



Version 2.0

User's Manual

Steven M. Goodreau, PhD¹; Emily D. Pollock, PhD¹; Eli S. Rosenberg, PhD²;
Elizabeth M. Rosenthal, MPH²; Meredith A. Barranco, MPH²; Richard L. Dunville,
MPH³

Lisa C. Barrios, DrPH³; Maria V. Aslam, PhD⁴; Li Yan Wang, MBA, MA³

Affiliations at the time of teen-SPARC development:

¹Dept. of Anthropology and Center for Studies in Demography and Ecology, Univ. of Washington, Seattle, WA; ²Dept. of Epidemiology and Biostatistics, Univ. at Albany School of Public Health, State Univ. of New York, Rensselaer, NY; ³Division of Adolescent and School Health, Centers for Disease Control and Prevention, Atlanta, GA; ⁴Office of the Director, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA.



CENTER FOR STUDIES IN DEMOGRAPHY & ECOLOGY
UNIVERSITY of WASHINGTON

This work was funded by the US Centers for Disease Control and Prevention's 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

Welcome!	4
Acknowledgments	6
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. <i>Welcome</i> worksheet and <i>Table of Contents</i>	18
D2. <i>Population sizes</i>	19
D3. <i>Sexual behavior</i>	24
D4. <i>Diagnoses</i>	27
D5. <i>Interventions</i>	28
D6. <i>Results</i>	30
D7. <i>Export</i>	31
D8. <i>Advanced options: Alternative population size calculator</i>	32
D9. <i>Advanced options: New partners per year calculator</i>	33
D10. <i>Advanced options: Transmissibility</i>	41
D11. <i>Advanced options: Infection prevalence among partners</i>	42
D12. <i>Advanced options: Expected lifetime medical cost</i>	43
D13. <i>Advanced options: Rates of contraception usage among females</i>	43
D14. Additional worksheets	46
E. Example Scenario 1: User with jurisdiction-specific YRBS data using SAS	47
F. Example Scenario 2: User without jurisdiction-specific YRBS data	53
G. Default parameter documentation	57
G1. <i>Population</i> worksheet	57
G2. <i>Sexual behavior</i> worksheet	57
G3. <i>Diagnoses</i> worksheet	58
G4. <i>Advanced options: Alternative population size calculator</i>	60
G5. <i>Advanced options: New partners per year calculator</i>	60
G6. <i>Advanced options: Transmissibility</i>	60

<i>G7. Advanced options: Infection prevalence among partners</i>	62
<i>G8. Advanced options: Estimated Lifetime Medical Cost per Diagnosis</i>	63
H. Technical note: Model calibration	65
I. Technical note: Final equations	67
J. Limitations	69
K. Citations	71

Welcome

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

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

It includes three age groups (13–15 years old, 16–17 years old, and 18 years old).

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.

Teen-SPARC Version 2.0 includes many additional functionalities. We have added four additional behavioral changes as possible inputs. New outputs include estimates of the number of expected pregnancies among females and costs associated with each model scenario. These additions are described in more detail in their respective sections in this manual.

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 national 2015 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 the Centers for Disease Control and Prevention (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). The teen-SPARC 2.0 release also includes the number of expected pregnancies among females and the costs associated with each model scenario. Due to the nature of the YRBS data underlying many parameters, these output numbers represent adolescents attending high school. Thus, they do not reflect all incident infections and diagnoses among adolescents in the jurisdiction.

The initial teen-SPARC release modeled two behavioral changes (numbers of sex acts, probability of condom use) known to change as a result of adolescent sexual health promotion campaigns. For instance, the *Safer Choices* program reduced the frequency of intercourse without condoms (a combination of both frequency of sex and probability of condom use).¹ Similarly, an evaluation of

the *Keep It Up!* Program among young men who have sex with men showed a reduction in the number of condomless anal sex acts, and a significant reduction when limited to casual partners specifically.² Teen-SPARC Version 2.0 models several additional behavioral change scenarios, including a delay in sexual debut, an increase in screening for gonorrhea and chlamydia, PrEP use among MSM, and changes in contraception methods among females.

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; transgender adolescents are included with the gender with which they identify (as indicated by their survey responses).
- The age distribution of the population of high school-attending students in the United States varies dramatically over the course of the year because 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 only slight differences in the overall population of high school-attending adolescents 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; 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+. Because relatively few people 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, HIV, and pregnancy, 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.

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 the University of Washington, especially Martina Morris.

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

Microsoft Windows and Microsoft Excel are copyrighted products of Microsoft Corporation, WA. SAS is a copyrighted product of the SAS Institute, NC. SPSS is a copyrighted product of International Business Machines Corporation, NY. Use of trade names and commercial sources is for identification only and does not imply endorsement.

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 definitions carefully.

- **MSM:** males who have sex with males. This includes males who have sex with males only and those who have sex with both males and females.
- **MSF:** males who have sex with females only.
- **FSM:** females who have sex with males. This includes females who have sex with males only and those who have sex with both males and females.
- **FSF:** females 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:** US Centers for Disease Control and Prevention.
- **DASH:** Division of Adolescent and School Health; a division of CDC's National Center for HIV, 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 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.
- **NSFG:** [National Survey of Family Growth](#) is 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. NSFG is developed and administered by CDC's National Center for Health Statistics. For this tool, NSFG data are used to develop one default parameter (number of sex acts per partner); 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).

- **Subgroups:** in the context of this tool, subgroups refers to the nine groups modeled; three SPGs (MSM, MSF, FSM) are crossed with three age groups (13–15, 16–17, 18).

B. Basic model structure and systems thinking

This tool aims to help health departments predict the impact of behavior change on the expected number and cost of STI diagnoses among adolescents, and pregnancies among females, in high school in their jurisdictions. This allows them to explore the relative impact of different forms of behavior change and assess how it might differ by STI, age, and SPG.

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 the

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

Although this core logic is simple, each of these four components is complex; the Guided Instructions section (Section D) details these complexities as they appear. The following factors further complicate the calculations (also addressed throughout Section D):

- Calculations are subdivided by three SPGs (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).
- Data are generally only available for diagnoses; this is an incomplete picture of actual infections.
- Some data sources limit the frame of reference to those attending high school (our interest), while others do not.
- The age distribution of high school students changes substantially over the course of the calendar year.
- In many cases, we are only interested in those who have ever had sexual intercourse because they are the only ones at risk for acquisition, but some parameter sources include everyone regardless of sexual debut.
- Gonorrhea and chlamydia can infect multiple sites in the body and can be transmitted through different types of contact with different probabilities.

Despite these complexities, the basic logic of the mathematical relationship above helps us predict upcoming incident infections. In a traditional dynamic epidemic model, there is a feedback loop whereby incident infections change prevalence, which changes both the number of susceptible people in the population and the probability of partner infection, and then further affects incidence. 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, and sexual contact patterns, 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 used for each. Here, we present the conceptual logic; 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 assumptions and resulting mechanics for the model. Users who prefer hands-on learning may wish to skip this section and proceed to the Guided Instructions and Examples (Sections D–F).

(1) Number of susceptible people. 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 Sizes* worksheet) is to first enter the number of enrolled high school students in the jurisdiction, followed by the proportion of high school students by sex and age, and 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 SPG (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 incidence estimate. 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 because it does not have a short duration. Instead, we use the percentage 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. 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 adolescents 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.

(2) Mean contact rate. 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 to account for 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, which are generally lacking for adolescent populations across many jurisdictions. Finally, there are also additional types of sexual acts that carry nonzero risk for 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. While some individuals have never had sex and are at no risk, others have had sex; 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 to obtain the age pattern in cumulative lifetime partners reported in data. For users with lifetime partner numbers, from YRBS or other sources, the tool provides a place to enter the information (the *New Partners per Year* section of the *Advanced Options* worksheet). The back-calculation is then performed on the *ppy_calc* worksheet, although users should never need to work directly in this worksheet. 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 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).

(3) 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 are 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 because there is limited data on adolescent sexual networks. 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 SPGs 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 contacts represents a much smaller fraction of FSM total contacts than it does for MSM; it is so small that the general risk for FSM can be ignored. Given the high disparity in HIV prevalence between MSM and other populations, this fact is much less true for HIV than for gonorrhea and chlamydia; however, we do not model HIV for FSM in this tool.

Given this, the methods for calculating partner prevalence are similar to those used for calculating prevalence in each subgroup (described above under *Number of susceptible people*). The only difference is that there we adjust to make the count among school-attending adolescents only (because they reflect the population of interest) and here we are interested in all adolescents (the partner pool). Because we wish to know 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.

(4) Probability of transmission. The probability of transmission per condomless sex act between a susceptible person and an infected person differs by STI and by act type. Sex act types are listed with the role of the individual with infection first and the susceptible individual second. For example, 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 better understanding or changing the values should first read Section H.

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/healthyouth/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 national 2015 YRBS dataset.

Users can use their preferred statistical software. Data analysis guidelines and SAS and SPSS program code can be accessed through the above link (information located in the *Combined YRBS Datasets and Documentation* and *National YRBS Datasets and Documentation* sections). 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 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 and is found on the tool website. The code is designed for users 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.

For those using the SAS-to-Excel code, the following code is useful to run for all sections because it makes the output match the tool formatting. The following sections also include SAS code chunks where relevant.

```
*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;
*"During the past 3 months, with how many people did you have sexual intercourse?";
    %let partners3months_question = Q63;
*"The last time you had sexual intercourse, what one method did you or your partner
use to prevent pregnancy?";
    %let pregnancyprevention_question = Q66;
*****;

*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";

```

```

*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 the tool file (teen-SPARC_2.0.xlsm).

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

Macros

This file uses macros as part of the navigation and model programming, and these macros need to be enabled for the model to run properly. There are a few options for enabling the macros:

- **Option 1:** When first opening the document, the user should see either a pop-up or a message bar below the toolbar asking whether 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 on 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.
- **Option 2:** Go to File > Info > Security Warnings. Here there are two 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 asked to verify your security preferences.
- **Note:** We do not recommend changing your advanced settings in the *Trust Center* because this will change your preferences for all Excel documents; this can make the user vulnerable to untrustworthy files with ill-intentioned macros.
- For more information, visit: <https://support.office.com/en-us/article/trusted-documents-cf872bd8-47ec-4c02-baa5-1fdb1a11b53>.

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 click to move through the tool or opportunities to update or revert cell values.
- Yellow cells and blue boxes indicate automatic calculations made to progress to the final outcomes of interest.

Defaults

- Nearly all white cells contain default numbers. The sources for these numbers are described in the Section G of this document. Users are free to leave default parameters in place or change those they have data for.
- Most worksheets have buttons on them, including the following:
 - *Reset values* will remove the data from the adjacent set of cells and leave them blank.
 - *Set to defaults* will change the values of the adjacent set of cells to the default values provided with the tool.
 - *Continue to next step* advances the user to the next worksheet in the tool.
 - *Back to Table of Contents* returns the user to the *Table of Contents* worksheet.
- Other buttons that appear on only one or two worksheets are described below in the instructions about those specific worksheets.
- **Very important:** although users generally 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 Sizes* worksheet
 - *Population Sizes* worksheet Step 1: total population of high school-attending adolescents
 - *Population Sizes* 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* to the *Population Sizes* worksheet by using the button at the top of the *Table of Contents* worksheet. If you are a returning user, 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 by 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 using data from the American Community Survey. SAS code using national YRBS variable names and numbers is also provided. Variable names and question numbers might 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 to obtain this number are as follows:

- (1) Go to data.census.gov.
- (2) In the search bar on the main page, search for “School Enrollment by Level of School for the Population 3 Years and Over”.
- (3) Select the table from the American Community Survey (there should only be one table option).
- (4) Click “Geos” and then navigate to and select your jurisdiction.
- (5) Click “Years” and select the ACS Year you want.
- (6) From the resulting table, copy the estimate listed for *Enrolled in grade 9 to grade 12* and enter it on the worksheet.

Step 2. Enter the weighted proportion of males and females by age group in your YRBS dataset. Instructions to obtain this are as follows:

- (1) Define a new variable that determines whether someone is 13–15, 16–17, 18, or none of the age groups 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;
```

- (2) Define a second age variable that does not include *none of the above*. Call this variable *age2*.

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

- (3) 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;
```

- (4) 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=yrebs_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;
```

- (5) 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;
```

- (6) Export your results into Excel for easy copying and pasting; replace *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=yrebs_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 SPGs 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 only).

- 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 SPG variable *sexpartgp*.

```
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 percentage of students with valid responses who appear in each of the nine subgroups of interest. Note that these do not need to total 100% because some students will be females who are only sexually active with other females, younger than 13 or older than 18, or will not be sexually active. The SAS code provided produces 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 provide the relevant proportions of sexually experienced adolescents in each age group.

```
proc surveyfreq data=yrebs_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; replace *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 might 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 data.census.gov.
- b. Click on *Advanced Search*.
- c. Under “Topic” select the “age and sex” box.
- d. Under “Geography” navigate to your jurisdiction’s geographical name (state, county, city, etc.).
- 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 single year age from 13 to 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 may 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.
- h. After you have filled in all cells, click *Proceed to Next Step*.

Users who do not have YRBS data may choose to enter population size estimates 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 choose 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 subgroup.** Note that these values are conditional on sexual experience (i.e., the denominator does not include those who have never had sexual intercourse). This must be considered when conducting calculations or interpreting numbers. For example, in the default values for MSF, the quantity is the same across all age groups (i.e., 1.32). However, because the proportion of the population that has ever had sexual intercourse increases with age, the total amount of sexual activity might not be the same across age groups.

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 may 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, including SAS code.

The national default numbers for 16–17-year-olds and 18-year-olds are the same; this is a coincidence and does not reflect an assumption of the model that requires them to be the same. Although users might have been expecting rates for 18-year-olds to be higher than those for 16–17-year-olds, it is important to remember that these numbers are all conditioned on having had sex at all, which is an increasingly large proportion of the population as adolescents age.

Users might be wondering whether 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 underestimate the number 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, this approach simply reallocates them to the new partners at the relevant age.

2. **Average number of acts of intercourse per partner that each subgroup averages (coital frequency).** This measure is 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:

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

(2) 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;
```

(3) 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;
```

(4) 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;

```

(5) Export your results into Excel for easy copying and pasting; replace *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 from YRBS data) or use other local data sources; 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 are typically 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).

Even though the outcome metrics are in terms of school-attending adolescents, these data should represent diagnoses among adolescents regardless of whether they are in school. While this may seem contradictory, 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. Note the following:

- 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.^{3,4}
- 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,⁵ so even if jurisdictional data do not allow them to be distinguished, the actual impact on estimates is likely 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 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 may wish to leave these defaults in place, but they can choose to alter them if desired; see the *Advanced Options* sections below.

For first-time users, 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

First-time users may wish to skip this worksheet for now and proceed to the *Results*. To do so, make sure the intervention boxes on the left-hand side 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 to compare scenarios. Interventions can take many forms; we do not lay out all possible implementation details because it is impossible to anticipate all types of interventions jurisdictions might consider. Instead, this tool allows the user to input the expected impact of an intervention, or several interventions simultaneously. 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.

The initial release of teen-SPARC modeled two behavior change scenarios: a change in the frequency of sex acts and in the rate of condom use. The 2.0 release added four behavioral change scenarios. First, a delay in sexual debut represented as a reduction in the size of the sexually active population. Second, an increase in screening for chlamydia and gonorrhea represented by an increase in the proportion of incident cases that get diagnosed. Third, preexposure prophylaxis (PrEP) usage among MSM represented by a reduction in the size of the MSM population susceptible to HIV infection. And fourth, a change in the proportion of females who use long-acting reversible contraceptives (LARCs) to prevent pregnancy (changes are relative to the rates of all other forms of contraception).

For any form of behavior change, users may model a change of the same impact across all relevant subgroups. Users may also model a change of different magnitude across any or all subgroups. This is done in the *Within each group* table, which is found on the right side of the *Interventions* worksheet. While most behavior changes can be specified by the full nine subgroups, some behavior changes are group-specific, such as PrEP use among MSM and LARC use among females. These group-specific behavior changes can be modified by age group. The change in the proportion of incident cases that get diagnosed because of changes in screening is a broader change and can only be increased among males or females but not by specific subgroup.

Percentages should be entered using decimal format for all behavior changes.

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

1. The checkboxes 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 any behavior change, if both the *Across all subgroups* and *Within each subgroup* numbers are filled in, the *Across all subgroups* value is used and the others ignored.
3. For any behavior change, if the *Within each subgroup* values are being used and some cells are left blank, the tool will assume a value of 0 (i.e., there is no behavior change in that group).

Once the behavior change 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. Additionally, because female use of LARCs is so low at baseline, users are allowed to increase this rate by more than 100% of its original value. For example, if users wished to increase LARC use by 200%, they would simply enter 2 in the appropriate cell.

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 in the upper left corner and results for that model will appear to the right. Results 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. The 2.0 release also added an estimate of the costs associated with each scenario and the number of expected pregnancies among females. Cases, pregnancies, and dollars are shown rounded to the nearest whole number, while percent averted is calculated on the unrounded numbers.

If a user wishes to change their baseline model, they may click *Reset Model* to empty the results cells. They can then change various parameters, return, and click *Calculate Baseline Model* again; prior results will be lost. If the user wishes to save 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 for developers and advanced users only; most users can ignore these numbers. For more information, see Section H.

Comparison scenarios

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

After setting up the comparison model, users should return to the *Results* worksheet and click one of the *Calculate Comparison Model* buttons (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 intervention's effects 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 percentage of cases averted relative to the baseline model. The number of cases averted is expressed as all incident cases (not just diagnosed). Given the model assumptions and allowed interventions, the percentage 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 to save any of the first four they wish to keep.

D7. Export

Users may wish to export model results 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 so this is not listed as an option, but users may select up to four comparison models to export in addition to the baseline model. Once all desired exports are checked, users should click the *Export to Excel* button. A *save as* popup will appear, where users can specify their filename and location.

Each model scenario (baseline and up to four comparison scenarios) has its own dedicated page where the parameters and results are calculated (e.g., *Baseline Params*). If users also wish to export these pages, for example to reference each model's parameters, they can copy and paste these worksheets into another Excel document.

D8. *Advanced Options*: Alternative population size calculator

The 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 Sizes* worksheet.

Step 1. Obtain 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 Sizes* worksheet. To use US Census Bureau numbers, follow these steps:

- (1) Go to data.census.gov.
- (2) Click on *Advanced Search*.
- (3) Under “Topic” select the “age and sex” box.
- (4) Under “Geography” navigate to your jurisdiction’s geographical name (state, county, city, etc.).
- (5) 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.
- (6) From this table, extract the number of females and males in each single year age from 13 to 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 may 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).
- (7) 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. Identify 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 values may be difficult to obtain for many jurisdictions; users may 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. Enter the percentage of students in each age group who have engaged in sexual intercourse according to the sex of their partners. 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 might also be difficult to obtain for non-YRBS jurisdictions, so users can accept the national defaults or scale the values up or 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 OptionsNew partners per year calculator* or you can access it directly from the *Advanced Options* worksheet by scrolling down to the section called *New partners per year calculator*.

This method was employed because the tool projects STI acquisition risk over the period of a year for adolescents of different ages, which requires knowing how many partners individuals have in a year. However, YRBS does not ask this question directly, nor does any other data source with sufficient sample sizes; thus, we use back-calculation to estimate this based on available information. Any jurisdiction that does have a local data source with direct information on numbers of new partners per year can enter it directly onto the *Sexual Behavior* worksheet.

It is crucial that you check the box in the upper left corner to let the model know you will be using this method.

The purpose of this method is to convert 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 detail the mechanics for doing so below; for interested readers, we outline the conceptual logic behind this process in the Technical Note at the end of this section.

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 Sizes* 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; replace *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 first 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; replace *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 First 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 feeds 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 is that reports of lifetime partner numbers by people in different ages (e.g., 15-year-olds and 16-year-olds) provides an estimate of the number of new partners each year. However, this is complicated 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. Thus, we use information from each respondent on both their current age and their age of first sexual intercourse to determine where they fit. We assume that partners are evenly distributed between the age of first sexual intercourse and the current age. Respondents provide partner information at multiple ages (e.g., a 17-year-old who first had sexual intercourse at age 15 provides information about the partners per year at ages 15, 16 and 17). When estimating new partners per year, we do not include information from people who report having had sexual intercourse for the first time at

their current age; determining an average mean number of partners per year for such people requires making a variety of assumptions that we deemed questionable. These people appear in the total number of people who have ever had sexual intercourse and are 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 (*A: Age of First Intercourse by Current Age, Cells = Counts of Persons*) provides the weighted counts from YRBS data for the number of respondents in 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 for each current age by age at first sex combination. Matrices A and B are copies of those that appear on the *Advanced Options* worksheet.

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 divided by 3).

Now, we wish to take the weighted average 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. The calculations 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, 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; the denominator is 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 sets 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 four sums, giving the sum of all 16. Matrix set F conducts a similar summation for the denominator but sums the population size weights from matrix set A only. 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.

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 *Transmissibility* section.

Transmissibility probabilities for *Neisseria gonorrhoea* and *Chlamydia trachomatis* are shown for four types of acts, including 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 because of 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* 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 parameters; 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 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 infection, two steps are required to convert the number of diagnoses reported to the prevalence of infections.

Step 1. 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.

Step 2. 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 $prevalence = incidence \times duration$ 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 (see Section G for source information).

Step 3. The 2.0 release also includes an additional section to the right of Steps 1 and 2. This section is automatically modified when a user has implemented the *increase in diagnoses due to screening* behavior change. The increase in diagnoses reduces the mean duration of infection for gonorrhea and chlamydia in the population and subsequently reduces the prevalence of disease in the partner pool. Users do not need to modify these cells.

D12. *Advanced Options*: Expected lifetime medical cost

Teen-SPARC Version 2.0 includes an estimate of the expected lifetime medical cost per diagnosis of each of the three STIs in the next year. These values were drawn from Chesson et al.⁶ and then converted to 2015 US dollars. This was done to reflect the year in which most of the default population size and sexual behavior data were collected. Users can choose to specify their own costs per diagnosis.

D13. *Advanced Options*: Rates of contraception usage among females

The 2.0 release now estimates the number of expected pregnancies among females in the next year based on rates of contraception use by contraception type and female age group. The baseline contraception rates are estimated among six categories using the 2015 YRBS, including no method, hormonal pills, condoms, LARCs, other hormonal contraceptives, and withdrawal or other methods. Users may notice that the rate of LARC use decreases very slightly with age in the data below, which is somewhat counterintuitive. This is because the contraception question is conditional on reporting at least one sex partner in the past 3 months (a small group at the youngest ages that increases over time) and a higher proportion of younger females chose not to report what type of contraception they used at last sex. Rather than make assumptions about the group that chose not to respond to this question, we instead estimated contraception usage solely among sexually active females who answered the question. The full method used to estimate the number of expected pregnancies is based on the demography literature and is described in more detail in the *Section H* of this manual.

On the main *Interventions* worksheet, users can change the rate of LARC use relative to all other contraceptive forms. In this section, however, users can specify alternative distributions for all methods by age group. For example, users could model a scenario in which all females aged ≥ 18 years exclusively used LARCs or model what would happen if those using no contraceptive method switched to using condoms.

If users wish to specify an alternative distribution for any age group, they must check the box on the left-hand side of this section. Otherwise, the default rates will be used or they will be overridden by the LARC behavior data from the *Interventions* worksheet. Each distribution by age group must sum to 1.

Note: The condom behavior change (second option on the *Interventions* worksheet) only applies to transmission of STIs. If users are interested in how a 10% increase in condom use would affect pregnancies, they can specify that change in the *Rates of contraception usage among females* section in the *Advanced Options* worksheet. This approach is necessary because YRBS asks two questions to assess condom use: “The last time you had sexual intercourse, did you or your partner use a condom?” and “The last time you had sexual intercourse, what one method did you or your partner use to prevent pregnancy?” Females can use a primary method of contraception like LARCs and also use a condom to prevent STIs, but we are not able to estimate the overlap based on these questions. Therefore, the baseline rates of condom use for STI prevention are

drawn from the first question, and the rates of contraception use for pregnancy avoidance are drawn from the second.

To extract primary contraception usage among females by age group for local jurisdictions:

- (1) Define a new binary variable for having at least one sex partner in the past 3 months. Call this variable *sex3mo*.

```
if 3<=&partner3months_question<=8 then sex3mo = 1;
else if &partner3months_question in (1,2) then sex3mo = 0;
```

- (2) Define a new categorical variable for pregnancy prevention method at last sex called *pregprev*, ignoring those who did not have a sex partner in the past 3 months. Recode responses of *I have never had sexual intercourse* and *Not sure* as missing.

```
if sex3mo = 1 and &pregnancyprevention_question not in (1,8,..) then do;
    pregprev = &pregnancyprevention_question-1;
end;
```

- (3) Conduct a crosstab on *sex* by *agegroup* (both created in the *Population sizes* worksheet; see Section D2) by *pregprev* and output to a new dataset.

```
proc surveyfreq data=yrbs_tool nomcar varheader=label;
    strata stratum; cluster psu; weight weight;
    tables sex*agegroup*pregprev/cl row;
    ods select CrossTabs;
    ods output crosstabs=yrbs_preg_out;
run;
```

- (4) 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_preg;
    set yrbs_preg_out;
    *Create proportion reporting each pregnancy prevention method from row
    percent and round;
    pregprevprop=round(rowpercent/100,0.00000001);
    *Limit to female respondents;
    if sex ne '1' then delete;
    *Remove total rows generated by crosstab;
    if agegroup in ('1','') or F_agegroup='Total' or pregprev=. then
    delete;
    keep agegroup pregprev pregprevprop;
run;

proc sort data = yrbs_preg; by pregprev agegroup; run;

proc format;
value lpregprev    1="No method"
                  2="Hormonal pills"
                  3="Condoms"
                  4="LARCs"
                  5="Other hormonal"
                  6="Withdrawal or other";
run;
```

```

data yrbs_preg_final;
merge
  yrbs_preg (where = (agegroup=1) rename = (pregprevprop=pregprev_age1))
  yrbs_preg (where = (agegroup=2) rename = (pregprevprop=pregprev_age2))
  yrbs_preg (where = (agegroup=2) rename = (pregprevprop=pregprev_age3))
;
by pregprev;
drop agegroup;
label pregprev_age1="13-15"
      pregprev_age2="16-17"
      pregprev_age3="18+"
      pregprev="Pregnancy prevention method";
format pregprev lpregprev.;
run;

```

- (5) Export your results into Excel for easy copying and pasting; replace *USERFILENAME* with the full file name of your choice.

```

ods tagsets.Excelxp file="&filepath\USERFILENAME.xls" style=excel
options(sheet_name='Primary contraception');
proc print data=yrbs_preg_final label noobs;
  title 'YRBS primary form of contraception among females by age
  group';
  footnote 'In the Advanced options sheet, copy the proportions for
  each form of contraception by age into the table under Rates of
  Contraceptive Usage among Females';
run;
ods tagsets.Excelxp close;

```

D14. 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 relevant subgroups include the following:

- 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, number of cases, percentage diagnosed, and percentage diagnoses for each STI
- Transmission probabilities for each sex type for each STI are also shown in the boxes at the right

- Lifetime medical costs per case of HIV, chlamydia, and gonorrhea and medical and social costs per teen pregnancy
- Contraception usage rates and pregnancy estimates (including the one-year failure rates for each contraception type, specified in a box on the right of this section)
- PrEP use among MSM (if any)

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 can be copied and pasted into an additional Excel worksheet to save the parameter lists.

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

This example uses 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. 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 data.census.gov 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.

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 and 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+
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 in each of the three SPGs (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									
Controlling for Age group=13-15									
Population by sex	Sexual partner groups	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

YRBS proportion of sexually experienced adolescents by age group and SPG				
1				
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

- 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			Total Pop
	13-15	16-17	18	
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*.

- Calculate jurisdiction-specific data on numbers of partners using the process outlined in D9. This requires momentarily jumping 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 corner.

1	Matrix 1: Age at first sex by current age, cells=counts of persons							
2								
3	Sexual partner groups=MSM							
4	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								
MSM	AGE	11	12	13	14	15	16	17
	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 that lays below these two matrices 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 from this calculator in your model*).

- 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

YRBS predicted condom use for each SPG and age group				
	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: Because there is little data on MSM condom use with females, we use a combined estimate for all male condom use with females and apply it to both MSM and MSF contacts with females.

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

Neisseria gonorrhoea			
Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:			
Diagnoses among MSM	<input type="text" value="121"/>		Pop Size
		Total among males	<input type="text" value="485"/>
Diagnoses among MSF	<input type="text" value="364"/>	Total among females	<input type="text" value="865"/>
Diagnoses among FSM	<input type="text" value="865"/>		

Chlamydia trachomatis			
Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:			
Diagnoses among MSM	<input type="text" value="119"/>		Pop Size
		Total among males	<input type="text" value="1,328"/>
Diagnoses among MSF	<input type="text" value="1,209"/>	Total among females	<input type="text" value="5,398"/>
Diagnoses among FSM	<input type="text" value="5,398"/>		

HIV		
Number of diagnoses per 12-month period among 13-18 year-old MSM in or out of school:		
<input type="text" value="29"/>	<input type="button" value="Set to Defaults"/>	<input type="button" value="Reset Values"/>

- Click *Skip to Results* and click *Calculate Baseline Model*.

Calculate Baseline Model	Gonorrhea	Chlamydia	HIV
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: OR Within each subgroup:

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

	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

10. Click *Proceed to Results* then click the first box labeled *Calculate Comparison Model*. Name your intervention.

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

	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%

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. The tool can still be used with publicly available data. In this example, we will determine the impact of 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 corner marked *Important!*
2. Find the population size for your jurisdiction data.census.gov), following the instructions in Section D8. Click on *Advanced 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 2019 Population Estimates by Single Year of Age and Sex (PEPSYASEX).
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	40,007	27,159	13,975	Total	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. Thus, we 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 gonorrhoea*, and HIV, as instructed. It is crucial that diagnosis information is properly scaled to your population.

Neisseria gonorrhoea			
Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:			
Diagnoses among MSM	<input type="text" value="16"/>		Pop Size
		Total among males	<input type="text" value="87"/>
Diagnoses among MSF	<input type="text" value="71"/>	Total among females	<input type="text" value="143"/>
Diagnoses among FSM	<input type="text" value="143"/>		

Chlamydia trachomatis			
Number of diagnoses per 12-month period among 13-18 year-olds in or out of school:			
Diagnoses among MSM	<input type="text" value="15"/>		Pop Size
		Total among males	<input type="text" value="256"/>
Diagnoses among MSF	<input type="text" value="241"/>	Total among females	<input type="text" value="887"/>
Diagnoses among FSM	<input type="text" value="887"/>		

HIV

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

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

Calculate Baseline Model	Gonorrhea	Chlamydia	HIV
Expected # of in-school incident cases in next year	331	2,264	6
Reset Model	Expected # of in-school diagnoses in next year	169	847
			3

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 and FSM and a 20% increase in condom use for MSM adolescents.

By what percent should the frequency of sex acts be decreased? (Note: please use decimal format.)

OR

WARNING:

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

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

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

OR

WARNING:

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

	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

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 percentage of incident cases averted from baseline. As a reminder, cases are shown rounded to the nearest whole number, while percent 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 Sizes* 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, Step 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, Step 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, Step 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. This parameter set requires 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. As already discussed in Section D9, this method was employed because the tool projects STI acquisition risk over the period of a year for adolescents of different ages, which requires knowing how many partners individuals have in a year. However, YRBS does not ask this question directly, nor does any other data source with sufficient sample sizes; thus, we use back-calculation to estimate this based on available information.

These default values are derived specifically from the national YRBS question “During your life, with how many people have you had sexual intercourse?” Our method began by partitioning respondents according to their response to other questions, including 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 SPGs. 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 SPG. The calculations for this occur on the *ppy_calc* worksheet of the tool; users do not normally need to look at this worksheet because the calculations will occur automatically when users enter 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 four partners will be assumed to have averaged two new partners per year. This is because their time since first sex is estimated at 2 years and $4 \text{ partners} / 2 \text{ years} = 2 \text{ new partners per year}$. We then average the number of lifetime partners adolescents have had 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 younger. We then subtract sequential estimates of lifetime partners (e.g., lifetime partners at 16 – lifetime partners at 15) 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 because 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 (coital frequency). 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 is available in Section D3, Step 3.

G3. Diagnoses worksheet

Gonorrhea and chlamydia: Estimates take the form of the total number of diagnoses by SPG. Our estimates begin with sex-specific rates from the CDC’s 2016 STD Surveillance Report⁷ for adolescents, taken from Figure 5 (chlamydia) and Figure 17 (gonorrhea).

Age	Sex	Reported rates per 100k	
		CT	GC
10–14 years old	Female	91.1	19.1
15–19 years old	Female	3,070.9	482.1
10–14 years old	Male	12.7	4.7
15–19 years old	Male	832.6	280.8

We then multiplied these rates by the population sizes for 13–14 years old and 15–18 years old 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

We then summed across ages, which provided us with our final number for FSM.

For MSM and MSF, we needed to partition the male cases according to these two SPGs. One option we considered was using data from Figures 11 and 24 in CDC’s 2016 STD Surveillance Report,⁷ which report on the percentage of MSM vs. MSW who test positive for each STI in 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; a higher infection rate for one group might be the result of higher rates of screening, especially for an STI like gonorrhea that often presents acute symptoms in males. The other option we considered 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 might differ substantially between adults and adolescents, especially as patterns of having ever had sexual intercourse differ between the two groups. Lacking explicit national data assigning STI cases among adolescent males by SPG, we decided to use data from a single large jurisdiction that includes this information (New York State); data was provided courtesy of the NY State Department of Health AIDS Institute. Here, the number of diagnosed cases of chlamydia among 13–18-year-old MSW was 16.3 times higher than MSM (these are absolute numbers, not rates); for gonorrhea, the same ratio was 4.8. We then calculated the numbers of cases in each group needed to observe these ratios using the total numbers for adolescent males calculated in the previous step above.

HIV: For HIV, we used the CDC’s 2016 HIV Surveillance Report.⁸ 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 and 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 would be a linear increase across ages, with effectively no diagnoses below 15 (i.e., the five ages 15–19 would be 1/15, 2/15, 3/15, 4/15, and 5/15 of the diagnoses, respectively). Removing the 19-year-olds means reducing the combined (MSM and MSM and 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

Total adolescent population size by age and sex group (regardless of whether in school).

This was derived from the US Census national data, using the instructions laid out above in Section D2, Step 4. The numbers here are identical to those in Step 4 of the main *Population Sizes* worksheet.

Proportion attending high school by sex and age group. This 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). Using other data sources that measure this number directly (e.g., 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 because the age distribution of the population in high school changes throughout the school year.

Proportion of adolescents who have ever had sexual intercourse. This was also derived from national YRBS data, using the instructions laid out above in Section D2, Step 3. The numbers here are identical to those in Step 3 on the *Population Sizes* 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 percentage of MSM 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 and using default values for all other parameters. For a full description of this process, see the Section H.

However, we wanted to compare our resulting values to estimates drawn from the literature to determine if our values were within a reasonable range; these results are discussed in Section H. Below, we detail the literature review.

Chlamydia trachomatis and Neisseria gonorrhoea

We conducted a literature review in PubMed for articles estimating penile-vaginal and penile-anal transmissibility risk of the two pathogens, either empirically or through model calibration, and for articles from which estimates are cited and used in the current modeling literature. We used combinations of the following keywords: *Chlamydia trachomatis*, *Neisseria gonorrhoea*, 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. Because we were interested in the per-contact risk of transmission, papers that estimated per-partnership transmission risk or yearly acquisition risk were excluded.

Our final set included 23 papers that had estimates for *Chlamydia trachomatis* transmissibility,⁹⁻³¹ and 19 papers that had estimates for *Neisseria gonorrhoea* transmissibility.^{11, 13-15, 17, 18, 22, 26, 29, 32-41} Nine papers considered both diseases and were included in both analyses. All papers are listed in Section J. We recorded four parameters of interest for each infection, including 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 percentage of cases that were asymptomatic.

The median and range of values obtained from this literature review are detailed in the table below.

	Chlamydia		Gonorrhoea	
	Median	Range	Median	Range
Female to male (insertive penile-vaginal)	0.11	0.0375–0.45	0.25	0.09–0.6
Male to female (receptive 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, we relied on a recent literature review and meta-analysis that used an analogous process for obtaining means and 95% confidence intervals for HIV.⁴² 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 convert the diagnosis data into prevalence 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. Below, we explain the sources of these default parameters.

Gonorrhea and chlamydia: proportion of infections diagnosed

Our primary source is Owusu-Edusei et al.⁴³ In the absence of clear estimates for proportion of infections diagnosed for the relevant age groups, we instead use their estimate 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; in most cases, these are similar or identical to those from Satterwhite et al.⁴⁴

Gonorrhea and chlamydia: mean duration of infection

We derive these sex-specific numbers from Satterwhite et al.,⁴⁴ 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 gets diagnosed and treated. These assumptions are detailed in the *Duration* page, and are used to estimate the new duration of infection in the presence of a screening intervention.

HIV: proportion aware of their infection

For this, we turn to CDC’s 2015 HIV Monitoring Supplemental Report (Table 9b). We focus on the percentage of infections diagnosed for all 13–24-year-olds (55.6%) because an intense search of the published and gray literatures revealed this is the definition closest to our target population (13–18-year-old MSM). Although it is not ideal, the effects of including a wider population definition are somewhat mitigated because 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 to partially cancel out in that 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, 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:

$$(\# \text{ of adolescents } 13\text{--}19 \text{ living with an HIV diagnosis}) * (\text{proportion of diagnoses among MSM } 13\text{--}19) / \text{new diagnoses among MSM } 13\text{--}19.$$

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. We use the 2016 CDC HIV Surveillance Report to obtain these numbers:⁸

- # of 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 and denominator from comparable numbers summed over 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: $5,680 * 0.84 / 1,428 = 3.34$.

G8. Advanced Options: Estimated lifetime medical cost per diagnosis

These parameters were drawn from Chesson et al.⁶ and were adjusted to 2015 USD.

Pathogen	Sex	Cost in 2018 USD	Cost in 2015 USD
Gonorrhea	Males	78	73.62
	Females	254	239.75
Chlamydia	Males	46	43.42
	Females	262	247.30
HIV	Males	420,285	396,702.16

G9. Advanced Options: Contraception Rates among Females

The default values for these rates are estimated from the 2015 YRBS female respondents (note that numbers might not add to 1 due to rounding). The per-method failure rates are listed below the usage rates.

Method	Contraception Usage Rates by Age Group		
	13-15	16-17	18+
No method	0.1659	0.1316	0.1892
Hormonal pills	0.1524	0.2260	0.2350
Condoms	0.4504	0.3959	0.3783
LARCs	0.0500	0.0450	0.0435
Other hormonal	0.0613	0.0773	0.0870
Withdrawal or other	0.1200	0.1243	0.0670

One-Year Contraception Failure Rates	
No method	0.0126
<i>Relative to "No method"</i>	
Hormonal pills	0.082
Condoms	0.153
LARCs	0.0024
Other hormonal	0.047
Withdrawal or Other	0.235

H. Technical note: Model calibration

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 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 SPGs modeled for that infection (MSM, MSF, and FSF for gonorrhea and chlamydia; MSM for HIV).

One challenge 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. While the former represents data that most jurisdictions are likely to have, the latter is more appropriate for our outputs because behavioral parameters primarily come from YRBS (which is among high school-attending adolescents only). In response to this, we produce a single additional set of outputs for the base model only, which reflects 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 four reasons. The first reason is that there is an enormous amount of uncertainty around many of the transmission probabilities per act in the literature, especially for gonorrhea and chlamydia. The second reason is that these are the only set of parameters in the model that we expect to be similar across diverse jurisdictions, unlike 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). The third reason 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 route (vaginal vs. anal, insertive vs. receptive), they allow us to neatly calibrate each of the different infections and SPGs.

Through back-calculation, we identified 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 might be the result of an assumption we made in our model because of available data. In the model, we assumed that adolescents are mostly partnering with people of like ages to themselves. When determining the probability that a person's partner is infected, we used the prevalence of infection in the relevant SPG in the model, which reflects prevalence among 13–18-year-olds. However, there is often asymmetric age-matching among opposite-sex partnerships within this age range, with male partners slightly older than females. Because prevalence generally increases with age across the 13–18 age range, this means our method is

overestimating actual prevalence in adolescent MSF partners and underestimating prevalence for adolescent FSM partners. This would explain the pattern in our calibrated transmission probabilities relative to the medians in the published literature. Given this, it might 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 and 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 by a factor of approximately 4.7. This might be because young people are more likely than others to have young, recently infected partners who are potentially more infectious. Another factor is likely the role of age mixing—because HIV is incurable, its prevalence increases sharply with age, especially over this age range. Even a small number of older partners could change the HIV prevalence estimates because of the low rates among adolescents. Once again, 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 intervention effect sizes without changing these calibrated numbers. However, advanced users might opt to either calibrate their local model more precisely or to explore the implications of different infection probabilities. Users can change the transmission probabilities in the *Transmissibility* section of the *Advanced Options* worksheet and rerun 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. 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 models are built 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 model, where there are no behavioral interventions or other changes. Once we introduce changes, we begin to see results that we could not necessarily have predicted without the model. The calibration helps ensure that the model accurately targets certain key features, allowing us to 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 inputs or calculated from inputs for any subgroup. These can be indexed to distinguish the values among subgroups, but we present a general version here for visual simplicity.

s = number of susceptible people

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 method of considering the probability a person does not get infected from 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

Our expression for new infections takes the following form:

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

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 natural clearance or treatment. Thus, we wish to consider the acts with this partner as separate possible exposures. The expression $\pi\beta$ represents the probability of acquiring an infection during one act of condomless intercourse; $(1 - \pi\beta)$ 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 person and multiplying this by the number of susceptible people 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 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

Our expression for new infections takes the following 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 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. We evaluate each relationship status once to see if the partner is infected; if infected, we then consider the probability of transmission over the subsequent acts. We do this separately for each partner. There are two ways to avoid transmission from a given partner: (1) for the partner not to be infected, or (2) for the partner to be infected but for transmission not to happen. The probability of the former is $(1 - \pi)$ and of the latter is $\pi(1 - \beta)^{n(1-c)}$. Adding these together gives the probability of no transmission from a single partner, and raising this probability to the power of r does so for r partners. Subtracting from 1 gives the probability of acquiring infection from any of r partners; multiplying this by s gives the number of new infections.

Again, this approximates a complex dynamic process for a static model.

Pregnancies

We use a method based on the demographic literature and simplified for single-year estimates. We define pregnancies of interest as *recognized* pregnancies, i.e., ending in birth, induced abortion, or known fetal loss. The underlying model estimates annual pregnancies for females of age a in the next year (P_a) as:

$$P_a = E_a D_a \left(1 - \prod_{m=1:6} (1 - f r_m)^{c a \varphi_{a,m}} \right)$$

with indices (a, m) representing age and contraception method. E is population size, D is proportion having had sexual intercourse, f is the probability per act of sexual intercourse of a pregnancy with no contraception method used; r is the relative failure rate for method m ; c is the mean number of sexual intercourse acts per year across all partners for females who have ever had sexual intercourse, and φ is the proportion of sexual intercourse acts using method m . The quantity c is itself a product of the number of partners and the acts per partner. The expression inside the product represents a traditional Bernoulli probability for repeated exposures, with pregnancy probabilities in the base and number of exposures in the exponent. The latter are partitioned by contraception type, and the former reduced by each type's effectiveness.

J. Limitations

All models have countless limitations, and this one is no exception. The following lists the most significant limitations and should be kept in mind when interpreting model results:

The tool requires data from multiple sources, which are not perfectly comparable. Although most behavioral data derive from YRBS, some is filled in from NSFG, with a different sampling scheme and population (not limited to high school students, only including ages 15+). This requires some assumptions (e.g., that the sexual acts per partner of 13–14-year-old high school students are similar to that of 15-year-olds). Moreover, parameters relating to pathogen biology and care were derived from the literature; while every effort was made to select comparable behavioral data, we were limited to what currently exists used assumptions to apply them to our population (see Section G). We note that a strength of modeling is that it forces one to make these kinds of assumptions explicit; a strength of publicly available modeling tools, such as this one, is that it allows users to make different assumptions when constructing their parameters or to use different data.

13–18-year-olds in high school are not representative of all 13–18-year-olds. This is not an issue if one uses this tool for its main purpose: as a means for exploring possible behavior change within high school settings. However, if one wishes to take these results as indicative of all adolescents, the issue might become significant. Most 13- and 14-year-olds are still in middle school, so those in high school might be expected to be precocious. Those aged 15–17 who have dropped out of school are likely to differ in behavior than those in school. And 18-year-olds who have completed school and moved to higher education, the workforce, or elsewhere, might differ from high school students.

The data on 13–14-year-olds are relatively thin. Related to the previous point, the number of 13–14-year-olds in high school (and thus surveyed in YRBS) is small, yielding a fair amount of uncertainty around their precise behavior. Although we could have opted to exclude them from the tool, we chose to include them because their behavior contributes to the overall STI burden and potential impact of behavior change in high schools. We have combined them with the 15-year-olds to ameliorate the impact of uncertainty in measurement. Each group is present in the YRBS data in proportion to their size of the actual high school population, so this should not bias our estimates for the group.

The method for estimating new partners per year is especially complex and requires assumptions. We consider this back-calculation method to be the best we could use given the form of data commonly available, the form needed to construct a model of this type, and the constraints of having the method work automatically in an Excel file when new inputs are provided. However, it is a fairly indirect measure of the quantity of interest with multiple sources of error. The assumption that the partners an individual has had between their age of first sex and their current age are evenly spaced across years is difficult to verify with existing data. We note that any user who has data on new partners per year in their jurisdiction, or who wishes to explore different assumptions, may do so by entering values directly into the *Sexual Behavior*

worksheet instead of using the *New Partners per Year Calculator* on the *Advanced Options* worksheet.

Other calculations are not directly measured. One example is determining partner prevalence. The process of calculating expected incidence for any subgroup requires estimating the current prevalence among their partners. We do not have direct information on partners' prevalence and have to assume it to be similar to that of the sexually active population in the ages represented in our tool. However, some high school students partner with those outside of high school. Another example is calculating number of sex acts per partner. The data available include one survey (NSFG) in which respondents report on number of sex acts in the last 4 weeks; we must assume this is representative of the last year and then divide by the average number of partners per year for individuals of that age from YRBS. This provides multiple potential forms for error.

K. Citations

1. Coyle K, Basen-Engquist K, Kirby D, et al. Safer choices: reducing teen pregnancy, HIV, and STDs. *Public Health Rep.* 2001; 116(Suppl 1):82–93.
2. Greene GJ, Madkins K, Andrews K, Dispenza J, Mustanski B. Implementation and evaluation of the Keep It Up! online HIV prevention intervention in a community-based setting. *AIDS Educ Prev.* 2016; 28(3):231–245.
3. Brown RT, Lossick JG, Mosure DJ, Smeltzer MP, Cromer BA. Pharyngeal gonorrhea screening in adolescents: is it necessary? *Pediatrics.* 1989; 84(4):623–625.
4. Roohvarg LB, Lovchik JC. Screening for pharyngeal gonorrhea in adolescents. A reexamination. *J Adolesc Health.* 1991;12(3):269–272.
5. Marrazzo JM. Barriers to infectious disease care among lesbians. *Emerg Infect Dis.* 2004;10(11):1974–1978.
6. Chesson HW, Spicknall IH, Bingham A, et al. The estimated direct lifetime medical costs of sexually transmitted infections acquired in the United States in 2018. *Sex Transm Dis.* 2021;48(4):215–221.
7. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2016.* Atlanta, GA: US Department of Health and Human Services, 2017. Available: https://www.cdc.gov/std/stats16/CDC_2016_STDS_Report_for508WebSep21_2017_1644.pdf.
8. Centers for Disease Control and Prevention. Diagnoses of HIV infection in the United States and dependent areas, 2016. HIV surveillance report; vol. 28. Atlanta, GA: US Department of Health and Human Services, 2017. Available: <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>.
9. 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. *Sex Transm Dis.* 2006;33(7):407–415.
10. Beck EC, Birkett M, Armbruster B, Mustanski B. A data-driven simulation of HIV spread among young men who have sex with men. *J Acquir Immune Defic Syndr.* 2015;70(2):186–194.
11. Bracher M, Santow G, Watkins SC. "Moving" and marrying: modelling HIV infection among newly-weds in Malawi. *Demog Res.* 2003;9:207–246.

12. Gray RT, Beagley KW, Timms P, Wilson DP. Modeling the impact of potential vaccines on epidemics of sexually transmitted Chlamydia trachomatis infection. *J Infect Dis.* 2009;199(11):1680–1688.
13. Hui BB, Gray RT, Wilson DP, et al. Population movement can sustain STI prevalence in remote Australian indigenous communities. *BMC Infect Dis.* 2013;13:188.
14. Johnson LF, Alkema L, Dorrington RE. A Bayesian approach to uncertainty analysis of sexually transmitted infection models. *Sex Transm Infect.* 2010;86(3):169–174.
15. Johnson LF, Dorrington RE, Bradshaw D, Coetzee DJ. The role of sexually transmitted infections in the evolution of the South African HIV epidemic. *Trop Med Int Health.* 2012;17(2):161–168.
16. Katz BP. Estimating transmission probabilities for chlamydial infection. *Stat Med.* 1992;11(5):565–577.
17. 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.* 2000;14(5):573–593.
18. Kretzschmar M, Van Duynhoven YTHP, Severijnen AJ. Modeling prevention strategies for gonorrhoea and chlamydia using stochastic network simulations. *Am J of Epidemiol.* 1996;144(3):306–317.
19. Kretzschmar M, Welte R, van den Hoek A, Postma MJ. Comparative model-based analysis of screening programs for Chlamydia trachomatis infections. *Am J Epidemiol.* 2001;153(1):90–101.
20. Low N, McCarthy A, Macleod J, et al. Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection. *Health Technol Assess.* 2007;11(8):1–165.
21. Lycke E, Löwhagen GB, Hallhagen G, Johannisson G, Ramstedt K. The risk of transmission of genital Chlamydia trachomatis infection is less than that of genital Neisseria gonorrhoeae infection. *Sex Transm Dis.* 1980;7(1):6–10.
22. 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. *Sex Transm Infect.* 2007;83(Suppl 1):i5–16.
23. Regan DG, Wilson DP, Hocking JS. Coverage is the key for effective screening of Chlamydia trachomatis in Australia. *J Infect Dis.* 2008;198(3):349–358.
24. Ruijs GJ, Schut IK, Schirm J, Schröder FP. Prevalence, incidence, and risk of acquiring urogenital gonococcal or chlamydial infection in prostitutes working in brothels. *Genitourin Med.* 1988;64(1):49–51.

25. Severijnen A, Laar MVD, Ossewaard J. Infecties met chlamydia trachomatis. Seksueel overdraagbare aandoeningen in Nederland. Bilthoven: RIVM (Rijksinstituut voor Volksgezondheid en Milieu); 1993.
26. Stigum H, Magnus P, Veierød M, Bakketeig LS. Impact on sexually transmitted disease spread of increased condom use by young females, 1987–1992. *Int J Epidemiol.* 1995;24(4):813–820.
27. Tu W, Ghosh P, Katz BP. A stochastic model for assessing Chlamydia trachomatis transmission risk by using longitudinal observational data. *J R Stat Soc Ser A Stat Soc.* 2011;174(4):975–989.
28. Turner KM, Adams EJ, Gay N, Ghani AC, Mercer C, Edmunds WJ. Developing a realistic sexual network model of chlamydia transmission in Britain. *Theor Biol Med Model.* 2006;3:3.
29. 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. *Sex Transm Infects.* 2003;79(5):363–367.
30. Vriend HJ, Lugner AK, Xiridou M, et al. Sexually transmitted infections screening at HIV treatment centers for MSM can be cost-effective. *AIDS.* 2013;27(14):2281–2290.
31. 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 Infect Dis.* 2013;13:436.
32. Chen MI, Ghani AC, Edmunds WJ. A metapopulation modelling framework for gonorrhoea and other sexually transmitted infections in heterosexual populations. *J Royal Soc Interface.* 2009;6(38):775–791.
33. 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. *Philos Trans R Soc Lon B Biol Sci.* 1999;354(1384):787–797.
34. Hethcote HW, Yorke JA. *Gonorrhea Transmission Dynamics and Control.* Berlin: Springer; 1984.
35. Holmes K, Johnson D, Trostle H. An estimate of the risk of men acquiring gonorrhea by sexual contact with infected females. *Am J Epidemiol.* 1970;91(2):170–174.
36. Hooper RR, Reynolds GH, Jones OG, et al. Cohort study of venereal disease. I: the risk of gonorrhoea transmission from infected women to men. *Am J Epidemiol.* 1978;108(2):136–144.
37. Marseille E, Kahn JG, Billingham K, Saba J. Cost-effectiveness of the female condom in preventing HIV and STDs in commercial sex workers in rural South Africa. *Soc Sci Med.* 2001;52(1):135–148.

38. Morin BR, Medina-Rios L, Camacho ET, Castillo-Chavez C. Static behavioral effects on gonorrhea transmission dynamics in a MSM population. *J Theor Biol.* 2010;267(1):35–40.
39. Platt R, Rice PA, McCormack WM. Risk of acquiring gonorrhea and prevalence of abnormal adnexal findings among women recently exposed to gonorrhea. *JAMA.* 1983;250(23):205–209.
40. 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. *Sex Transm Dis.* 2016;43(9):542–548.
41. 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. *Epidemiol Infect.* 2015;143(8):1575–1584.
42. 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):1509–1519.
43. Owusu-Edusei K, Chesson HW, Gift TL, et al. The estimated direct medical cost of selected sexually transmitted infections in the United States, 2008. *Sex Transm Dis.* 2013;40(3):197–201.
44. Satterwhite CL, Torrone E, Meites E, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis.* 2013;40(3):187–193.