';
          footnote 'Copy into Matrix 2 on the "Advanced Options Page" under
          "New Partners per Year Calculator"';
    run;
ODS tagsets.ExcelXP CLOSE;
```

d. Enter the resulting means into *Matrix 2: Age of Fist Intercourse by Current Age, Cells = Average Number of Lifetime Partners*.

Step 3: Determine the proportion of MSM's partnerships that are with females. The default value (14%) is included, and is derived from national YRBS data. Users can accept this, or calculate their own value if they have data from their jurisdiction (this cannot be calculated from standard YRBS questions).

Technical note: the user's inputs of two matrices related to lifetime partner numbers are used to calculate estimates for new partners per year by age. This occurs behind the scenes on the final sheet, named *ppy_calc*. Users ***do not*** need to access this sheet, or make any changes to it; it conducts the calculations automatically, and then feed the results into the *Sexual behavior* worksheet (assuming that the checkbox in the New Partners per year Calculator is checked).

For those interested in the logic of the calculations, the basic premise relies on the fact that reports of lifetime partner numbers by people in different ages (e.g., 15-year olds and 16- year olds) provides us an estimate of the number of new partners each year. However, this is complicated in this case by the fact that we are interested in the mean number of partners among the sexually experienced population only – and that population keeps getting larger with increasing age as more new people enter the category of sexually experienced, and start out with a single partner. We thus use information from each respondent on both their current age and their age of first sexual intercourse, in order to determine where they fit. We make the assumption that partners are distributed evenly between the age of first sex and the current age. Thus individuals can contribute to information about multiple ages – e.g., a 17-year old who first had sexual intercourse at age 15 provides information about the partners per year for 15, 16 and 17 year olds. Note that in estimating new partners per year, we do not include information from people who report having had sexual intercourse for the first time at the same age that they currently are; determining an average mean number of partners per year for such persons requires making a variety of assumptions that we deemed questionable. (These persons do appear in the total number of persons who have ever had sexual intercourse, however, assumed to have partner acquisition rates equaling others of the same age). Additionally, people who reported they have never had sexual intercourse are not represented in this matrix.

On the *ppy_calc* worksheet, the first set of matrices (labeled *A: Age of First Intercourse by Current Age, Cells = Counts of Persons*) provides the weighted counts from YRBS data in terms of how many respondents there are with every combination of current age (rows) and age of first intercourse (columns). Note, then, that nobody appears above the diagonal, i.e., reporting an age at first sex later than their current age.

The next matrix set (B) tells us the mean number of lifetime partners reported by all those with that current age/age at first sex combination. Both A and B are simply copies of those that appear on the *Advanced Options* page.

Matrix set C calculates the partners per year for each of the combinations (e.g., for those who first had sexual intercourse three years before their current age, their new partners per year is their total lifetime partners/3).

Now, we wish to take the weighted average of all of the reports about new partners per year for a given age. For instance, new partners per year at age 14 comes from 16 different groups' reports: those who are currently 15, 16, 17 or 18 and who first had sexual intercourse at 11, 12, 13 or 14. These calculations are complicated by the fact that they each need to be weighted by the number of observations in each of these combinations (as found in matrix set A). For any given age, then, we need a numerator that has each of the relevant reports of new partners per year, multiplied by their weight in cell A, and then all added together; and for the denominator we need the total number of people contributing observations, i.e., the sum of the weights in A for all relevant cells. The numerator calculation occurs in matrix set D and E, while the denominator occurs in matrix set F. Specifically, matrix set D holds the cell-specific product of matrices C and A as an intermediate step; in matrix set E, we then calculate the sum of this for all relevant groups. Using the example of age 14 from above, we need to sum 16 numbers from matrix set D for our numerator; each cell entry in the column for age 14 sums four of these, while the red total cell at the bottom then sums those 4 sums, giving the sum of all 16. Matrix set F then conducts a similar summation for the denominator, but sums the population size weights from matrix set A only. To get our final numbers, we divide the means in the red cells from E by those from F, taking weighted averages across multiple ages to get our final three age groups.

D10: *Advanced options*: Transmissibility

The transmissibility parameters represent the per-act probability of transmissions that lead to a base model that *calibrates* using data at the national level—that is, a model in which input diagnoses match output diagnoses in the absence of any interventions or behavior change. In most cases, the values for these transmissibility parameters lie within the range of published estimates for per-act transmission probabilities. More information can be found in *Section H: Technical note: Model calibration.*

Given this, we anticipate that few users will wish to alter these parameters. Those who do so should be mindful of the potential impacts on results.

From the *Table of Contents* worksheet, users can access these parameters by clicking on *Advanced Options: Transmissibility*, or directly from the *Advanced Options* worksheet by scrolling down to the section on Epidemiological Factors – Transmissibility.

Transmissibility probabilities for *Neisseria gonorrhea* and *Chlamydia trachomatis* are shown for four types of acts: penile-vaginal (receptive risk), vaginal-penile (insertive risk), penile-anal (receptive risk), and anal-penile (insertive risk). For HIV, only the last two are included, given the focus on MSM.

After changing these values, users will need to go to the *Results* worksheet and click on either *Calculate Baseline Model* or *Calculate Comparison Model* in order to incorporate these new parameters into calculations.

D11: *Advanced options*: infection prevalence among partners

This alternative option for specifying conversion factors to translate the number of diagnoses reported to the prevalence of infections makes most sense for users who have their own parameter; in this case, these parameters might come from a variety of different types of sources. From the *Table of Contents* worksheet, users can access this method by clicking on *Advanced Options: Infection Prevalence among Partners* or you can access it directly from the *Advanced Options* worksheet by scrolling down to the section on *Infection Prevalence among Partners*.

This section is split into three subsections, one for each infection. For each of them, converting the number of diagnoses reported to the prevalence of infections takes two steps. First, diagnoses must be converted to incidence using the estimated proportion of all cases that get diagnosed. For gonorrhea and chlamydia, these numbers should be inputted for females and males separately.

The second step is to convert incidence to prevalence. For gonorrhea and chlamydia, this is done by dividing incidence by the average duration of infection in years. These numbers are also provided for females and males separately. Note that these durations reflect the weighted average of duration for those who are diagnosed and treated and those who are not. This method does not work for HIV, which is a lifelong infection; the well-known approximate relationship *prevalence = incidence x duration* will fail when focusing on adolescents alone because HIV infection extends well beyond adolescence even for those infected as adolescents. We have instead provided a default number that reflects the ratio between incidence and prevalence observed among adolescents within national surveillance data (again, see Section G for source information).

D12: Additional worksheets

The additional worksheets include the full parameter worksheets for the Baseline Model and Comparison Models 1-4. Once the user has calculated results for these models, data will populate for the full parameters.

It is important to note that data should not be manually entered into cells on these worksheets.

For each model, parameters shown for each of the nine subgroups include:

- Population sizes
- Total number of acts and condom use during each sex type (receptive penile-vaginal sex, penile-vaginal sex, penile-anal sex, anal-penile sex) per year
- Prevalence of each STI (gonorrhea, chlamydia, and HIV) in partner pools
- Risk for each STI from condomless sex (for each sex type per year)
- Total acquisition risk for each STI
- Current prevalence, # cases, % diagnosed, and # diagnoses for each STI
- Transmission probabilities for each sex type for each STI are also shown in the boxes at the right

On these additional parameter worksheets, light gray boxes refer to inputted parameters, blue boxes are calculated parameters, and green boxes refer to the outcomes of interest.

The full parameters are summarized in the *to_export* worksheet, which is the spreadsheet version of the PDF export.

E: Example Scenario 1: User with jurisdiction-specific YRBS data using SAS

This example walks users through the use of the tool with national YRBS 2015 default parameters to explore scenarios involving varying increases in condom use within different age groups. Outputs from the SAS result viewer, Excel output from SAS, and inputs into the tool are shown below. Here we use a fictitious jurisdiction, which we name *Abigail County*.

1. Fill out the population size as outlined in Section D2, step 1. This entails finding the total number of high-school-attending adolescents through American Factfinder, and entering it in the section labeled Step 1:

Step 1: Please enter the total number of high-school attending adolescents (grades 9-12) in your jurisdiction.

350,713 | Set To Defaults | Reset Values

2. Find the weighted proportion of males and females by age group, as outlined in Section D2, step 2. Use the Percent column, and enter as a decimal. The SAS output will list each age group separately, while the Excel output from SAS will be formatted with all six age/sex combinations together and in the right order:

Table of Age group by Population by sex

Age group	Population by sex	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	95% Confidence Limits for Percent		Row Percent
13-15	Male	802	61224	5287	17.4569	1.0044	15.3278	19.5861	48.5595
	Female	981	64856	5470	18.4926	1.1083	16.1430	20.8422	51.4405
	Total	1783	126080	9851	35.9495	1.7165	32.3108	39.5883	100.0000

YRBS weighted proportion males and females by age group

	Population by sex	13-15	16-17	18+
1				
2				
3	Population by sex	13-15	16-17	18+
4	Male	0.1746	0.2425	0.0843
5	Female	0.1849	0.2439	0.0698

Age Group

	13-15	16-17	18+
Males	0.1746	0.2425	0.0843
Females	0.1849	0.2439	0.0698

3. Find the proportion of adolescents by sex and age who have ever had sexual intercourse into each of the three sexual partnership groups (MSM, MSF, FSM), as outlined in Section D2, step 3. Use the Row Percent column. Ensure you are using the table for the correct age group. Again, the SAS output will be split among multiple sections, while the Excel output from SAS will be compiled in order:

Table of Population by sex by Sexual partner groups including none

Controlling for Age group including none=13-15

Population by sex	Sexual partner groups including none	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	95% Confidence Limits for Percent		Row Percent
Male	MSM	10	535.38370	176.60575	0.4246	0.1349	0.1387	0.7106	0.8745
	MSF	147	11048	1380	8.7626	1.1810	6.2589	11.2663	18.0450
	FSM	0	.	.	.	.	.	.	.
	None of the above	645	49640	4952	39.3723	1.6733	35.8250	42.9196	81.0805
	Total	802	61224	5287	48.5595	1.7135	44.9270	52.1921	100.0000

1	YRBS proportion of sexually experienced adolescents by age group and SPG			
2				
3	Sexual partner groups	13-15	16-17	18+
4	MSM	0.0087	0.0325	0.0465
5	MSF	0.1805	0.339	0.5165
6	FSM	0.1604	0.4101	0.543

	Age Group 13-15	16-17	18+
Group 1: MSM	0.0087	0.0325	0.0465
Group 2: MSF	0.1805	0.339	0.5165
Group 3: FSM	0.1604	0.4101	0.543

4. Find the total number of adolescents in the jurisdiction in each age group, using the instructions in Section D2, Step 4. This entails using American Factfinder again. Enter the results into the cells labeled Step 4:

	Age Group 13-15	16-17	18	Total Pop
Males	122,845	86,943	50,772	
Females	128,549	87,351	41,494	
Total By Age	251,394	174,294	92,266	517,954

Click “Proceed to Next Step.”

5. Calculate jurisdiction-specific data on numbers of partners using the process outlined in D9. This requires jumping momentarily from the *Sexual behavior* worksheet to the *Advanced options* worksheet. Copy the appropriate data from the two matrices into the *New Partners per Year Calculator* cells, located below the *Alternative Population Size Calculator* on the *Advanced options* worksheet. Make sure to check the box marked "Important!" in the upper left-hand corner.

Matrix 1: Age at first sex by current age, cells=counts of persons

Sexual partner groups=MSM

	Current age	11	12	13	14	15	16	17
5	13	0	0					
6	14	42.01	47.17	25.81				
7	15	47.38	83.33	35.67	0			
8	16	531.74	112.51	0	254.24	440.17		
9	17	112.26	0	105.44	454.32	323.98	47.44	
10	18	470.94	76.17	0	61.49	367.53	332.31	65.77

Matrix 1: Age of First Intercourse by Current Age, Cells = Weighted Counts of Persons

	AGE	11	12	13	14	15	16	17
MSM	13	0.00	0.00					
	14	42.01	47.17	25.81				
	15	47.38	83.33	35.67	0.00			
	16	531.74	112.51	0.00	254.24	440.17		
	17	112.26	0.00	105.44	454.32	323.98	47.44	
	18	470.94	76.17	0.00	61.49	367.53	332.31	65.77
MSF	AGE	11	12	13	14	15	16	17
	13	0	0					
	14	0.00	260.54	379.91				
	15	650.99	1140.98	1142.25	3055.89			
	16	570.43	555.08	2400.20	2116.43	4052.31		
	17	825.42	1280.64	1310.11	1561.97	2641.76	5828.36	
	18	717.04	607.41	1087.53	1592.23	3103.09	3610.26	4434.41
FSM	AGE	11	12	13	14	15	16	17
	13	0	0					
	14	149.58	253.55	561.09				
	15	600.02	208.51	1278.65	2831.14			
	16	262.56	253.77	1025.76	3218.46	6148.76		
	17	176.77	310.32	947.00	3723.97	5632.98	7234.76	
	18	136.23	169.74	461.17	1829.61	3844.79	2611.98	4239.56

Matrix 2: Age of First Intercourse by Current Age, Cells = Average Number of Lifetime Partners

MSM AGE	11	12	13	14	15	16	17
13	0.00	0.00					
14	1.00	2.00	2.00				
15	6.00	1.00	4.00	0.00			
16	5.15	5.00	0.00	1.92	3.47		
17	3.83	0.00	1.00	3.81	3.70	1.00	
18	5.49	4.00	0.00	6.00	4.33	2.81	3.00

MSF AGE	11.00	12.00	13.00	14.00	15.00	16.00	17.00
13	0.00	0.00					
14	0.00	6.00	2.51				
15	2.58	4.61	3.00	2.31			
16	5.71	3.42	3.81	3.13	1.81		
17	4.95	5.55	3.97	3.29	2.50	1.90	
18	5.72	5.04	4.89	4.18	3.99	2.25	1.65

FSM AGE	11.00	12.00	13.00	14.00	15.00	16.00	17.00
13	0.00	0.00					
14	3.74	2.67	1.72				
15	4.11	2.02	3.52	2.10			
16	2.63	1.88	3.23	2.60	1.77		
17	2.80	5.13	4.62	3.69	2.89	1.81	
18	4.38	5.00	4.02	4.78	3.54	2.44	1.67

Notice that the matrix in blue below these on the Advanced Options worksheet ("Expected Number of New Partners per Year") will have filled in with new numbers. Confirm once again that you have checked the box at the top of this section ("Check this box to use the partner data…")

6. Return to the *Sexual behavior* worksheet. Find the probability of condom use per sex act by subgroup as described in Section D3: Step 3. Use the mean of the *Proc Logistic* estimate for each of the nine subgroups:

Estimate

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Mean	Standard Error of Mean
13-15 MSM	0.2585	0.3501	16	0.74	0.4710	0.5643	0.08608

1	**YRBS predicted condom use for each SPG and age group**			
2				
3	**Sexual partner groups**	**13-15**	**16-17**	**18+**
4	MSM	0.5643	0.463	0.4047
5	MSF	0.7876	0.7116	0.6606
6	FSM	0.6672	0.5717	0.5128

		Age Group		
		13-15	16-17	18+
Group 1: MSM	Vaginal-Penile	0.7876	0.7116	0.6606
	Penile-Anal	0.5643	0.4630	0.4047
Group 2: MSF	Vaginal-Penile	0.7876	0.7116	0.6606
Group 3: FSM	Penile-Vaginal	0.6672	0.5717	0.5128

Note: since there is little data on MSM's condom use with females specifically, we use a combined estimate for all males' use of condoms with females and apply it to both MSM's and MSF's contacts with females.

7. Now go to the *Diagnoses* worksheet to enter jurisdiction-specific diagnoses of *Chlamydia trachomatis*, *Neisseria gonorrhea*, and HIV as instructed. For gonorrhea and chlamydia, it is crucial that diagnosis information is properly scaled to your population (i.e., that you do not just accept defaults) since the absolute numbers at the level of the US population will be much larger than any local jurisdiction.

Neisseria gonorrhea

Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:

Diagnoses among MSM	121		Pop Size
		Total among males	485
Diagnoses among MSF	364	Total among females	865
Diagnoses among FSM	865		

Chlamydia trachomatis

Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:

Diagnoses among MSM	119		Pop Size
		Total among males	1,328
Diagnoses among MSF	1,209	Total among females	5,398
Diagnoses among FSM	5,398		

HIV

Number of diagnoses per 12-month period among 13-18 year-old MSM in or out of school:

29

Set to Defaults | Res et Values

8. Skip to results and click “Calculate Baseline Model”:

Calculate Baseline Model / Reset Model		Gonorrhea	Chlamydia	HIV
Calculate Baseline Model	Expected # of in-school incident cases in next year	1,928	12,654	45
Reset Model	Expected # of in-school diagnoses in next year	985	4,859	25

9. Now, return to the *Interventions* worksheet to implement an intervention. In this example, condom use is increased differently in different age and partnering groups. Thus, make sure to clear the “across all subgroups” box.

By what percent should condom use be increased? (Note: please use decimal format. For example, 5% should be entered as 0.05)

Across all subgroups: [ ]

WARNING:
The intervention will default to the "across all subgroups" setting unless the input box has been cleared

OR

Within each subgroup:

Age Group	13-15	16-17	18+
Increase among MSM	0.15	0.1	0.05
Increase among MSF	0.1	0.075	0.025
Increase among FSM	0.1	0.075	0.025

Reset Values

10. Click “Proceed to Results,” then click the first box down labeled “Click Comparison Model.” Name your intervention.

11.

	Gonorrhea	Chlamydia	HIV
Expected # of in-school incident cases in next year	1,757	11,542	43
Expected # of in-school diagnoses in next year	898	4,467	24
Incident in-school cases averted from baseline	172	1,112	2
% of in-school cases averted from baseline	8.9%	8.8%	4.9%

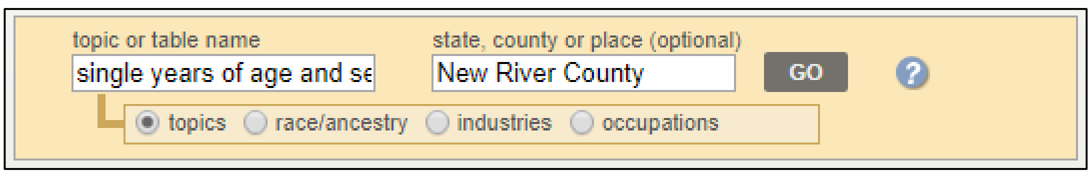
The results shown indicate the number and percent of incident cases that would be expected to be averted as a result of the intervention.

F: Example Scenario 2: User without jurisdiction-specific YRBS data

This example demonstrates how to use the tool if you do not have access to YRBS data for your jurisdiction. Using publicly available data, the tool can still be used. In this example, we will be finding the impact an intervention that decreases the frequency of sex acts, using the fictional jurisdiction of *New River County*.

1. Choose the *Advanced Options* worksheet. Make sure you are looking at the *Alternative Population Size Calculator* section at the top of the worksheet. Check the box in the upper left-hand corner marked "Important!"

2. Find the population size for your jurisdiction on American Fact Finder (factfinder.census.gov), following the instructions in Section D8. Click on Advanced Search → Show Me All. Search for topic and table name "age and sex" and then enter your jurisdiction.



3. Select the most recently available relevant data containing numbers by single year of age and sex – in this case, it is the 2017 Population Estimates by Single Year of Age and Sex (PEPSYASEX)

	ID	Table, File or Document Title	Dataset	About
	PEPSYASEX	Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States, States, and Puerto Rico Commonwealth: April 1, 2010 to July 1, 2017	2017 Population Estimates	

4. Find the number of males and females in each of the tool's three age groups. Users may need to manually add cells to find totals for the tool's age groups. Enter this information into the white cells in Step 1 of the *Alternative Population Size Calculator* on the *Advanced Options* worksheet.

	Age Group			
	13-15	16-17	18	
Males	20,318	13,771	7,113	
Females	19,689	13,388	6,862	
				Total
Total	40,007	27,159	13,975	81,141

5. The New River County Health Department does not have jurisdiction-specific information to inform local estimates of the “proportion of adolescents who attend high school by sex and age group”, so we keep the defaults for Step 2.

6. However, they do have data on the proportion of adolescents who have ever had sexual intercourse, so we enter the jurisdiction-specific information in Step 3.

	Age Group		
	13-15	16-17	18+
MSM	0.01376	0.0221	0.02772
MSF	0.23489	0.43733	0.52521
FSM	0.19341	0.43292	0.55293

7. Proceed to the *Sexual behavior* worksheet. New River County estimates that condom use in MSM is 5% lower than the national average. We thus calculate new rates for the penile-anal section of the *Sexual Behavior* worksheet, by multiplying the default numbers by 0.95. There is no information on condom use for vaginal-penile sex, so we leave the national defaults in all other cells.

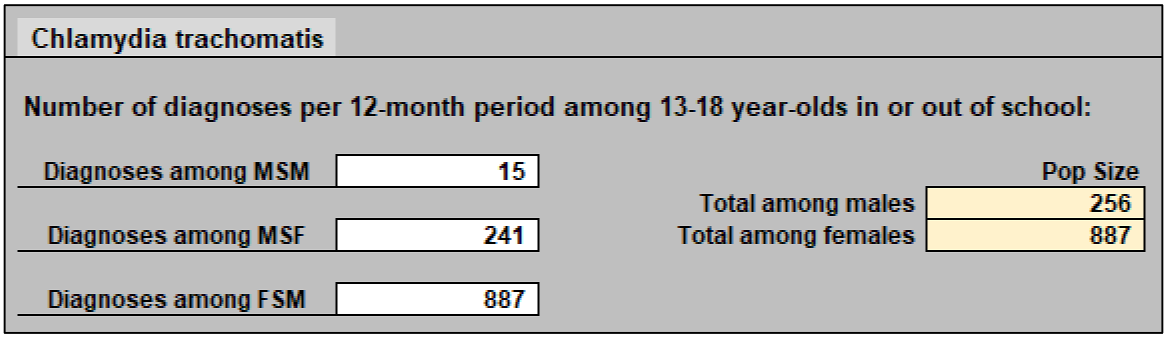
		Age Group		
		13-15	16-17	18+
Group 1: MSM	Vaginal-Penile	0.6930	0.6526	0.6091
	Penile-Anal	0.5196	0.4760	0.4317
Group 2: MSF	Vaginal-Penile	0.6930	0.6526	0.6091
Group 3: FSM	Penile-Vaginal	0.6113	0.5669	0.5204

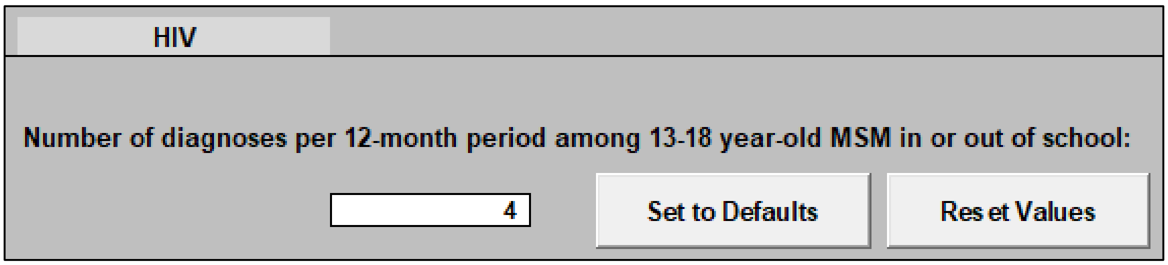
8. Now go to the ‘Diagnoses’ worksheet to enter jurisdiction-specific diagnoses of *Chlamydia trachomatis*, *Neisseria gonorrhea*, and HIV as instructed. It is crucial that diagnosis information is properly scaled to your population.

Neisseria gonorrhea

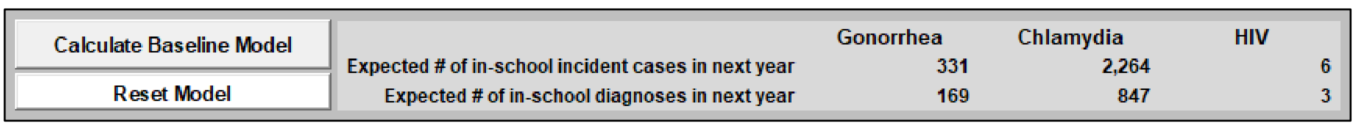
Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:

Diagnoses among MSM	16		Pop Size
		Total among males	87
Diagnoses among MSF	71	Total among females	143
Diagnoses among FSM	143		

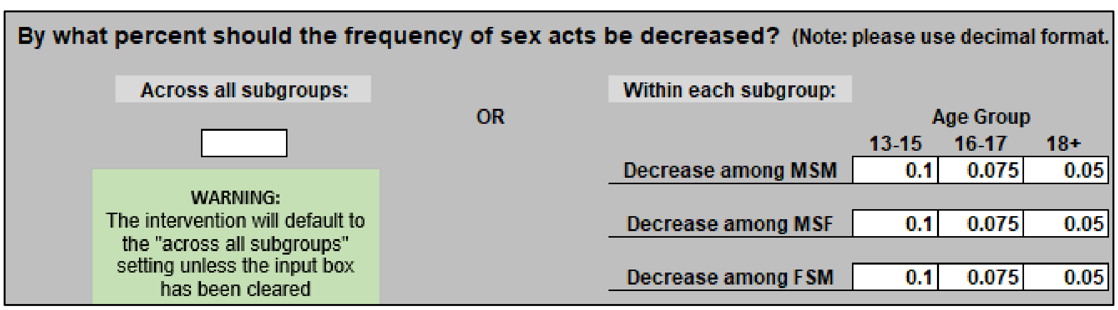




9. Go to the *Results* worksheet and click "Calculate Baseline"



10. Now we can model behavior change in the *Interventions* worksheet. In this example, we model a differential decrease in sex acts for different age groups. We will also implement a 10% increase in condom use for MSF/FSM, and a 20% increase in condom use for MSM adolescents:



By what percent should condom use be increased? (Note: please use decimal format. For example, 5% s

Across all subgroups:	OR	Within each subgroup:	Age Group 13-15	16-17	18+
		Increase among MSM	0.2	0.2	0.2
		Increase among MSF	0.1	0.1	0.1
		Increase among FSM	0.1	0.1	0.1

WARNING:
The intervention will default to the "across all subgroups" setting unless the input box has been cleared

11. Proceed to results, and "Calculate Comparison Model".

	Gonorrhea	Chlamydia	HIV
Expected # of in-school incident cases in next year	264	1,835	5
Expected # of in-school diagnoses in next year	135	693	3
Incident in-school cases averted from baseline	67	429	1
% of in-school cases averted from baseline	20.2%	18.9%	16.5%

12. The model now shows the number and percent of incident cases averted from baseline. As a reminder, cases are shown rounded to the nearest whole number, while % averted is calculated on the unrounded numbers.

G. Default parameter documentation

In this section we document the sources and derivations for the default parametrizations supplied with the tool. We proceed in order by worksheet.

G1. *Population* worksheet

The total number of high-school-attending students in the US (16,985,786) was obtained from the 2012-2016 American Community Survey 5-Year Estimates, using the steps outlined above in Section D2, part 1.

The proportions of high school students by age and sex were obtained from national YRBS data, using the instructions laid out above in Section D2, part 2.

The proportions of adolescents who have ever had sexual intercourse were also derived from national YRBS data, using the instructions laid out above in Section D2, part 3. Note that the numbers here are identical to those in step 3 of the "Alternative population size calculator" on the *Advanced Options* worksheet.

The total adolescent population size by age group (regardless of whether in school) was derived from the US Census national data, using the instructions laid out above in Section D2, step 4. Note that the numbers here are identical to those in step 1 of the "Alternative population size calculator" on the *Advanced Options* worksheet.

G2. *Sexual behavior* worksheet

Average numbers of new partners per year is the parameter set that requires by far the most amount of data manipulation and back-calculation from commonly available data sets. (From the users' point of view, these calculations are mostly taken care of behind the scenes; however, here we explain what those steps do so that users can interpret the default values provided with the tool.

These default values are derived from national YRBS data from the question "During your life, with how many people have you had sexual intercourse?". Our method began by partitioning respondents out according to their response to other questions: their current age, and their age at first intercourse. For each combination of these two values (e.g., current age 17, first intercourse at 15), we tabulated the number of people with that combination and the mean number of lifetime partners for those with that combination. We do so separately for each of the three sex partnering groups. The SAS code to generate these matrices is provided in Section D9.

We then convert these numbers into estimates for the annual number of new partners by age and sex partnering group. The calculations for this occur on the *ppy_calc* worksheet of the tool; users do not normally need to look at this worksheet, since the calculations will occur automatically when the user enters their data in the New Partners per Year calculator on the *Advanced Options* worksheet. In summary: we assume that partners are spread out evenly over the time between first intercourse and current age. (For example, someone who is currently 16 and first had sexual intercourse at 14 and has had 4 partners will be assumed to have averaged 2 new partners per year. This is because their time since first sex is estimated at 2 years, and 4 partners / 2 years = 2 new partners per year. We then average the lifetime partner numbers adolescents have by a given age, weighted across all of those contributing data to that estimate. (For example, for the estimate of partners per year among 15 year olds, data will be contributed by everyone who is currently 16 or older and who first had sexual intercourse at age 15 or earlier). We then subtract sequential estimates of lifetime partners (e.g., lifetime partners at 16 – lifetime partners at 15) numbers to get new partners per year by age. Note that in doing these calculations, we ignore those who report age at first sex is the same age as their current age, since it is difficult to calculate an expected annual rate under this condition. A variety of exploratory calculations indicated that the overall numbers are highly insensitive to this exclusion. For more information, see the Technical Note at the end of Section D9.

Mean number of sex acts per partner. These numbers cannot be derived from YRBS, which lacks information on sex frequency. Instead, we derived these numbers from the National Survey of Family Growth (NSFG). We used questions on respondent's reported age, sex and sex of partners to assign them to our nine subgroups. We then used the answer to the question "how many times have you had sexual intercourse in the last four weeks?" and rescaled this number up to a year. We then divided by our estimate of the number of new partners per year (derived above) to obtain an estimate for the number of acts per partner.

Probability of condom use. This parameter is derived directly from the national YRBS data using the question "*The last time you had sexual intercourse, did you or your partner use a condom?*" and calculating the proportion of valid responses (*yes* or *no*) that were *yes*, for each subgroup. SAS code for this is available in Section D3 step 3.

G3. *Diagnoses* worksheet

Gonorrhea and chlamydia: Estimates here take the form of the total number of diagnoses by sexual partnering group. Our estimates begin with sex-specific rates from the CDC's 2016 STD Surveillance Report [4] for adolescents, taken from Figures 5 for chlamydia and Figure 17 for gonorrhea:

		Reported rates per 100k	
Age	Sex	CT	GC
10-14 year old	Females	91.1	19.1
15-19 year old	Females	3,070.9	482.1
10-14 year old	Males	12.7	4.7
15-19 year old	Males	832.6	280.8

We then multiplied these rates by the population sizes for 13-14 years old and 15-18 years by sex from the US Census Bureau's 2016 estimates:

Ages	Males	Females
13-14	4,182,575	4,020,859
15-18	8,632,209	8,254,310

and summed across ages. This provided us with our final number for FSM.

For MSM and MSW, we needed to partition the male cases according to these two sexual partnering groups. One option we considered was using data in Figures 11 and 24 from the same STI Surveillance Report, [4] which report on the percent of MSM vs MSW who test positive for each STI at a set of jurisdictions across the country. However, this does not account for any difference in the rates at which these two groups appear at clinics in the first place; a higher infection rate for one group should show up in large part in higher rates of seeking screening, especially for an STI like gonorrhea that often presents acute symptoms in males. Another was to use the ratio of gonorrhea rates estimated for adult MSM and MSW (e.g., Figure BB of the same report). However, this ratio may differ substantially between adults and adolescents, especially as patterns of having ever had sexual intercourse differ between the two groups. Lacking explicit data at the national level assigning STI cases among adolescent males by sexual partnering group, we turned to data from a single large jurisdiction that does include this information (New York State), provided courtesy of the NY State Department of Health AIDS Institute. Here, the number of diagnosed cases of chlamydia among 13-18-year MSW was 16.3 times higher than MSM (recall these are absolute numbers, not rates). For gonorrhea, the same ratio was 4.8. We calculated the numbers of cases in each group needed to obtain these ratios with the total numbers for adolescent males calculated in the previous step above.

HIV

For HIV we used the CDC's 2016 HIV Surveillance Report.[5] Table 5a lists new diagnoses for MSM by age category. We used 2015 numbers for consistency with other parameters. The "total" table lists 10 and 1,375 new diagnoses for MSM aged 13-14 and 15-19, respectively, and 0 and 43 for the MSM/IDU in the same two age groups. We needed to convert the 15-19 numbers into estimates for just 15-18. We assumed that there were be roughly a linear increase across ages, starting from effectively none below 15, such that the five ages 15 through 19 would contribute 1/15, 2/15, 3/15, 4/15, and 5/15 of the diagnoses, respectively. Removing the 19-year-olds thus means reducing the combined (MSM and MSM/IDU) 15-19 number (1,375 + 43 = 1,418) to 10/15 of its initial value, or to 945. Adding back in the ten 13-14-year-old diagnoses yields 955.

G4. *Advanced options*: Alternative population size calculator
The *total adolescent population size by age and sex group* (regardless of whether in school) was derived from the US Census national data, using the instructions laid out above in Section D2, step 4. Note that the numbers here are identical to those in step 4 of the main *Population* worksheet.

The *proportion attending high school by sex and age group* was derived by using three sets of numbers from the YRBS and Census, all of which appear on the *Population sizes* Worksheet: Step 1 (total high school population) * Step 2 (proportion of HS students by age and sex group) / Step 4 (total number of adolescents regardless of school attendance). Note that using other data sources that measure this number directly (for example the Census Bureau's Current Population Survey, taken in October), yield numbers that are different and inconsistent with the numbers used elsewhere in this tool, since the age distribution of the population in high school changes throughout the school year.

The proportions of adolescents who have ever had sexual intercourse were also derived from national YRBS data, using the instructions laid out above in Section D2, part 3. Note that the numbers here are identical to those in step 3 on the *Population* worksheet.

G5. *Advanced options*: New partners per year calculator
For the two matrices entered here, see *Average numbers of new partners per year* entry in Section G2, as well as the code in Section D9.

Also included in this section is the measure for the percent of MSM's partners that are female. This was derived from an analysis of NSFG data. Males were categorized as MSM if they reported lifetime sexual intercourse with another male. Respondents were asked for both the lifetime number of male partners and lifetime number of female partners, allowing for an estimate of the average proportion of all partners that were female.

G6*. Advanced options*: transmissibility
Per-act transmission probabilities were used as "tuning parameters" in the model, i.e., set to values that yielded the same number of expected incident infections in the coming year as were in the inputs, using default values for all other parameters. For a full description of this process, see the *Section H: Technical note: model calibration*.

However, we still wished to compare our resulting values to estimates drawn from the literature, to determine if our values lay within a reasonable range or were above or below that range, and if so, why. These results are also discussed in Section H. Here we explain the literature review that we conducted, and its results:

Chlamydia trachomatis and Neisseria gonorrhea

We first conducted a literature review in PubMed for articles estimating penile-vaginal and penile-anal transmissibility risk of the two pathogens empirically or through model calibration, and which estimates are cited and used in the current modeling literature. We used the following keywords used in various combinations: Chlamydia trachomatis, Neisseria gonorrhea, transmission, transmissibility, acquisition risk, infection, model, MSM. Additionally, any papers that were referenced in the papers found by this search (but not themselves included in the search results) and that estimated or used transmissibility parameters were also evaluated. Most commonly, papers were excluded because they estimated per-partnership transmission risk or yearly acquisition risk, where we were interested in the per-contact risk of transmission.

Our final set included 23 papers that had estimates for CT transmissibility, [6-28] and 19 papers that had estimates for GC transmissibility.[8, 10-12, 14, 15, 19, 23, 26, 29-38] Nine papers considered both diseases, and thus were included in both analyses. All papers are listed in *Section J: Citations*. We recorded four parameters of interest for each infection: insertive penile-vaginal transmissibility (the per-act probability that the female partner infects the male), receptive penile-vaginal transmissibility (the per-act probability that the male infects the female), insertive penile-anal transmissibility (the per-act probability that the receptive male infects the insertive male) and receptive penile-anal transmissibility (the per-act probability that the insertive male infects the receptive male). For each paper, we also made note of what, if any, parameters were used for duration of each infection, incubation period, and the percent of cases that are asymptomatic.

The median and range of values obtained from this literature review are:

	Chlamydia		Gonorrhea	
	Median	Range	Median	Range
Female to male (ins. penile-vaginal)	0.11	0.0375 – 0.45	0.25	0.09 – 0.6
Male to female (rec. penile-vaginal)	0.12	0.049 – 0.45	0.50	0.2 – 0.8
Insertive penile-anal	0.17	0.1 – 0.32	0.35	0.2 – 0.5
Receptive penile-anal	0.17	0.1 – 0.4	0.50	0.25 – 0.6

For HIV, the process was much simpler, since a recent literature review and meta-analysis already conducted an analogous process for HIV with both means and 95% confidence intervals.[39] Estimates were drawn from Table 1 of that paper:

	Mean (per 10k exposures)	95% CI (per 10k exposures)
Insertive penile-anal	11	4-28
Receptive penile-anal	138	102-186

G7. Advanced options: **Infection prevalence among partners**

Estimating infection prevalence among partners begins with the diagnoses numbers and rates described in Section G3. For gonorrhea and chlamydia, we then convert this using data on *proportion of infections diagnosed* and *average duration of infection*, as explained in Section B, under *Probability partner is infected*. For HIV, we used the *proportion aware of their status* and *ratio of prevalence to incidence*, described in the same section. Here we explain the sources of these default parameters:

Gonorrhea and chlamydia: proportion of infections diagnosed

Our prime source here is Owusu-Edusei et al. 2013.[40] In the absence of clear estimates for proportion of infections diagnosed for the relevant age groups, we instead use their estimated for proportion treated as an approximate value. We calculate the overall proportion as a weighted average across those symptomatic and asymptomatic, using their numbers in the first table of the Supplement, which are in most cases similar or identical to those from the source manuscript Satterwhite et al. 2008.[41]

Gonorrhea and chlamydia: mean duration of infection

We derive these sex-specific numbers from Satterwhite et al.,[41] using their numbers for 15-24 year olds (pp. 188-9). Their method takes weighted averages across both symptomatic and asymptomatic cases, incorporating the probability that each of those case types get diagnosed and treated.

HIV: proportion aware of their infection

For this we turn to the CDC's 2015 HIV Monitoring Supplemental Report, Table 9b. We focus on the percent of infections diagnosed for all 13-24 year-olds (55.6%), since this is the definition closest to our target population (13-18 year-old MSM) available through an intense search of the published and gray literatures. We note that, although it is not ideal, the effects of including a wider population definition are somewhat mitigated since this is a ratio measure rather than an absolute measure; that is, the more expansive population definition occurs in both the numerator and denominator. Moreover, we would expect the two forms of bias we have to partly cancel out; i.e., we expect our target age range (13-18) to have a lower rate of diagnosis than the additional ages included (19-24), but our subgroup of interest (MSM) to have higher rates of diagnoses than the other subgroup included (given greater awareness of HIV risk and more frequent testing among this subgroup).

HIV: ratio of prevalence to incidence

Because HIV is a lifelong infection, and most years lived with HIV will occur outside of adolescence, even for those infected as adolescents, we cannot use a method based on duration analogous to that used for gonorrhea and chlamydia. Instead, we estimated an empirical ratio of prevalence to incidence among an adolescent MSM population, as:

(# adolescents 13-19 living with an HIV diagnosis) * (proportion of adolescent diagnoses 13-19 that are among MSM) / new diagnoses among MSM 13-19

We include 13-19 and diagnoses rather than incident and prevalent cases because of data availability. Converting from diagnoses to all cases should involve the same conversion factor in both the numerator and the denominator, meaning the ratio is unchanged. Similarly, including 19-year-olds in both the numerator and denominator should have a similarly minor impact on the final ratio. To obtain these numbers we again turn to the 2016 CDC HIV Surveillance Report,[5] which gives:

- # adolescents 13-19 living with an HIV diagnosis, 2015: 5,680 (Table 20a, = 721+4,959)
- % of adolescent 13-19 diagnoses that are among MSM = 84% (numerator from Table 5a = 10+1375+0+43); denominator from comparable numbers summed over all risk groups in Tables 5a and 7a = 1,696)
- new diagnoses among MSM 13-19: 1,428 (Table 5a, 10+1375+0+43)

Putting these pieces together yields: 5,680 * 84% / 1,428 = 3.34.

H. Technical note: Model calibration

In order for a model to yield meaningful predictions about the impact of any changes (behavioral or otherwise) on infectious disease burden, the model must first produce realistic outcomes in the absence of those changes. The process of ensuring that the model does so is called *model calibration*.

We elected to calibrate our model by making sure that, using all our data-derived inputs and no interventions, it produced incident cases in the coming year that were similar to those inputted for the previous year. We did this using national-level data. We elected to separately calibrate each of the three infections for each of the sexual partnering groups modeled for that infection (MSM, MSF and FSW for gonorrhea and chlamydia; MSM for HIV).

One wrinkle is that our diagnosis inputs ask for diagnoses among all 13-18 year olds whether in or out of high school, while our main model output yields diagnoses among 13-18 year olds in school only. This is because the former represents the data that most jurisdictions are likely to have access to; while the latter is more justified for our outputs since behavioral parameters mostly come from YRBS, which is among high-school-attending adolescents only. To get around this, we produce a single additional set of outputs for the base model only, which reflect estimated diagnoses regardless of school status collapsed across age groups; these appear at the far right of the *Results* worksheet, under the label "Calibration – all diagnosed cases". Most users will not need to use these numbers.

The parameters we chose to "tune," i.e., to change until the model yielded calibrated results, were the transmission probabilities per act. We did so for a few reasons: one is that there is an enormous amount of uncertainty around many of these in the literature, especially for gonorrhea and chlamydia. Another is that these are the only set of parameters in the model that we expect to be similar across diverse jurisdictions. This is in contrast to parameters around sexual behavior, existing diagnoses, and infection duration (which, although partly a function of basic pathogen biology, is also a function of treatment rates). Another is that we do not expect users to have their own inputs for these. Finally, because these parameters are specific to each pathogen (unlike sexual behavior), and to each routes (vaginal vs. anal, insertive vs. receptive), they allow us to neatly calibrate each of the different infections and sexual partnering groups.

We identified through back calculation the unique set of transmission probabilities that yielded calibrated results at the national level. We then compared these probabilities to the range that we had identified from the published literature.

For gonorrhea and chlamydia, all final parameter values lay within the published range for the act type. Those for MSM (anal sex) tended to be relatively close to the median of values we found in the literature, while those most relevant for MSF (insertive penile-vaginal sex) tended to be on the lower end of the range and those for FSM (receptive vaginal-penile sex) tended to be on the high end. We expect this is in large part because of an assumption we made in our model that was necessary given available data, but does not fully reflect reality. In the model, we assumed that adolescents are mostly partnering with people of like ages to themselves; this is because, when deciding on the probability that a person's partner is infected, we used the prevalence of the infection in the relevant sexual partnering group in the model, which reflects prevalence among 13-18 year olds. In reality, there is often asymmetric age matching among opposite-sex partnerships within this age range, with male partners slightly older than females. Since prevalence generally increases with age across the 13-18 age range, this means our method is overestimating actual prevalence in adolescent MSF's partners, and underestimating it for adolescent FSM's partners. This would explain the pattern in our calibrated transmission probabilities relative to the medians in the published literature. Given this, it may be useful for users to think of our transmission parameters as composites, reflecting both the true transmission probability per act for that pathogen and sexual act type, as well as a separate component that reflects the unmeasured ratio between expected and actual prevalence of infection for the sexual partners of the relevant group.

For HIV, we needed to increase our numbers considerably, by a factor of roughly 4.7. Some of this may be because young people are more likely than others to have partners who are also young, and thus recently infected, and thus potentially more infectious. An additional factor is likely to be that here age mixing is especially important – since HIV is incurable, its prevalence increases sharply with age, especially over this age range. Even a small number of older partners could throw off the estimates of HIV prevalence by a fair bit, given the low rates among adolescents. Once again, then, we suggest that these parameters be thought of as composites, reflecting both true transmission probability per act and the unmeasured ratio between expected and actual prevalence of infection among sexual partners.

We expect that most users will benefit from accepting our calibrated numbers, and estimating their intervention effect sizes without changing these. However, advanced users might opt to do so, to either calibrate their local model more precisely or to explore the implications of different infection probabilities. They can do so by changing the transmission probabilities in the *Transmissibility* section of the *Advanced Options* worksheet, and rerunning the base model with their jurisdiction-specific data until the numbers in the Calibration section of the *Results* worksheet match the diagnoses input on the *Diagnoses* worksheet. Note that, because MSM engage in both anal sex with other MSM and vaginal sex with FSM, it is easiest to first calibrate the parameters for vaginal sex by matching the MSF and FSM diagnoses, and then calibrating the parameters for anal sex by matching the MSM diagnoses. In our approach, we assumed a constant ratio between insertive and receptive anal sex, increasing or decreasing each by the same factor.

Users of models often wonder: why build the model in the first place, if one is then going to change some aspects of it until it matches what is expected for output? Doesn't that mean the model just tells us what we already knew? The answer is that this is only true for the base case, where there are no behavioral interventions or other changes. Once we introduce those, we begin to obtain results that we could not necessarily have predicted without the model. The calibration helps us to make sure that the model is on target in certain key features, so that we can then interpret those predictions about change more clearly.

I. Technical note: Final equations

We can use the following notation to represent the various quantities that are either input or calculated from inputs for any subgroup. These can be indexed to distinguish the values among subgroups, but we present simply a general version here for visual simplicity:

s = number of susceptibles
r = number of new partners per year
α = number of acts per partner
π = infection prevalence among partners
c = probability of using condoms per act
β = probability of transmission per condomless discordant contact

To calculate the number of new infections for a subgroup, we use the standard methods of considering the probability a person does *not* get infected over any of the acts they have with any of their partners, and then subtract this from 1 to get the probability they are infected. We do so somewhat differently for the two curable bacterial STIs than for HIV.

Gonorrhea and chlamydia

Here our expression for new infections takes the form

$$s\left(1-(1-\pi\beta)^{\alpha r(1-c)}\right)$$

For these two infections, there is a reasonable possibility that a susceptible person's partner's status may change over the course of the partnership, either from a new infection or from natural clearance or treatment. Thus, we wish to consider the acts with this partner as separate possible exposures. The expression πβ represents the probability of acquiring an infection during one act of condomless intercourse, and $(1-\pi\beta)$ thus represents the probability of avoiding infection during one such act. There are $\alpha r(1-c)$ such acts, and avoiding infection in all of them has probability $(1-\pi\beta)^{\alpha r(1-c)}$. Subtracting this from 1 gives the probability of acquiring infection across all acts for an average susceptible, and multiplying this by the number of susceptibles gives the number of new infections.

This expression is an approximation based on the static nature of the model; a full dynamic model would be needed to consider all of the dynamics of transmission within relationships, and the likelihood that infection appears or clears for one partner or another over the course of each relationship.

HIV

Here our expression for new infections takes the form

$$s\left(1-\left((1-\pi)+\pi(1-\beta)^{n(1-c)}\right)^{r}\right)$$

This is more complicated than the expression for gonorrhea and chlamydia because it recognizes that, with lower overall incidence and without a cure, the HIV status of a susceptible person's partner is less likely to change over the course of the relationship than for the other infections. Thus, we evaluate each relationship status once to see if the partner is infected or not, and then, if so, consider the probability of transmission over the subsequent acts. We do so separately for each partner. There are two ways to avoid transmission from a given partner: for them not to be infected, or for them to be infected but for transmission not to happen regardless. The probability of the former is $(1-\pi)$ and of the latter $\pi(1-\beta)^{n(1-c)}$. Adding these together gives the probability of no transmission from a single partner, and raising this to the power of *r* does so for *r* partners. Subtracting from one thus gives the probability of acquiring infection from any of *r* partners; multiplying this by *s* gives the number of new infections.

Again, this is an approximation of a more complex dynamic process for a static model.

J. Citations

1. Brown RT, Lossick JG, Mosure DJ, Smeltzer MP, Cromer BA. Pharyngeal gonorrhea screening in adolescents: is it necessary? *Pediatrics* 1989; **84**(4): 623-5.

2. Roochvarg LB, Lovchik JC. Screening for pharyngeal gonorrhea in adolescents. A reexamination. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 1991; **12**(3): 269-72.

3. Marrazzo JM. Barriers to infectious disease care among lesbians. *Emerging infectious diseases* 2004; **10**(11): 1974-8.

4. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2016. In: https://www.cdc.gov/std/stats16/CDC_2016_STDS_Report-for508WebSep21_2017_1644.pdf USDoHaHS, editor. Atlanta; 2017.

5. Centers for Disease Control and Prevention. HIV Surveillance Report, vol. 28: Diagnoses of HIV Infection in the United States and Dependent Areas, 2016. http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. 2017.

6. Andersen B, Gundgaard J, Kretzschmar M, Olsen J, Welte R, Øster-Gaard L. Prediction of costs, effectiveness, and disease control of a population-based program using home sampling for diagnosis of urogenital Chlamydia trachomatis infections. *Sexually Transmitted Diseases* 2006; **33**: 407-15.

7. Beck EC, Birkett M, Armbruster B, Mustanski B. A Data-Driven Simulation of HIV Spread Among Young Men Who Have Sex With Men. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2015; **70**: 186-94.

8. Bracher M, Santow G, Watkins SC. "Moving" and marrying: Modeling HIV infection among newly-weds in Malawi. *Demographic Research* 2003; **9**: 207-46.

9. Gray Richard T, Beagley Kenneth W, Timms P, Wilson David P. Modeling the Impact of Potential Vaccines on Epidemics of Sexually Transmitted <i>Chlamydia trachomatis</i> Infection. *The Journal of Infectious Diseases* 2009; **199**: 1680-8.

10. Hui BB, Gray RT, Wilson DP, et al. Population movement can sustain STI prevalence in remote Australian indigenous communities. *BMC infectious diseases* 2013; **13**: 188.

11. Johnson LF, Alkema L, Dorrington RE. A Bayesian approach to uncertainty analysis of sexually transmitted infection models. *Sexually Transmitted Infections* 2010; **86**: 169-74.

12. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The role of sexually transmitted infections in the evolution of the South African HIV epidemic. *Tropical Medicine & International Health* 2012; **17**: 161-8.

13. Katz BP. Estimating Transmission Probabilities for Chlamydial Infection. *Statistics in Medicine* 1992; **11**: 565-77.

14. Korenromp EL, Van Vliet C, Grosskurth H, et al. Model-based evaluation of single-round mass treatment of sexually transmitted diseases for HIV control in a rural African population. *AIDS (London, England)* 2000; **14**: 573-93.

15. Kretzschmar M, Van Duynhoven YTHP, Severijnen AJ. Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations. *American Journal of Epidemiology* 1996; **144**: 306-17.

16. Kretzschmar M, Welte R, van den Hoek A, Postma MJ. Comparative Model-based Analysis of Screening Programs for Chlamydia trachomatis Infections. *American Journal of Epidemiology* 2001; **153**: 90-101.

17. Low N, McCarthy A, Macleod J, et al. Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection. *Health technology assessment (Winchester, England)* 2007; **11**.

18. Lycke EL, G B ; Hallhagen, G ; Johannisson, G ; Ramstedt, K. The risk of transmission of genital Chlamydia trachomatis infection is less than that of genital Neisseria gonorrhoeae infection. *Sexually transmitted diseases* 1980; **7**: 6-10.

19. Orroth KK, Freeman EE, Bakker R, et al. Understanding the differences between contrasting HIV epidemics in east and west Africa: results from a simulation model of the Four Cities Study. *Sexually transmitted infections* 2007; **83 Suppl 1**: i5-16.

20. Regan David G, Wilson David P, Hocking Jane S. Coverage Is the Key for Effective Screening of <i>Chlamydia trachomatis</i> in Australia. *The Journal of Infectious Diseases* 2008; **198**: 349-58.

21. Ruijs GJ, Schut IK, Schirm J, SchrÃ¶der FP. Prevalence, incidence, and risk of acquiring urogenital gonococcal or chlamydial infection in prostitutes working in brothels. *Genitourinary Medicine* 1988; **64**: 49.

22. Severijnen A, M.v.d. Laar, and J. Ossewaard. Infecties met chlamydia trachomatis. In: RIVM MvdLB, editor. Seksueel overdraagbare aandoeningen in Nederland; 1993.

23. Stigum H, Magnus P, Veierød M, Bakketeig LS. Impact on sexually transmitted disease spread of increased condom use by young females, 1987-1992. *International Journal of Epidemiology* 1995; **24**: 813-20.

24. Tu W, Ghosh P, Katz BP. A stochastic model for assessing Chlamydia trachomatis transmission risk by using longitudinal observational data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2011; **174**: 975-89.

25. Turner KM, Adams EJ, Gay N, Ghani AC, Mercer C, Edmunds WJ. Developing a realistic sexual network model of chlamydia transmission in Britain. *Sexually Transmitted Infections* 2004; **80**: 354-62.

26. Vickerman P, Watts C, Alary M, Mabey D, Peeling RW. Sensitivity requirements for the point of care diagnosis of Chlamydia trachomatis and Neisseria gonorrhoeae in women. *Sexually transmitted infections* 2003; **79**: 363-7.

27. Vriend HJ, Lugnér AK, Xiridou M, et al. Sexually transmitted infections screening at HIV treatment centers for MSM can be cost-effective. *Aids* 2013; **27**: 2281-90.

28. Xiridou M, Vriend HJ, Lugner AK, et al. Modelling the impact of chlamydia screening on the transmission of HIV among men who have sex with men. *BMC Infectious Diseases* 2013; **13**: 436.

29. Chen MI, Ghani AC, Edmunds WJ. A metapopulation modelling framework for gonorrhoea and other sexually transmitted infections in heterosexual populations. *Journal of the Royal Society Interface* 2009; **6**: 775-91.

30. Garnett GP, Mertz KJ, Finelli L, Levine WC, St Louis ME. The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from Newark, New Jersey. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 1999; **354**: 787-97.

31. Hethcote HW, Yorke JA. Gonorrhea Transmission Dynamics and Control. 1984; **56**.

32. Holmes K, Johnson D, Trostle H. An estimate of the risk of men aquiring gonorrhea by sexual contact with infected females. *American Journal Epidemiology* 1970; **91**: 170.

33. HOOPER RR, REYNOLDS GH, JONES OG, et al. Cohort Study of Venereal Disease. I: The Risk of Gonorrhea Transmission From Infected Women to Men. *American Journal of Epidemiology* 1978; **108**: 136-44.

34. Marseille E, Kahn JG, Billinghurst K, Saba J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Social Science & Medicine* 2001; **52**: 135-48.

35. Morin BR, Medina-Rios L, Camacho ET, Castillo-Chavez C. Static behavioral effects on gonorrhea transmission dynamics in a MSM population. *Journal of Theoretical Biology* 2010; **267**: 35-40.

36. Platt R, Rice PA, McCormack WM. Risk of acquiring gonorrhea and prevalence of abnormal adnexal findings among women recently exposed to gonorrhea. *JAMA: The Journal of the American Medical Association* 1983; **250**: 3205-9.

37. Xiridou M, Lugnér A, De Vries HJC, et al. Cost-Effectiveness of Dual Antimicrobial Therapy for Gonococcal Infections Among Men Who Have Sex With Men in the Netherlands. *Sexually Transmitted Diseases* 2016; **43**: 542-8.

38. XIRIDOU M, SOETENS LC, KOEDIJK FDH, VAN DER SANDE MAB, WALLINGA J. Public health measures to control the spread of antimicrobial resistance in Neisseria gonorrhoeae in men who have sex with men. *Epidemiology and Infection* 2015; **143**: 1575-84.

39. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. *Aids* 2014; **28**(10): 1509-19.

40. Owusu-Edusei K, Chesson HW, Gift TL, et al. The estimated direct medical cost of selected sexually transmitted infections in the United States, 2008. *Sexually Transmitted Diseases* 2013; **40**: 197-201.

41. Satterwhite CL, Torrone E, Meites E, et al. Sexually Transmitted Infections Among US Women and Men. *Sexually Transmitted Diseases* 2013; **40**: 187-93.