

Estimating national adult prevalence of HIV-1 in Generalized Epidemics

You are now ready to begin using EPP to generate HIV prevalence estimates for use in the Spectrum program.

Introduction

REMEMBER

The UNAIDS/WHO Estimation and Projection Package (EPP) uses available surveillance data to estimate the [*time trend*](#) of adult prevalence of HIV-1 at the national level. EPP can be used for either concentrated or generalized epidemics, defined as follows:

Generalized Epidemic

- HIV is firmly established in the general population.
- Although sub-populations with higher risk may continue to contribute disproportionately to the spread of HIV, sexual networking in the general population is sufficient to sustain an epidemic independent of sub-populations at higher risk for infection.
- HIV prevalence is consistently over 1% in pregnant women.

Concentrated Epidemic

- HIV has spread rapidly in at least one defined sub-population, but is not well-established in the general population. Most often more than one sub-population with higher risk is affected.
- HIV prevalence is consistently over 5% in at least one defined sub-population. HIV prevalence is below 1% in pregnant women in urban areas.
- This epidemic state usually suggests active networks of risk exist within and between the sub-populations.

EPP estimates the **time trend** of HIV prevalence by fitting a simple [*epidemiological model*](#) to the surveillance data provided by HIV sentinel surveillance systems.

National HIV epidemics are usually composed of multiple epidemics in different populations and different geographic areas. To reflect this, one of the fundamental principles underlying EPP is that epidemic curves can be developed separately for different populations and then combined to produce a single epidemic curve which estimates HIV prevalence at a national level. There is no limit to the number of sub-national epidemic curves you can create as long as sufficient data exists to fit a curve for each sub-population.

In generalized heterosexual epidemics the normal way to divide the populations for modeling is by geographic subdivisions such as

- urban and rural regions,
- national subregions, e.g., provinces or states

Separate estimates and time trends are developed for each of these geographic units, and they are then combined within EPP to produce a national estimate for HIV prevalence and its trends over time.

In developing the approach to be used in fitting, the UN Reference group on Estimates, Modelling and Projections (www.epidem.org) has determined that a model with four parameters is well suited to fitting HIV epidemic curves. The shape of the resulting epidemic curve and the influences of these four parameters are discussed in more detail in articles available at <http://www.epidem.org/publications.htm>

Briefly, the four parameters are:

- **t₀** – The start year of the HIV/AIDS epidemic.
- **r** – The force of infection. A large value of r will cause prevalence to increase rapidly while a small value will cause it to increase slowly.
- **f₀** – The initial fraction of the adult population at risk of infection. The parameter determines the peak level of the epidemic curve.
- **phi** – The behaviour adjustment parameter. This parameter determines how the proportion of new entrants in the adult population who are at risk of HIV infections changes over time. If phi is negative, people reduce their risk in response to the epidemic and the curve shows a sharper prevalence decline after the peak. If phi is zero, the proportion at risk remains constant and the prevalence declines after the peak as people die. If phi is positive, risk actually increases over time and prevalence falls less quickly or stabilizes at a high level.

In *low-level or concentrated* epidemics (not discussed in this section), a fifth parameter, **d**, is included to account for the exit or turnover of people from the higher risk population groups.

Some general observations about the two types of epidemics often discussed in HIV surveillance.

In *generalized epidemics*, the prevalence among pregnant women is usually a fairly good indicator of prevalence in the total adult population. Prevalence is usually higher in urban areas than in rural areas. Therefore, most countries with generalized epidemics can be modeled as consisting of two distinct sub-populations (denoted in EPP by an icon of a rising and falling prevalence curve) : an urban population and a rural population.

In *concentrated and low-level (low prevalence) epidemics* HIV infection is primarily concentrated in certain sub-populations. In these countries prevalence among pregnant women is **not** a good indicator of total adult prevalence. Therefore, countries with

concentrated epidemics are usually modeled by combining epidemic curves for several sub-populations.

The next sections explain how to prepare the input data, run the model, find an appropriate fit to your data, and save the resulting national time trends in HIV for a **generalized** epidemic.

Preparing the necessary input data – what’s needed

1. Demographic data about your country

EPP has demographic information for most countries in the world, based on population estimates from the UN Population Division. These data are updated regularly by UNPOP. This version of EPP 2005 uses the 2004 Revision of World Population Prospects.

However

- You can examine the demographic information for your country in EPP and modify it as necessary.
- If there is no demographic information for your country, you can enter this data directly into EPP on the Define Pops tab or on the User Preferences tab.

2. Data about the HIV epidemic in your country

For generalized heterosexual epidemics surveillance data from antenatal clinics are most appropriate.

3. Data about the surveillance classifications within your country

For generalized epidemics, surveillance sites are most often classified as urban or rural with some countries adding a peri-urban category. In EPP the generalized epidemic template uses urban and rural, but you can create as many subdivisions as necessary for your own country. In some countries wide geographic variations in prevalence lead people to use provincial or state classifications for their surveillance. You will have to review your surveillance sites and their classifications to choose the most appropriate sub-divisions for your country.

4. Site specific HIV prevalence data and sample sizes

EPP 2005 makes use of individual site data and sample sizes at each site for its fits when this is available. Thus you will need to use site specific data as well. HIV prevalence data by site and by year, with sample sizes, will be entered into EPP and used for developing the epidemic curves.

Warning!! You will need to have population sizes for each of the surveillance classifications you are using.

5. Population sizes for each surveillance classification you wish to use

Since good information is often lacking regarding the characteristics of surveillance sites you may find that some data needs to be re-classified. Where

possible, the classification of surveillance sites should match the census classifications. That is, sites in towns that are classified as urban by the census should be classified as urban in the EPP data file. This makes it possible for you to use national census or population projection data to estimate the populations in the different sub-divisions you use for the country, and also ensures that the urban and rural prevalence estimates are appropriate for the population figures used in combining the urban and rural estimates into a national estimate.

The EPP model contains epidemiological parameters that describe the dynamics of HIV-1 infection. It is not appropriate for projecting HIV-2 prevalence. **Only surveillance data relating to HIV-1 should be used.**

Epidemic Structures in EPP

EPP is generally used to estimate and project the national HIV/AIDS epidemic in a particular country. A national epidemic is often made up of many smaller epidemics among special populations or in different geographic areas. HIV will start spreading earlier in some groups or places than in others. The pace and ultimate magnitude of the epidemic will also differ by population sub-group (or sub-population) or geographic classification, e.g., by urban and rural. The overall national picture of the HIV epidemic with its sub-populations and sub-epidemics is known in EPP as the NATIONAL EPIDEMIC STRUCTURE.

Within the national epidemic structure, there are **TWO major divisions**, geographic sub-epidemics and sub-populations. Sub-populations are the actual groups of people in which HIV spreads. They can be defined in a way which is appropriate for your country, e.g, people in urban areas and people in rural areas. Sub-epidemics are a way of combining these epidemics in specific sub-populations into larger geographic units. They are made up of HIV epidemics in one or more sub-populations that you wish to group together because they come from the same geographic area.

- *Sub-epidemics* should be viewed as FOLDERS which hold one or more sub-populations with HIV epidemics. Because they are made up of sub-populations, they do not have population data or characteristics themselves, but just use the HIV prevalence trends from the sub-populations of which they are composed. They are represented in the EPP interface by a GLOBE. 
- *Sub-populations* should be viewed as FILES, which actually contain the data about demographics and HIV prevalence for a specific group of people. Therefore they do contain HIV trend data and population characteristics, along with a mathematical curve that fits through those HIV trends. In the EPP interface they are represented by a GRAPH icon, to indicate that they are associated with a real HIV prevalence curve. 

Surveillance data are usually collected to monitor the status of the epidemic in different sub-populations. The particular sub-populations  selected for each country will depend on the surveillance data available.

Most commonly these sub-populations will include:

In *generalized epidemics*:

- people in *urban and rural* areas,

or

In *concentrated epidemics*:

- female sex workers, (SW)
- men who have sex with men (MSM),
- injecting drug users (IDU),
- STI patients,
- TB patients,
- blood donors,
- migrants,
- military personnel,
- truck drivers and
- low risk men and women

EPP allows you to calculate the individual HIV epidemic curves for each important sub-population and then combine these curves easily to form a national estimate of HIV prevalence trends.

More complicated epidemic structures can also be defined, but one of these two general types will be appropriate for most countries.

Understanding the EPP screens

All screens have a HELP button. Please use it to get more information about that part of the program.

All screens have a COMMENT button. These buttons will open a text box that allows you to make a record of your data sources, the decisions you make, and other useful information along the way.

When you start EPP you should see a screen that looks like the one shown below in figure 1.

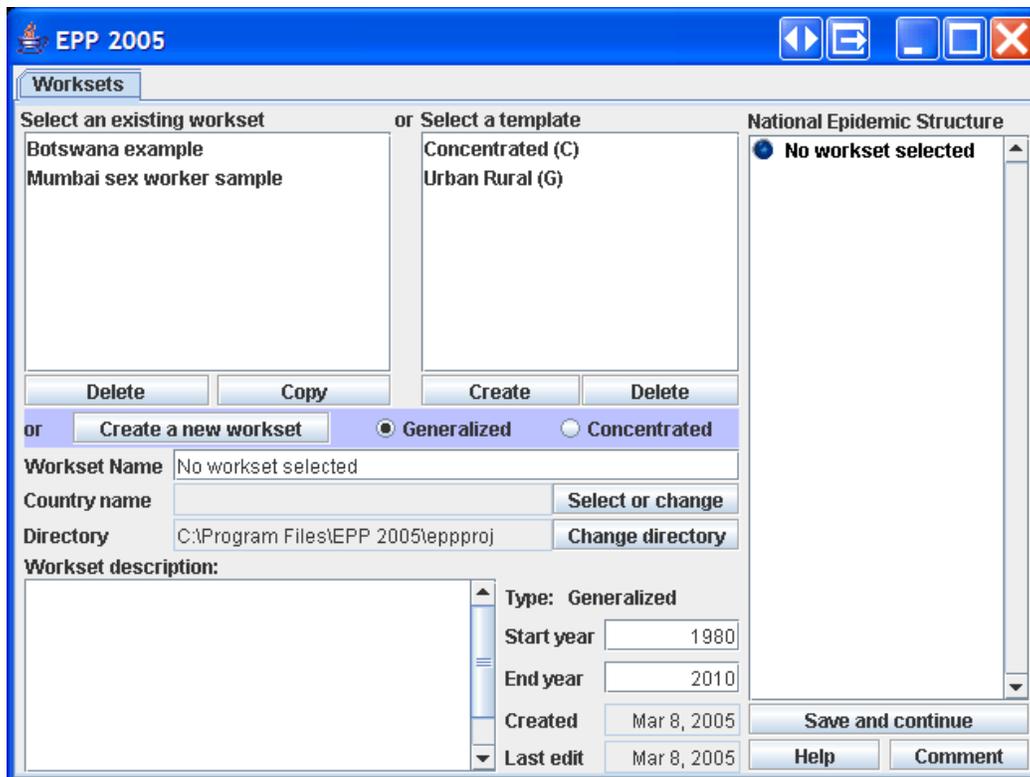


Figure 1 – The EPP Workset tab

The EPP interface – taking you through the steps

Each of the key functions of the EPP program is kept on a separate “tab” to keep the interface clean and uncluttered. The tabs, which appear across the top of the screen after you select a workset are shown in Figure 2 below. The tabs are:

- 1) WORKSET;
- 2) DEFINE EPI;
- 3) DEFINE POPS;
- 4) ENTER DATA;
- 5) PROJECT;
- 6) CALIBRATE;
- 7) RESULTS; and
- 8) PREFS.

The first 7 tabs correspond to the basic steps in creating a national estimate and projection:

1. Set up a file to contain your national epidemic (the workset in EPP)
2. Define what sub-populations and sub-epidemics are needed in your country
3. Provide the sizes and demographic characteristics of each key sub-population
4. Enter HIV prevalence data
5. Fit the prevalence data and make a short-term projection

6. Calibrate your projection to match HIV prevalence measured in large scale surveys
7. Explore your results and their implications

And the final (Prefs) is used to set various default values – this will be discussed in more detail later.

NOTE: The tabs mentioned above ONLY appear AFTER you have selected a workset to use, and chosen to SAVE AND CONTINUE.

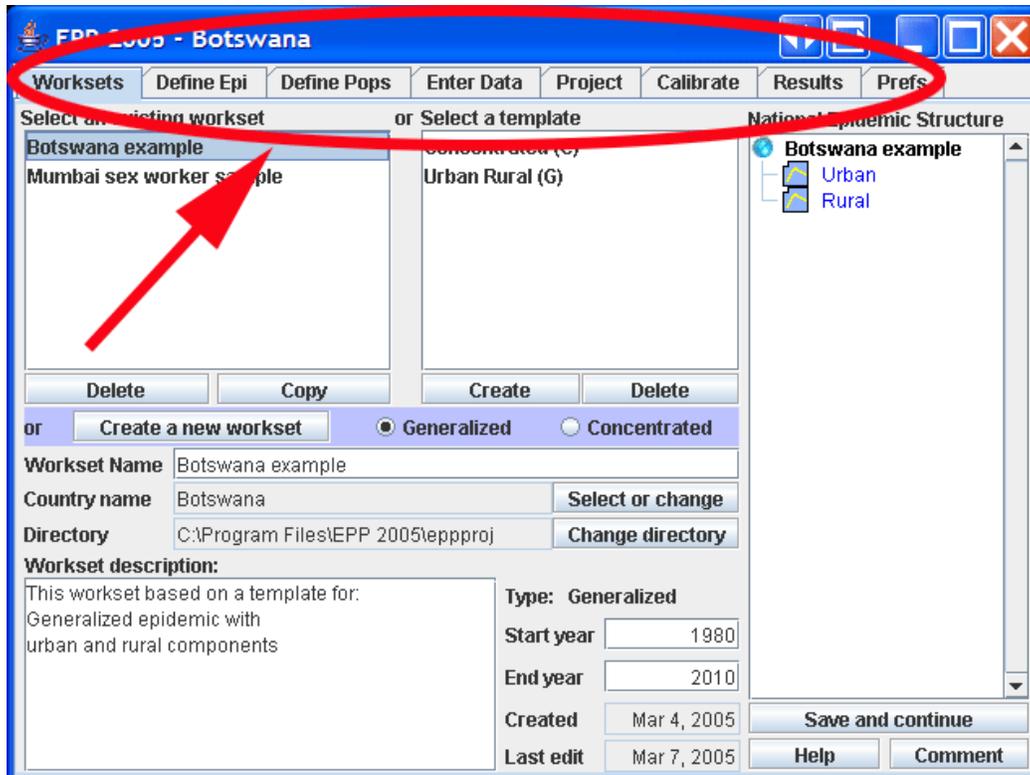


Figure 2 – The other EPP tabs

The functions and behavior of each tab are described in more detail in the following pages. You create a projection and calculate the HIV prevalence estimates required by Spectrum by working through these tabs in order. Once you have entered the data required on each tab for a given sub-population you will hit “Save and Continue” and EPP will automatically move you to the next sub-population. Once you have entered data for all sub-populations, it will automatically move you to the next tab. Thus, the program helps to step you through the process of preparing estimates and projections.

The Worksets Tab – Starting a national projection

This is the opening panel when you first run the program and the starting point for all national projections of HIV. The *workset* is where EPP stores all the information about your national epidemic. Each workset represents one national epidemic and can contain multiple sub-epidemics and sub-populations and their associated data and

model fits. The user can define as many worksets as desired - each is stored as a separate file on your hard drive and shows up as a separate entry in the list under “Select an existing workset” on the Worksets Tab. On the Worksets Tab you can either choose to use an existing workset or start a new one.

In order to start with EPP, you must load or create a workset to contain your national epidemic. There are four ways to do this.

Option ONE – Using an existing workset

- ***Select an existing workset*** by clicking on the workset name in the list of existing worksets.
- ***Click “Save and continue”*** to use this workset and move on to the next tab.

The list of existing worksets is found in the box on the upper left corner of the screen under the label “Select an existing workset”. When you select a workset from the list, the relevant information for that workset (country name, start and end years, creation and last edit date) will appear in the bottom part of the window. The example of selecting an existing workset for Botswana appears below (figure 3). The actual structure of the national epidemic will appear under “National Epidemic Structure”, which in this case consists of two sub-populations, “Urban” and “Rural”. If you have entered data in this program before for the Botswana workset, this data will be loaded automatically when you select it out of the list.

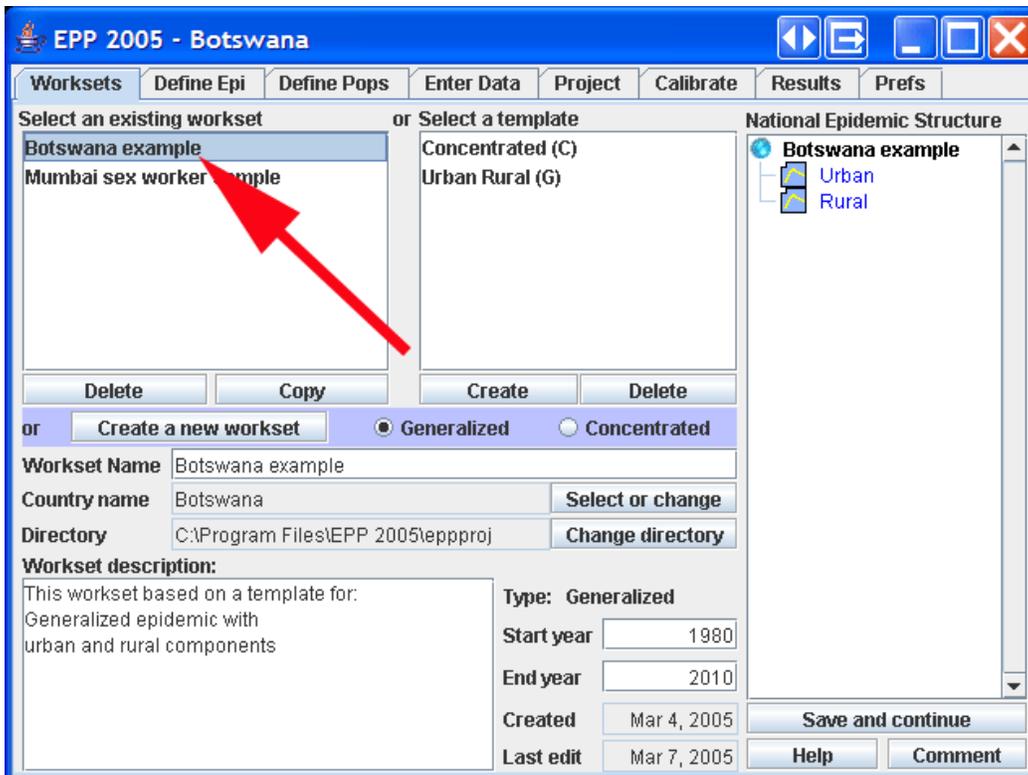


Figure 3 – The existing workset list

If you have not entered any data before, the screen will appear as below (Figure 4). Notice that the workset list indicates that there are no files in the directory. To get started you will need to create a workset (Option TWO), select a template (Option THREE) or create a template (Option FOUR)

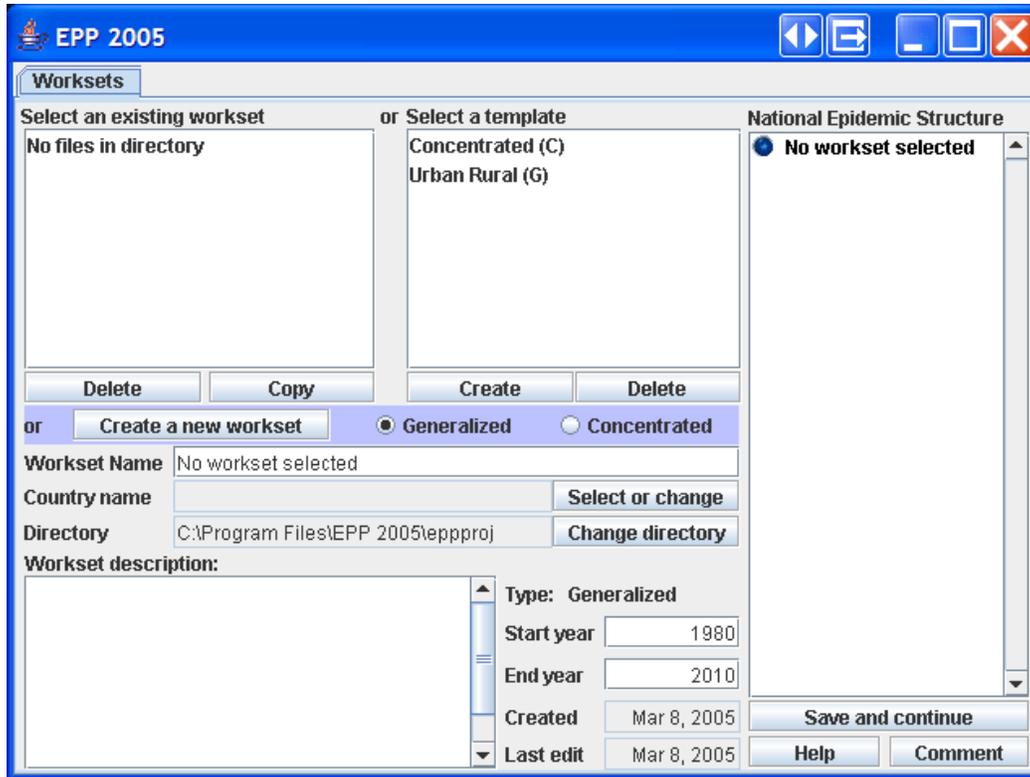


Figure 4 – Starting with no worksets in the current directory

Option TWO – Creating a completely new workset

- **Create a new workset** by taking the following steps:
 1. **Select either a “Generalized” or a “Concentrated” epidemic** by clicking next to those words in the purple area in the middle of the page (this area is circled in Figure 5).
 2. **Click on the “Create a new workset” button** which will actually create the new workset. The arrow in Figure 5 shows you where this button is.
 3. **Enter a name for the workset** by typing in the area next to “Workset name” and then hitting <ENTER>, e.g., in Figure 6 the name “Jamaica first try” has been typed in for the workset name.
 4. **Select a country for your workset** from the pop-up list which appears when you hit <ENTER> after typing the workset name or if you click on “Select or change” next to the country name.
 5. **Change the start year and end year of the projections** if desired by typing in new years next to “Start year” and “End year”. Most users will not need to change the values provided here.

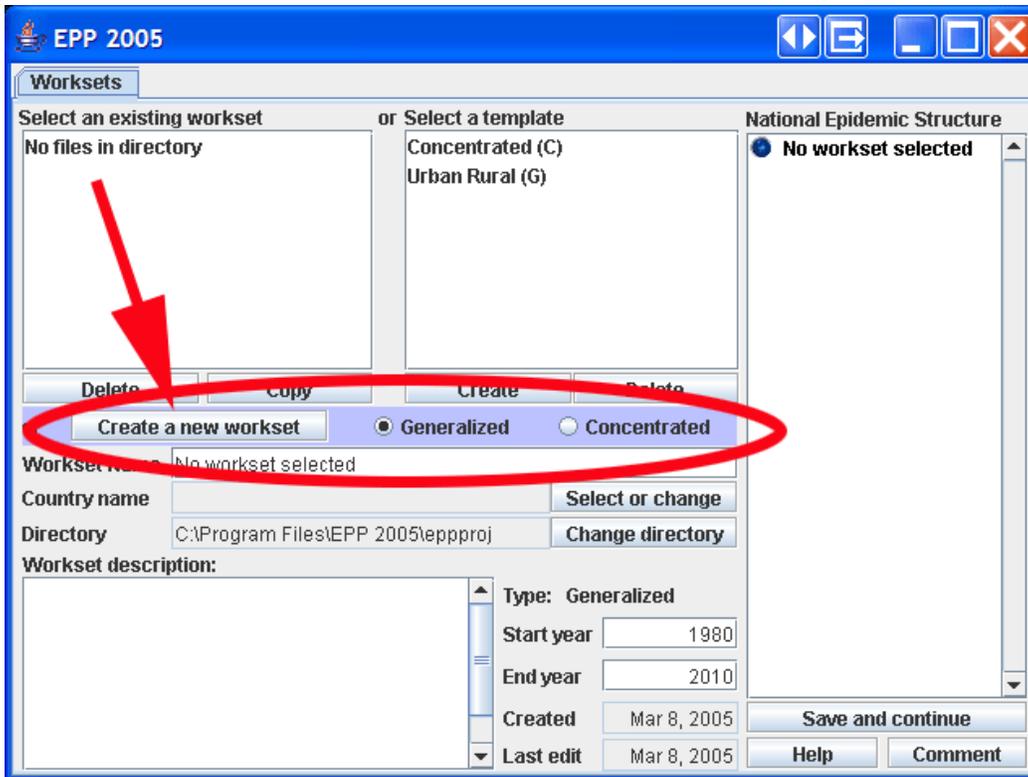


Figure 5 – Creating a new workset

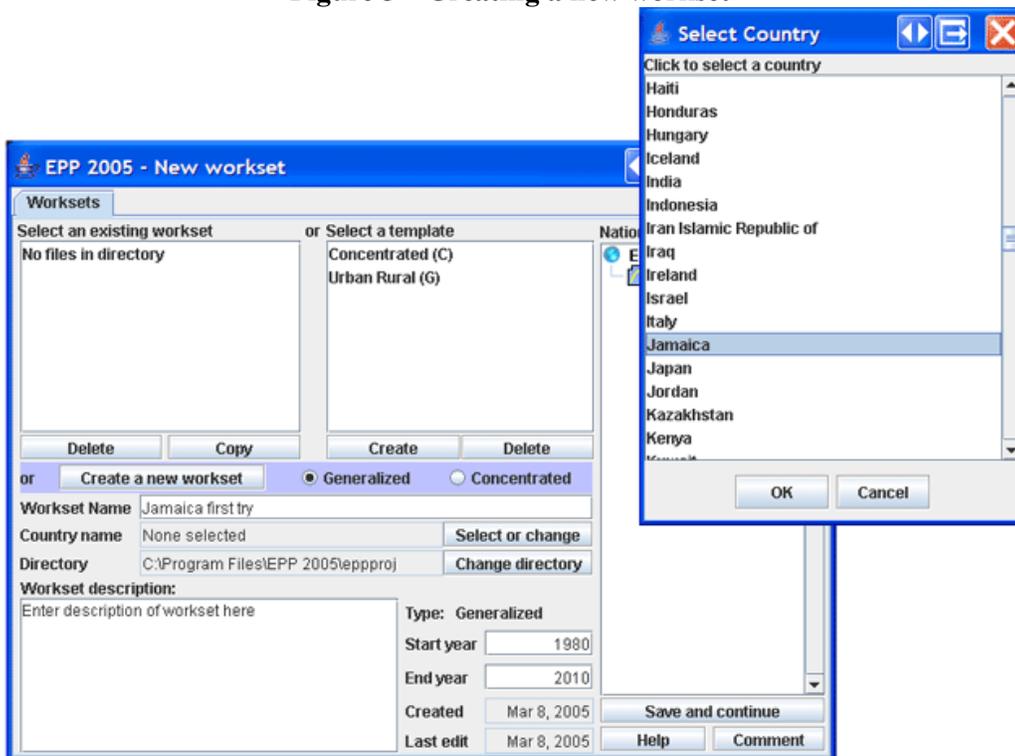


Figure 6 – Naming your workset and choosing a country

6. *Enter a brief description of your workset* by typing into the “Workset description” area near the bottom of the page.
7. *Save your workset* by clicking on “Save and continue”.

NOTE: *Be sure to follow the steps in the order listed above. Decide if you wish to have a “Generalized” or “Concentrated” workset and select the correct type BEFORE clicking on “Create a new workset”. This is necessary to ensure the workset you create is of the correct type.*

When you have completed these steps, you will now automatically be taken to the Define Epi tab to provide details about the sub-populations involved in your epidemic.

Figure 7 shows what the screen looks like after you have completed the first 6 steps above, but before you have hit “Save and continue”. Notice in particular that you have a National Epidemic Structure, pointed to by the red arrow, consisting of one-subpopulation “Remaining Pop”. EPP 2005 automatically adds this sub-population in to ensure that you remember to put your entire national population into one or more sub-populations. You will be able to add additional locally relevant sub-populations on the Define Epi tab.

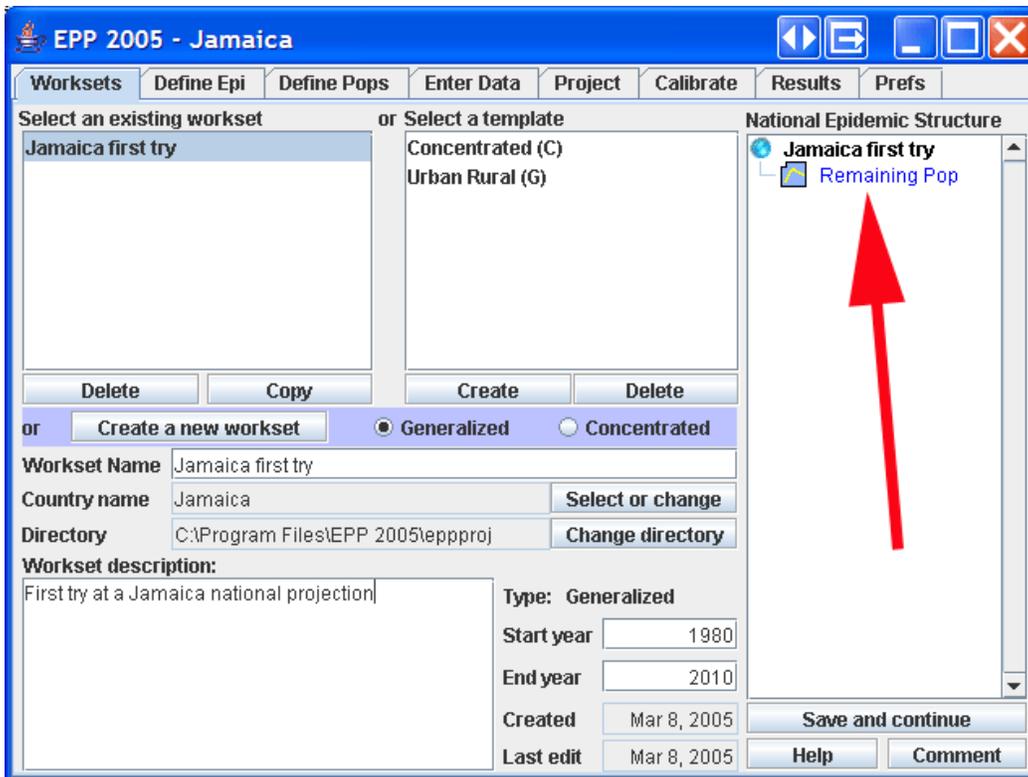


Figure 7 – The default epidemic structure in new EPP worksets

Option THREE – Creating a new workset from a template

One often desires to use a pre-defined form for an epidemic. For example in most Sub-Saharan African countries the epidemic is usually modeled with an urban sub-population and a rural subpopulation. EPP provides for this by providing two predefined forms of the epidemic in what are called *templates*:

- *Urban Rural (G)*. This is an epidemic consisting of one urban sub-population and one rural sub-population. This is most appropriate in generalized epidemics where HIV prevalence in antenatal women is above one percent.
- *Concentrated (C)*. This is for an epidemic where HIV remains primarily focused in populations with greater risk for HIV: injecting drug users (IDUs), men who have sex with men (MSM), female sex workers and their clients.

If your epidemic fits one of these patterns, the easiest way to specify the sub-populations  is to choose one of these templates in the template list in the red box in Figure 8. In this figure the *Urban Rural (G)* template has been selected and you can see that the National Epidemic Structure consists of an  urban sub-population and a  rural sub-population.

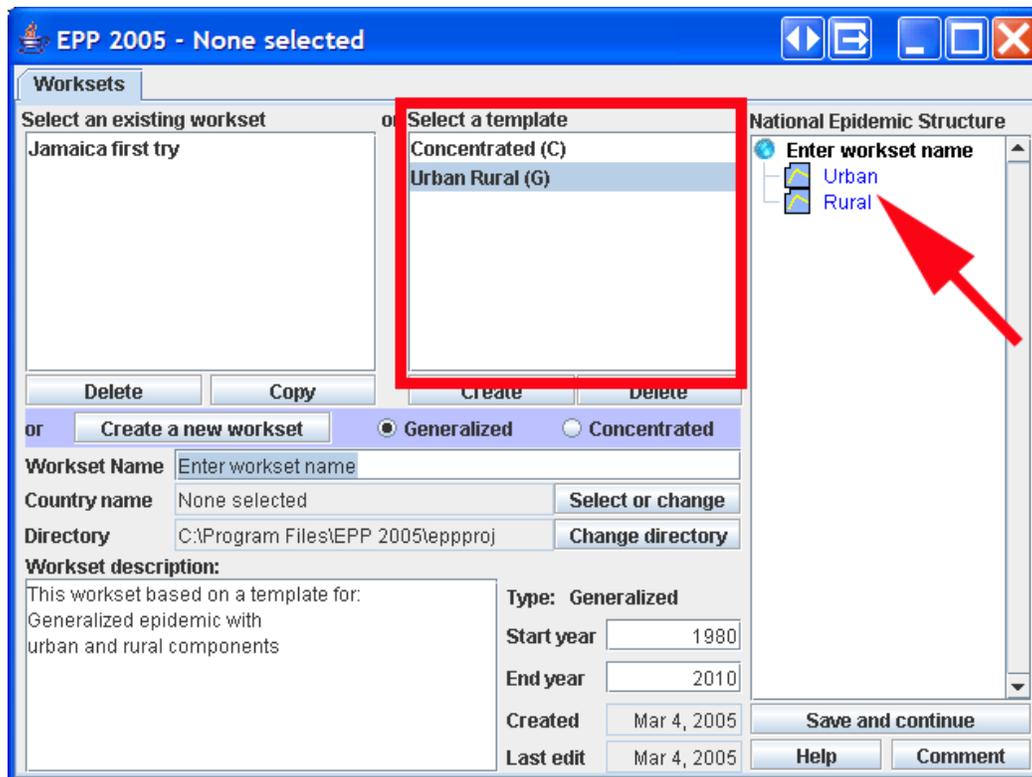


Figure 8 – The EPP template list and the *Urban Rural (G)* template for generalized epidemics

If one of these templates fits your needs, then creating a new workset is easy. The steps are:

1. **Select the template by clicking on its name in the** list underneath the words “Select a template”.
2. **Enter a name for the workset** by typing in the area next to “Workset name” and then hitting <ENTER>.
3. **Select a country for your workset** from the pop-up list which appears when you hit <ENTER>.
4. **Change the start year and end year of the projections** if desired by typing in new years next to “Start year” and “End year”. Most users will not need to change the values provided here.
5. **Enter a brief description of your workset** by typing into the “Workset description” area near the bottom of the page.
6. **Save your workset** by clicking on “Save and continue”

You will now be automatically taken to the Define Epi tab. Should you have a generalized national epidemic that requires more than just urban and rural sub-populations , you will be able to identify those additional sub-populations on the DEFINE EPI tab.

FOR ADVANCED USERS

Option FOUR - Creating your own templates

Occasionally you might want to create a new template that you will be able to use over and over. For, example, this might be the case if you have a split other than Urban/Rural that you would like to use repeatedly in several regions of the same country.

To create your own template, take the following steps:

1. **Click on the “Create” button** underneath the template list (this button is circled in Figure 9). This will bring up a window for “Creating a new template”, which is shown in Figure 10.
2. **Follow the steps outlined on this window:**
 - a. **Enter a name for your template** and type <ENTER>
 - b. **Select either “Generalized” or “Concentrated”.**
 - c. **Decide if you want this template to apply to one specific country.** If you click “No” here, the template can be used to generate similar epidemics for any country.
 - d. **Build the epidemic using the instructions for the Define Epi tab** (see later in this document for the details of doing this)
 - e. **Enter any comments you might have** by clicking on the “Comment” button and typing in your comments. These will appear in the description area when you use the template to create a new workset.
 - f. **Click on “Save”** to save the template and return to the Workset tab. Your new template will now appear in the list of templates and is ready to be used as described above in Option THREE.

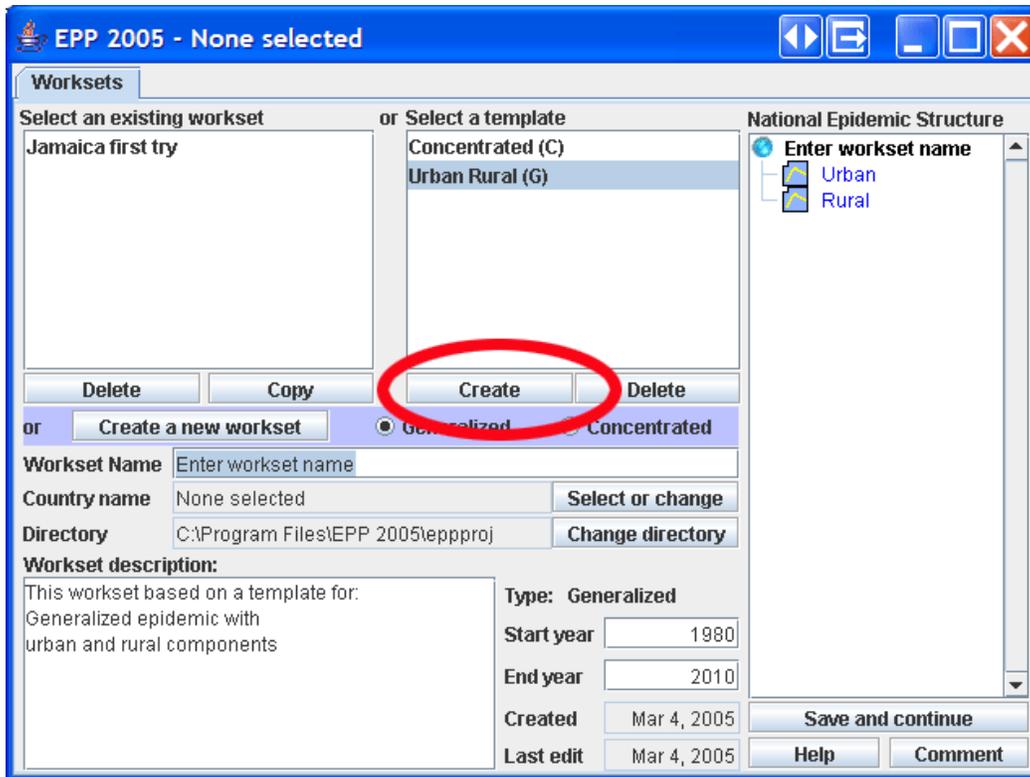


Figure 9 – The button to start creating your own template

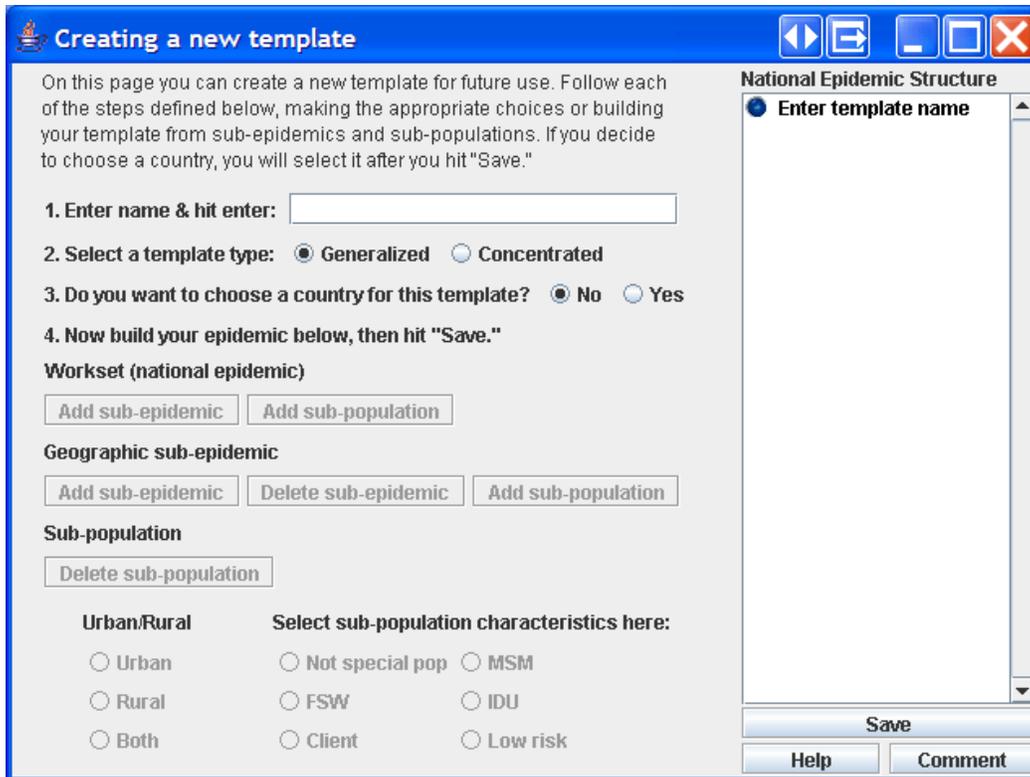


Figure 10 – The window for template creation (instructions included)

For more information, click on the “Help” button which shows an example of building a new template.

A few details about the entries on the Workset tab

1. *Workset name.* The workset name can be any legitimate Windows filename. The workset files themselves are stored as *<Workset name>.epr* in the *eppproj* directory. YOU DECIDE THIS NAME. Typically it is the name of the country for which you are developing estimates.
2. *The Workset type.* You will also notice that the type of workset (generalized or concentrated) must also be chosen . This must be done BEFORE creating the Workset. ONCE THE WORKSET IS CREATED, THE TYPE CANNOT BE CHANGED. Selecting the proper one of these two choices is important for screens you will use later in the program. You will **not** be able to come back to this page to re-set this important component at any later point in the program.
3. *Country Name* – this is normally chosen from the United Nations list of existing countries when the projection is first created. After you type in a Workset name, press <ENTER>. A pop-up window will provide you with a list of countries from the United Nations Population Division. Choose the appropriate one for your estimates. If the country is not listed, choose “Country not listed” and you will be asked to enter a name for it. In this case, you will later need to provide the demographic parameters. For countries in the list, UN Population Division has already provided the demographic data.
4. *Directory* – this is where the workset files are stored on the computer. The default will be **C:\Program Files\EPP 2005\eppproj** after a normal install. The user can change this by hitting the “Change directory” button and moving to another directory.
5. *Description* – a text description of this projection. This is created when you create or edit the projection. This description can be of arbitrary length. If it is too large to fit in the space provided, scrollbars will automatically appear.
6. *Projection start year and projection end year* – these are the years in which the calculation begins and ends. Note this is NOT necessarily the same as the year in which the HIV epidemic starts, but represents the years over which you are calculating populations and HIV spread. Normally you should start projections in 1980 and end them no later than 2010. The UNAIDS Reference Group strongly recommends against doing long term projections with EPP.
7. *Created* – date on which the workset file was first created
8. *Last edit* – date on which the workset file was last changed in any way. This can include editing projection information, adding or changing data, running a new fit, or changing any of the model results. Everytime you hit a “Save and continue”, this information is updated.

The Define Epi Tab – Laying out your epidemic’s structure

On the Define Epi tab (short for “Define Epidemic”) you will make any changes to the country’s epidemic structure that are desired. Recalling that any country’s epidemic is made of up smaller epidemics in sub-populations, your goal on this page is to lay out all of the sub-populations that are contributing to the epidemic in a substantial way.

For many countries with generalized epidemics, one urban and one rural sub-population are sufficient if the national epidemic is fairly similar in different parts of the country.

If you have chosen to build a generalized epidemic based on the Urban Rural (G) template, the screen will appear as in Figure 10 when you first come onto this page.

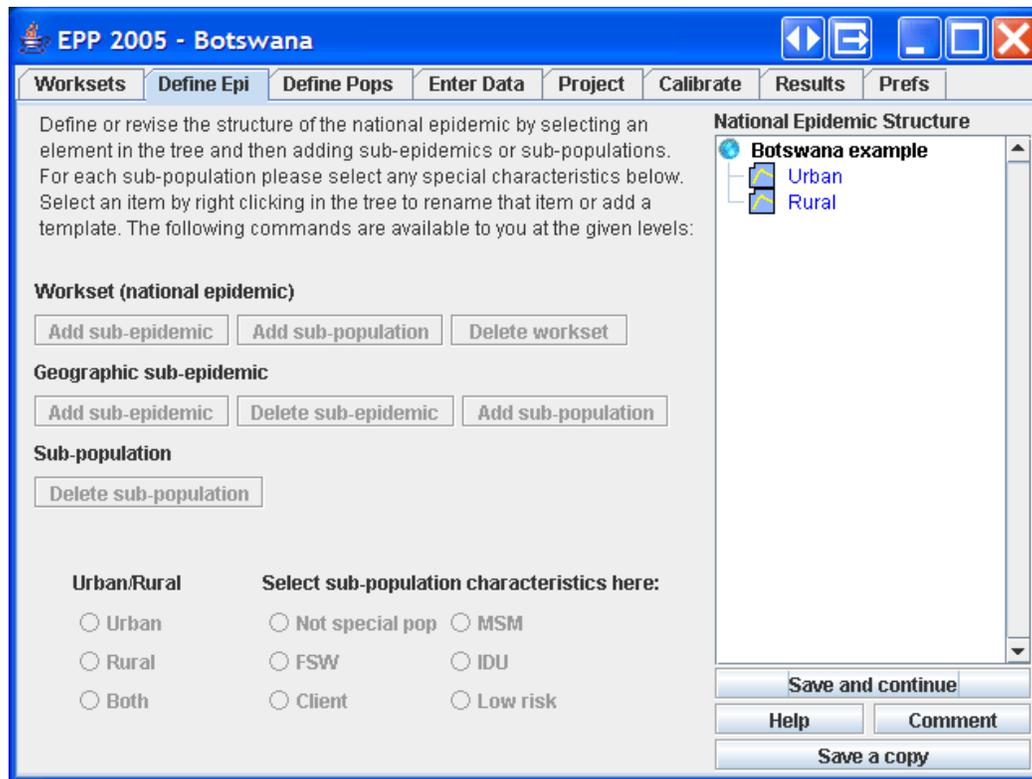


Figure 10 – The epidemic structure of many generalized epidemics as seen on the Define Epi tab

This tab shows an epidemic structure in Botswana that is defined by one  Urban sub-population and one  Rural sub-population. The total size of these sub-populations must add up to the national total population. Similarly the national HIV epidemic will be the sum of the number of people living with HIV in urban areas and the number in rural areas.

If you have selected a previously prepared country file by choosing that country on the first tab of this program, or decided to create a new file from a template, then the epidemic structure should already be correct.

You can now click the “Save and continue” button to go to the Define Pops tab.

REMINDER: The user must have appropriate data for these sub-populations, including the sizes of populations within these geographic divisions and HIV prevalence data over time for these geographic divisions. You should never add a sub-population for which you do not have both of these types of data.

NOTE: If you have created your own template or workset, the category “Remaining population” may appear. When you create your own new workset, by default the program will provide you with a sub-population  named ‘Remaining Pop’. You can change this name by right clicking on Remaining Pop in the epidemic structure tree and choosing *Rename*.

NOTE: At this point you may also wish to save a copy of the workset that you have developed. To do this, click on the ‘Save a copy’ button and provide a name for the copy of your workset. Should you have need to return to this stage of using EPP on this workset, you may access your work so far under the new name you supplied by going to the Workset tab and loading it from the workset list.

FOR ADVANCED USERS

In order to deal with the full complexity of different national epidemics, EPP has been designed to allow the user the flexibility to add as many sub-epidemics and sub-populations as necessary to reflect the current national situation. As mentioned earlier, EPP allows you to define sub-populations which are groups of people among whom HIV spreads, and sub-epidemics which are geographic collections of sub-populations which are related to each other. On the Projection tab, a separate epidemic curve or *projection* will be calculated for each sub-population by fitting the HIV prevalence trends in that sub-population. The curves from all the sub-populations making up a sub-epidemic will then be combined to calculate the total number of infections and HIV prevalence in the sub-epidemic.

The generalized epidemic template (*Urban Rural (G)*) has a single national “epidemic” composed of two sub-populations  (urban and rural). However, if you wanted to make separate estimates for different regions of the country, each with data on urban and rural sub-populations , you might create a structure like this:

- o East Region
 - Urban
 - Rural
- o West Region
 - Urban
 - Rural
- o North Region
 - Urban
 - Rural

- o South Region
 - Urban
 - Rural

The easiest way to do this is to start by creating a new workset without using templates. For example, let's assume on the Workset tab, we create a new workset for the country of Juminta, called "Regions of Juminta". After we create it and click on "Save and continue" on the Workset tab, we will be taken to the Define Epi tab, which will now appear as in Figure 11.

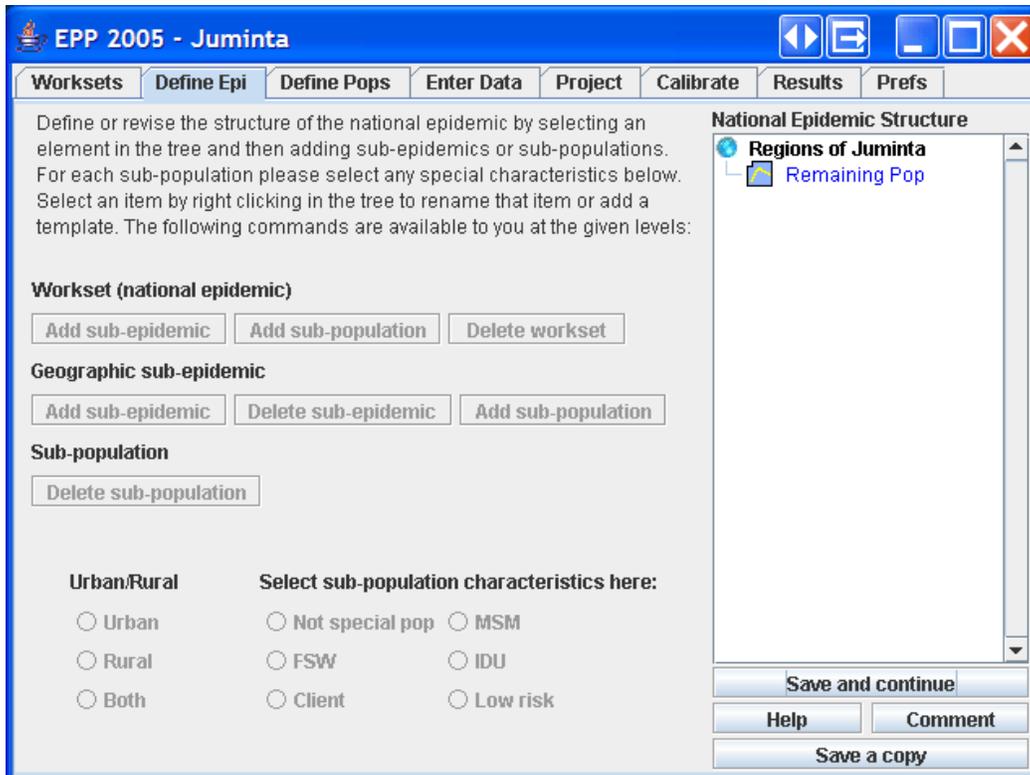


Figure 11 – The Define Epi tab after creating a new workset

We can now modify this structure by adding and deleting sub-epidemics  and sub-populations . To add a sub-epidemic for the East region, take the following steps:

- **Click on the workset name** "Regions of Juminta" in the epidemic tree. The screen will now appear as in Figure 12. Notice that the workset is selected (as you can see by the background changing to light blue) and the buttons under "Workset (national epidemic)" are now activated.
- **Click on "Add sub-epidemic"** – this is the button pointed to by the arrow in Figure 12.
- **Type in the name of the desired region "East" and click "OK"**

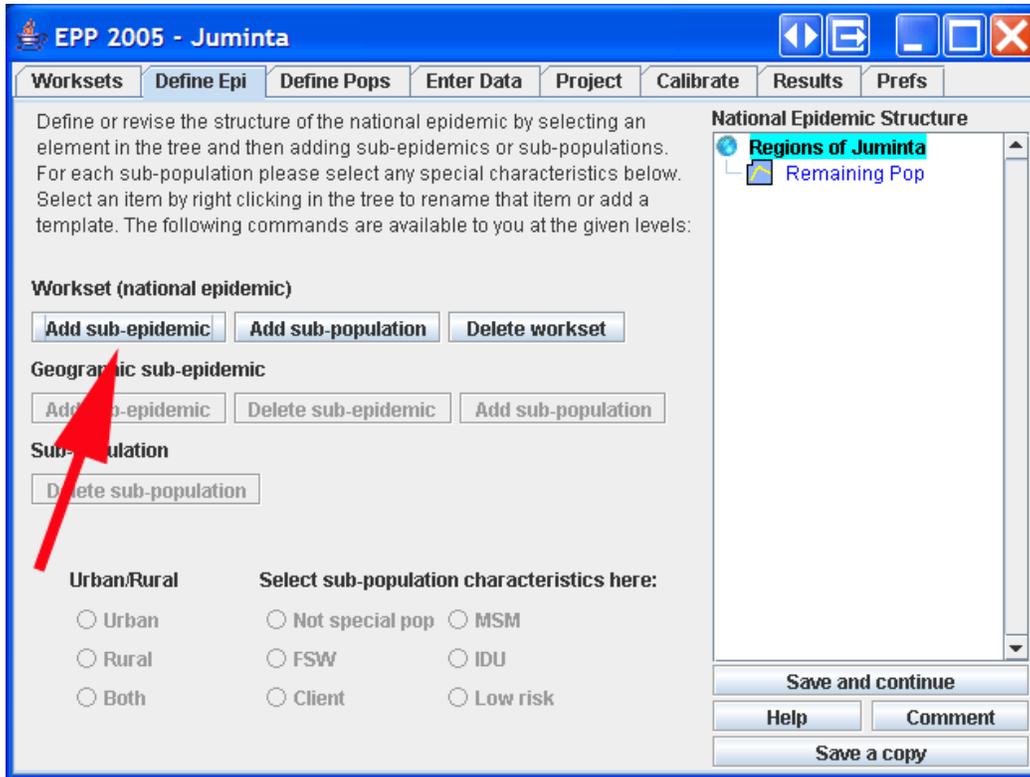


Figure 12 –After selecting the workset in the National Epidemic Structure tree

The screen will now appear as in Figure 13. To add the Urban and Rural sub-populations:

- *Click on “East” in the epidemic structure tree.*
- *Click on “Add sub-population”*
- *Type the name “Urban” and click “OK”*
- *Click on “Add sub-population” again*
- *Type the name “Rural” and click “OK”.*

When you are done, your epidemic will now appear as in Figure 14. The final step then is to remove the Remaining Pop under “East”. If you try to just remove it, you will be told you cannot delete Remaining Pop. So you must do 2 steps:

- *Select “Remaining Pop” in the epidemic structure tree.*
- *Right click and select “Rename”. Give it any name, e.g., “a”*
- *Then click on “Delete sub-population”.*

When you have completed this for both “Remaining Pops”, your structure will look as in Figure 15. Now you could continue to do this three more times for each additional region (“West”, “North”, and “South”), but that’s a lot of work. So there’s an easier way!

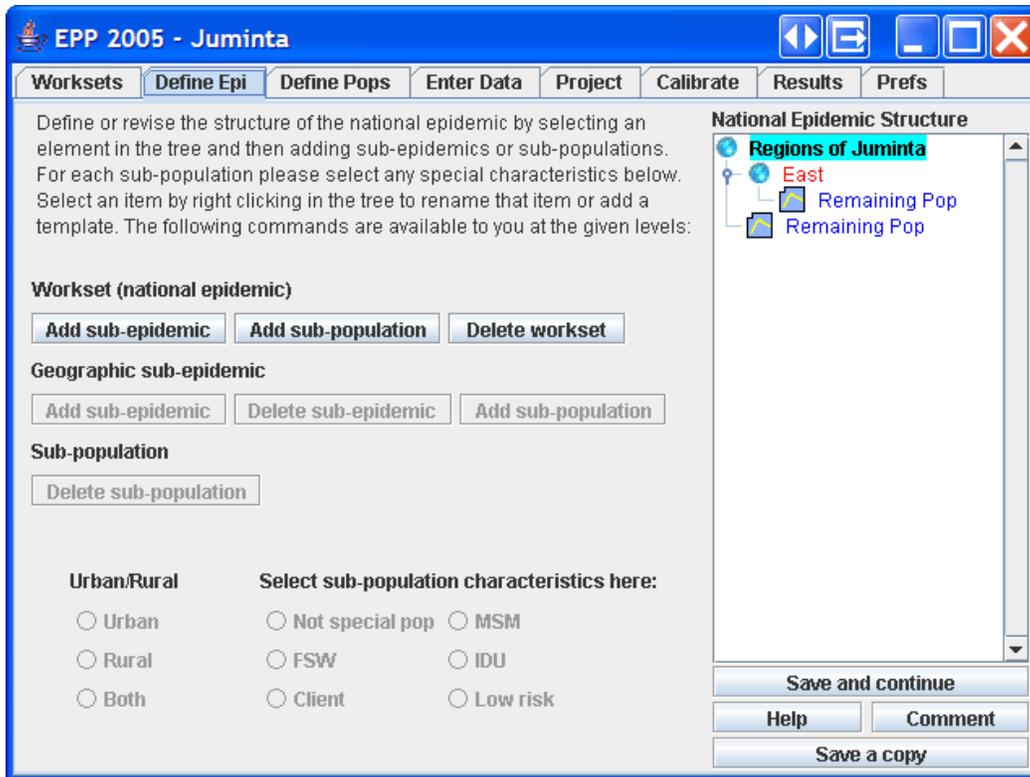


Figure 13 – After adding the sub-epidemic in the “East”

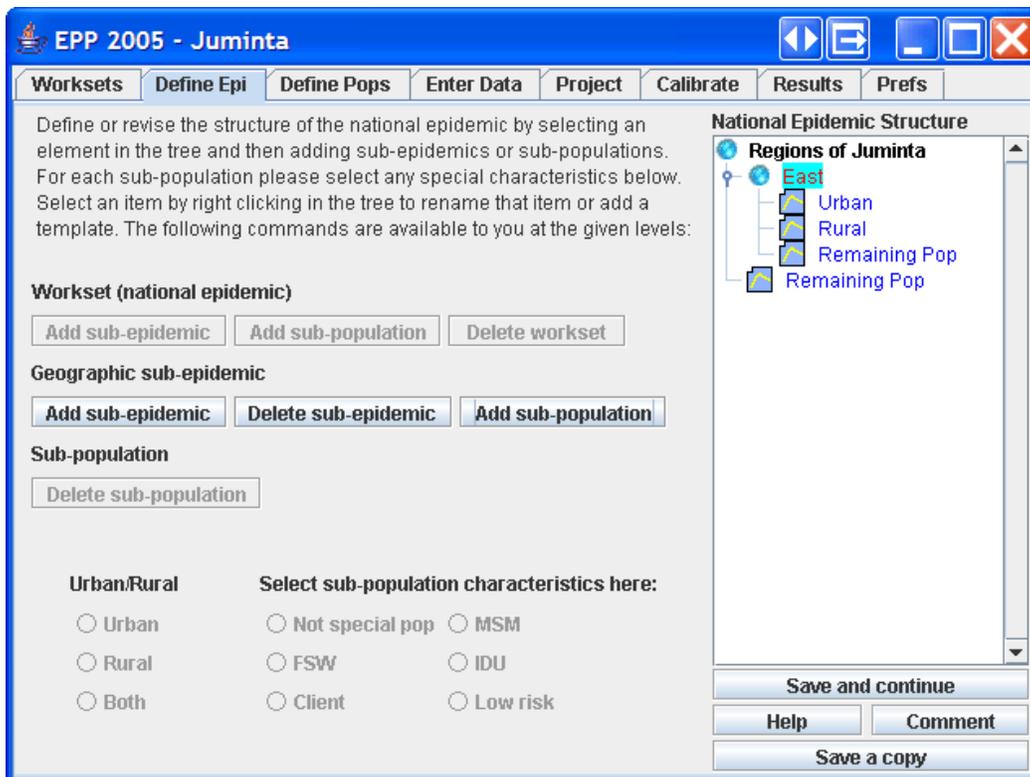


Figure 14 – After adding Urban and Rural sub-populations to East.

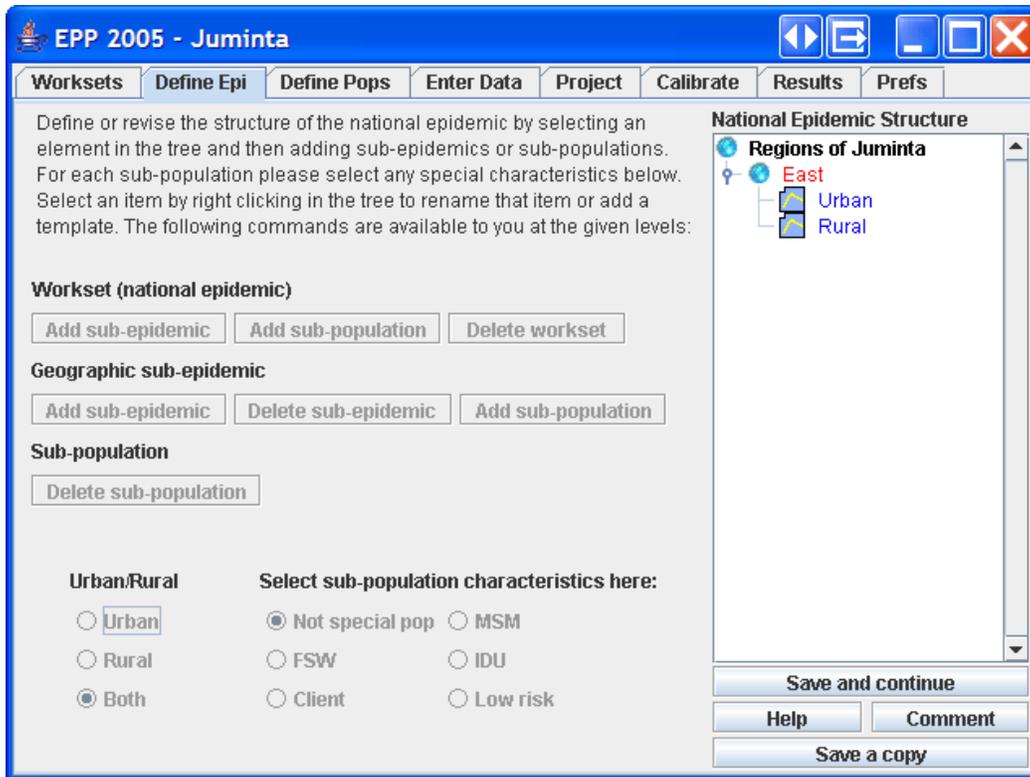


Figure 15 – The finished “East” sub-epidemic

To add the other two regions, which are going to have Urban and Rural sub-populations as well:

- *Click on the workset name in the epidemic structure tree.*
- *Right click and a menu will pop up (see Figure 16).*
- *Select the Urban Rural (G) template.*
- *Type in a name for the region, e.g., “North” and click OK.*

The screen will now appear as in Figure 17. Repeat this process twice for the “West” and “South” regions and you will have created your national epidemic with four sub-epidemics (one for each region), each of which is composed of Urban and Rural sub-populations.

SHORTCUT: You may create sub-epidemics with the full set of sub-populations in a template by right clicking on the workset and choosing an existing template (e.g. Urban Rural (G)) containing the sub-populations you desire.

IMPORTANT REMINDER

You will need to know the size of the populations and have HIV prevalence data for each of the sub-populations you create. Keep this in mind as you add new sub-populations. Do not add regional sub-epidemics with sub-populations unless you have regional data.

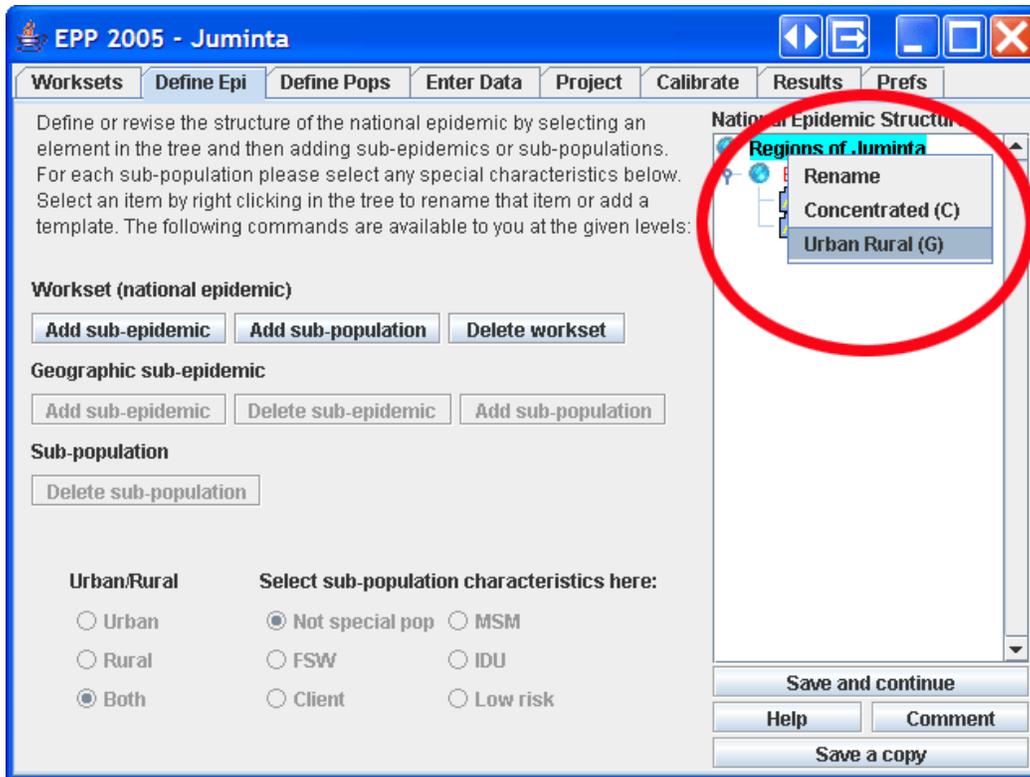


Figure 16 - Right clicking on the workset brings up a list of templates

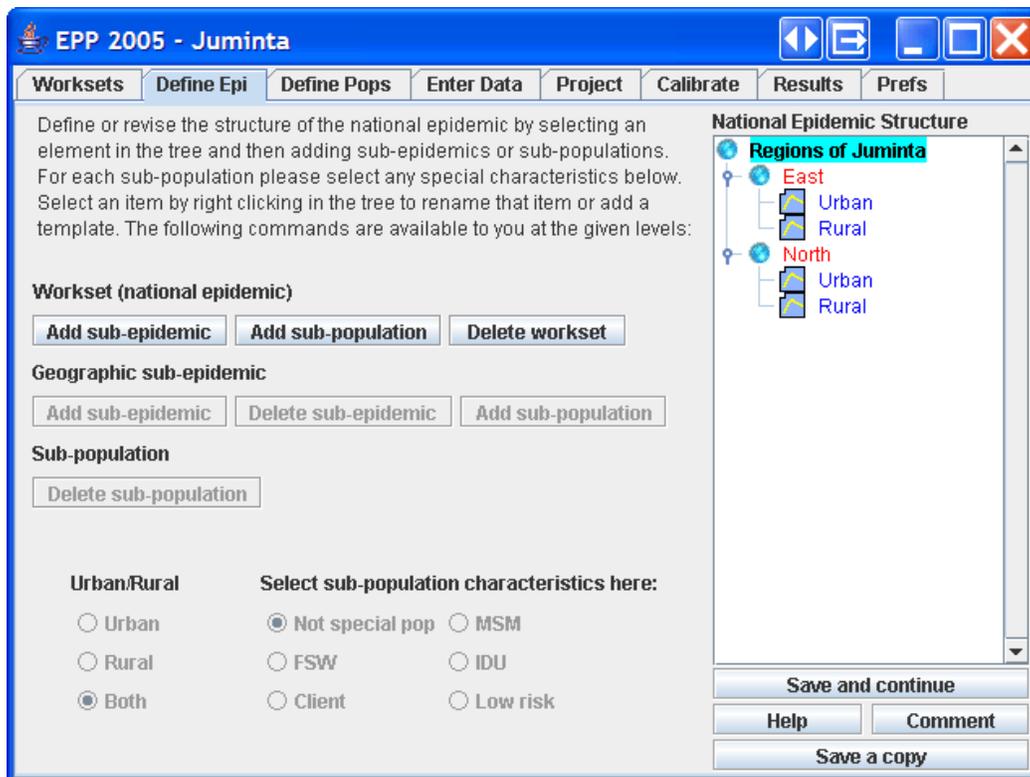


Figure 17 – Selecting one of the templates and giving it a name, adds it as a sub-epidemic

The final step – defining sub-population characteristics

Once you have built your epidemic structure you now have the opportunity to specify sub-population characteristics. Note that characteristics only apply to sub-populations, NOT to sub-epidemics.

These characteristics include:

- Region of residence
 - Urban
 - Rural
 - Both (that is, both urban and rural are included)
- Special population groups (primarily used in concentrated epidemics)
 - Not a special population
 - Female sex workers (FSW)
 - Clients of sex workers (Client)
 - Men who have sex with men (MSM)
 - Injecting drug users (IDU)
 - Low risk men or women (Low risk)

These characteristics are used later in EPP for defining demographic characteristics of sub-populations, calibration and running checks on your assignments of population and HIV prevalence levels. For generalized epidemics, the important characteristics are the Urban/Rural/Both choices. Be sure to set them appropriately. On the Calibration tab, these will be used to adjust the overall prevalence in your country to match national survey data.

To highlight and set the characteristics for a specific sub-population:

- ***Select the sub-population in the epidemic structure tree, the buttons will now become visible as in Figure 18.***
- ***Click on the characteristics you want for that sub-population***
- ***Repeat until all sub-populations have the correct characteristics.***

If you used one of the default templates to create your workset, these characteristics will already be set correctly. If you have created a new workset or a new template and have chosen to define the sub-populations yourself, you should choose the correct characteristics for each of your individual sub-populations. For generalized epidemics, you will choose either ***urban, rural or both***, along with ***Not special pop.***

Once you are satisfied with the epidemic structure, click the “Save and continue” button to advance to the Define Pops tab. Before you move on to the next screen, EPP will prompt you about whether you assigned characteristics. Click YES (if you have done so) and move on.

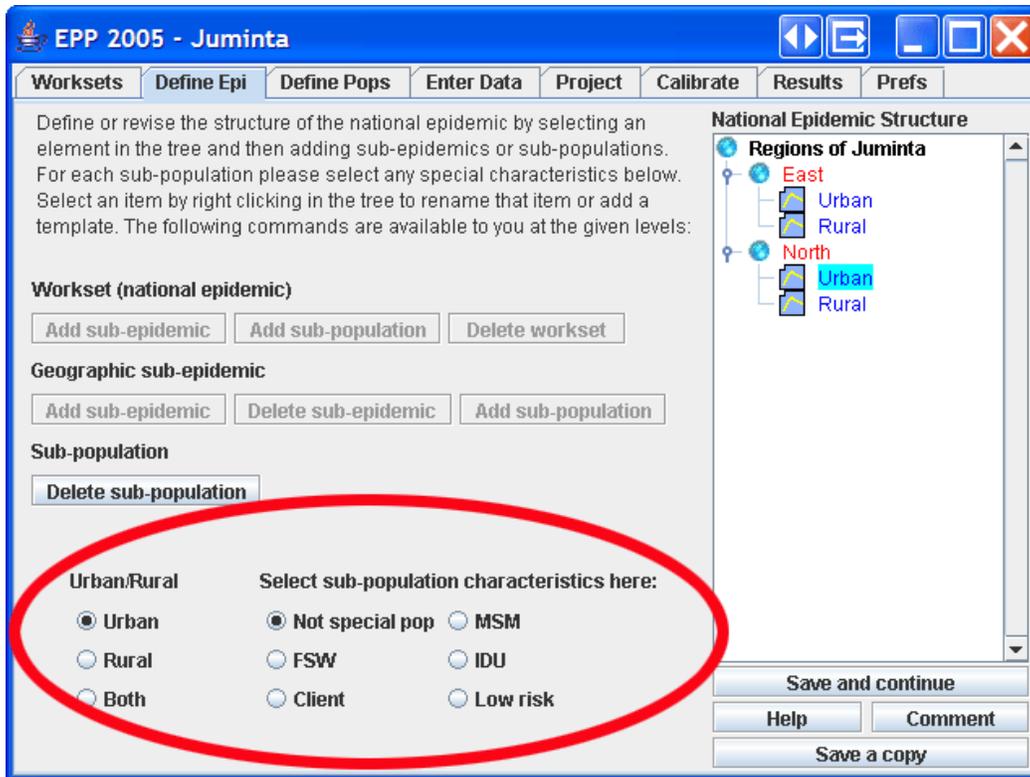


Figure 18 – The sub-population characteristics which appear when you click on a sub-population in the epidemic structure tree

The Define Pops Tab – Setting sub-population sizes

After clicking on "Save and continue" on the Define Epi tab, you will move on to the Define Pops tab (shown in Figure 19). In this tab you provide demographic information about each of your sub-populations. On entering the tab, EPP will automatically move to the first sub-population  in the epidemic structure. In Figure 19, it starts with Urban, which is highlighted in light blue.

If you are working with a country that was included in EPP's country list (from the United Nations Population Division) then you should only have to add sub-population  size data on this tab – all of the other demographic information will already be set for you. You will note that in this example, the National Population has already been set (1,309,367) along with birth rates, survival to 15, mortality and growth rate. The total national population has also been placed in the area "Population still to assign" because you have yet to add people to your Urban and Rural sub-populations. NOTE: All populations on this page are the total number of people ages 15 and older.

Notice that the Population 15 and over for the Urban sub-population, which is pointed to by the arrow in the figure, is still zero. Your job on this page is to fill in this number for each sub-population  in your workset. The total population in all of your sub-populations should sum to the national population. The number remaining to assign is displayed in the "Population still to assign" box at the top of the page.

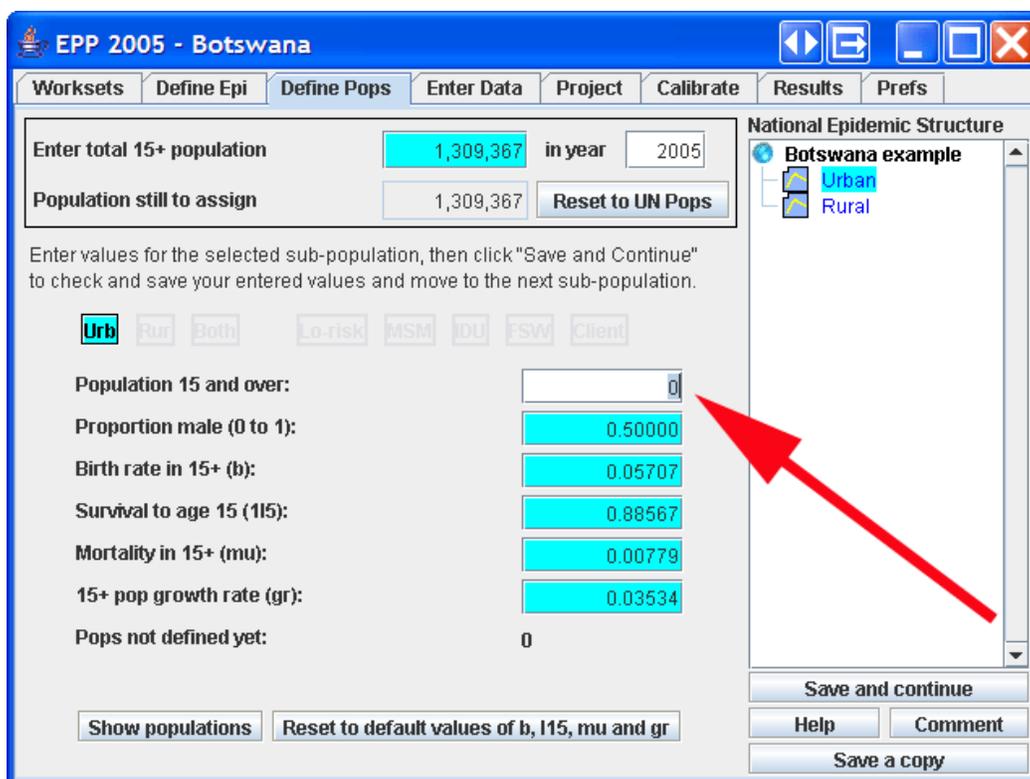


Figure 19 – Your first visit to the Define Pops tab

To enter the population for the Urban region, just type the value in the box next to the arrow in the Figure and hit <ENTER>. For example, if we type 250,000 for the urban population, the screen will now look as in Figure 20. Notice that the “Population still to assign” has been reduced to 1,059,367 which is the total national population of 1,309,367 minus the Urban population of 250,000. Now click on “Save and continue” to save the Urban population size and you will automatically be taken to the Rural sub-population. Enter 1,059,367 for the Rural sub-population size and click on “Save and continue” and you will be taken to the Enter Data tab.

If your population is not equally balanced between males and females, you should change the “Proportion male” below the sub-population size. If the sub-population is all male, this value should be 1.0. If it is all female, it should be 0.0. If you chose specific sub-population characteristics, e.g., female sex workers (FSW), that involve gender, this value will be changed accordingly.

IMPORTANT REMINDER

Always hit “Save and continue” after entering the sub-population size and other sub-population demographic characteristics to move on to the next sub-population. Otherwise the data will not be saved and you will need to re-enter it.

Once you have finished entering data for every sub-population and the population still to assign has been reduced to zero, hitting “Save and continue” will move you on to the Enter Data page.

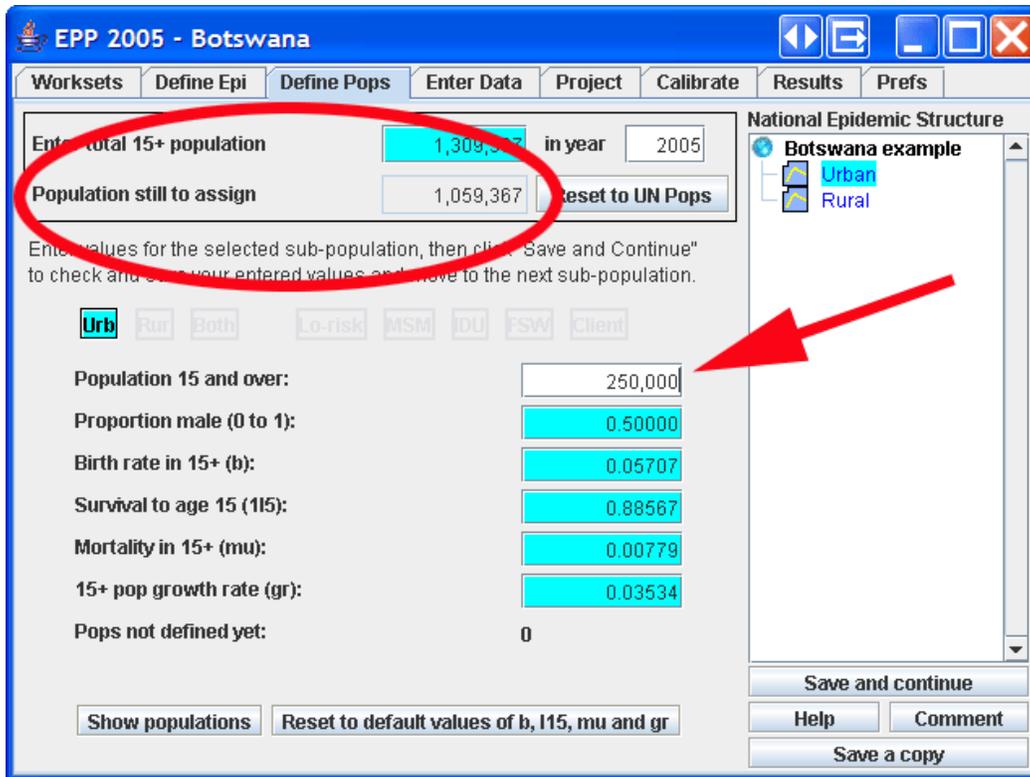


Figure 20 – After adding in 250,000 urban residents

IF YOU ARE CREATING A NEW WORKSET OR USING DEMOGRAPHIC DATA NOT AVAILABLE THROUGH UN POP

You will need to enter the following data on the Define Pops tab:

FOR THE WORKSET

- *Total population in the workset:* This is the number of adults 15 and older in the national population in the year specified in the next box.
- *Year:* The year for which the population size is provided. Generally this should be the latest year for which you have surveillance data or the current year.

FOR EACH SUBPOPULATION

- *Population 15 and over:* The number of adults 15 years and older in the sub-population.
- *Proportion male:* The percent of the sub-population that is male.
- *Birth rate in 15+(b):* The annual number of births divided by the sub-population 15 and older.

- *Survival to age 15 (I15)*: The proportion of births that survive to age 15 in that sub-population.
- *Mortality in 15+ (mu)*: The proportion of the sub-population that dies each year from causes other than AIDS.
- *15+ growth rate (gr)*: The annual growth rate of the sub-population.

When you finish entering the data for a sub-population click the “Save and continue” button to advance to the next population. Once you have defined all the populations, the “Save and continue” button will take you to the Enter Data tab.

The Enter Data Tab – Adding HIV prevalence trends

The Enter Data tab is the place to enter your HIV surveillance data. If you are using a pre-existing workset the data may already be entered. In that case you can review the data for completeness. If it is complete you can advance to the Projection tab by hitting “Save and continue” once for each sub-population, reviewing the next set of data and repeating this until you have reviewed all data. After the last sub-population, you will automatically be taken to the Projection tab.

If you are starting with no data, the screen will appear as in Figure 21.

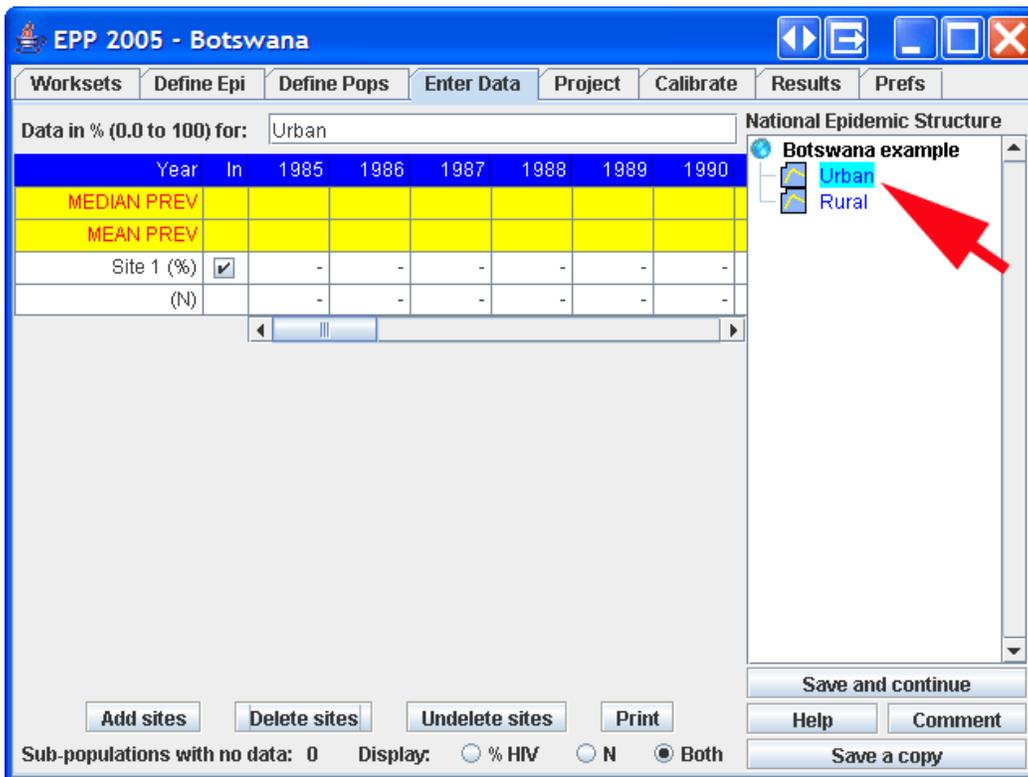


Figure 21 – The Enter Data tab before data is entered

ENTERING NEW DATA

There is a separate data entry spreadsheet for each sub-population . In figure 21, Urban is highlighted by clicking on that sub-population in the epidemic structure tree, and this spreadsheet is then the one for inputting data for Urban sites.

Surveillance data are entered in the alternating pairs of white or light blue cells of the spreadsheet. Each pair of lines refers to one surveillance site. In Figure 21, there is only one surveillance site. The first row for a given site is for the HIV prevalence, expressed in percent (that is, a number between 0 and 100). The second row for that site is for entering the sample size at that site for the given year. Use a separate pair of rows for each surveillance site, entering both the HIV prevalence and the sample size. If you do not have sample sizes, currently EPP defaults to a sample size of 300. You may over-write this value by just double-clicking in the sample size cell and typing over it. Setting a different default sample size will be discussed later in the Prefs tab section.

You can add new sites by clicking the “Add sites” button. You will be asked how many sites you wish to add and when you click “OK” that number of sites will be added to the table.

Because EPP 2005 uses a [maximum likelihood](#) approach and has added levels for estimating the best fit to your prevalence data, sites have become much more important in the analysis. Thus, if you have site specific data, you should always enter it in on this page. In maximum likelihood estimations, sample sizes serve as weights in the calculations. For more information see Appendix H or link to [BGIM: Maximum Likelihood Estimation Primer](#)

Because in EPP you may be calculating what are known as LEVEL fits, (see appendix F for more detail on level fits) all data points from one site must be input on ONE line.

Data may be copied/pasted from an Excel spreadsheet. In order to do this, you may need to turn off the N (sample size) option on this sheet. This is done by choosing “%HIV” in the Display portion of this page. NOTE: The program defaults to BOTH. See the Help page (push the “Help” button on the Enter Data tab) for more information on this.

ENTERING SITE DATA

The steps in entering data for a site are as follows:

- ***Enter the site name*** in the **first** column by clicking on the box and typing the name of your site. This is important if you want to come back and review your data for correctness later on.
- ***Decide if you want to include this site in the fits by setting the checkbox in the second column.*** There may be situation where you wish to remove a site from the fit because you think its data may be incorrect or you want to see how much that site is influencing the national fit. To remove a site from the

fitting, just double-click on the checkbox and that site will not be included in fitting. The data will still be there, but it will just not be used in the fitting procedure. If a site is de-selected, double-clicking on the checkbox will include it in the fits once again.

- **Enter your HIV prevalence data.** For each site, enter the HIV prevalence data for all years in which data was collected. If a year is missing, that's no problem, just leave the cell blank.
- **Enter your sample sizes for each year.** If you do not know the sample size, the EPP default sample size will be used. However, sample sizes do affect the fitting procedure, so include them if you have them.

Note that a prevalence estimate of 12% should be entered as 12, not as 0.12.

After entering data for a number of sites the screen might look something like this (figure 22):



Figure 22 – After entering a few year’s worth of data with the default sample size

The rows that are highlighted in yellow contain the *annual medians* and *means* for all sites entered for that year. If you hit <ENTER> or <TAB> each time you enter a value, these medians and means values are automatically updated by the spreadsheet each time a value is entered into a cells.

The color-coding of cells on this page is as follows:

- Yellow cells – are calculated by the program from the data you enter.
- Light blue or white cells (alternating for multiple sites)– are user-entered data. In this case this includes both site names and actual prevalence data values.

When you first open it, the program provides you with TWO white rows (one for prevalence data and one for sample size). Should you have other data from other sites, you should click on the ‘Add sites’ button at the bottom of the screen and add the appropriate number. Adding more sites would yield a screen like the one below (figure 23). Note that sites are color coded in alternating blue and white to help you see which HIV and sample size data goes with which site.



Figure 23 – The Enter Data tab after adding two more sites and additional data

Once you have entered all the HIV surveillance data and sample sizes for one sub-population, click the "Save and continue" button to save the data and advance to the next (if applicable) sub-population. When you are finished entering data for all sub-populations, click 'Save and continue' button and the program will advance you to the Projection tab.

Ideally you should have data for all sub-populations in the workset. If you do not you may still continue, however, EPP will prompt you with a screen which appears as

below (Figure 24) and you will be unable to do fits for the sub-populations with missing data.

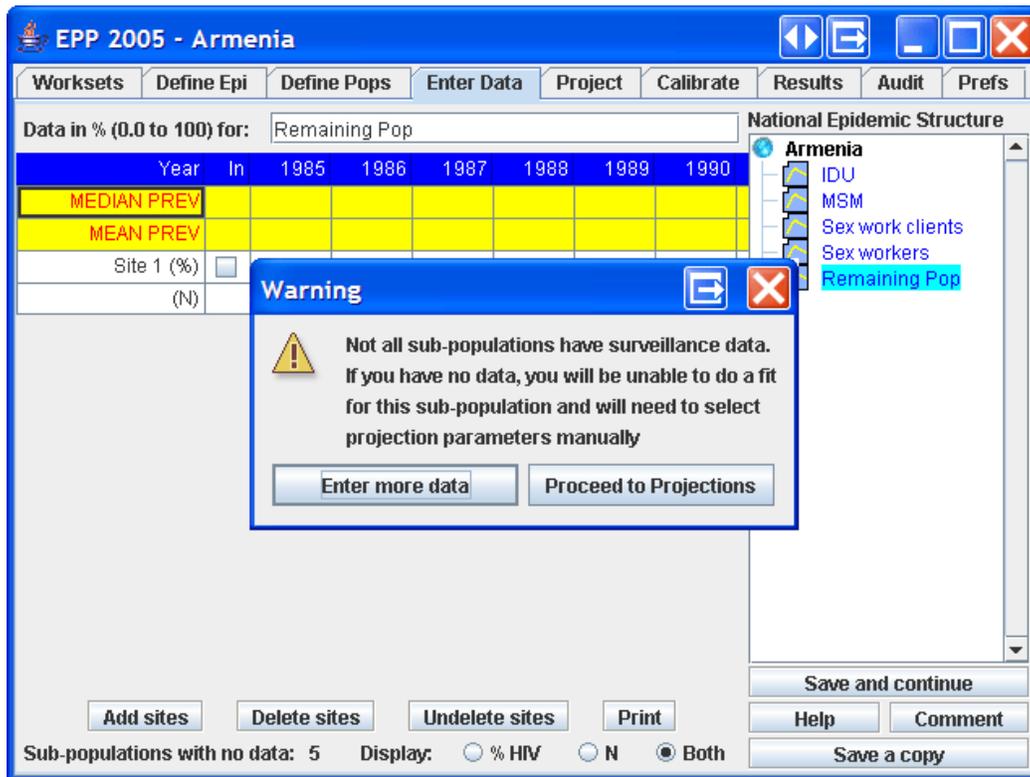


Figure 24 – The warning that appears if you do not have data for all sub-populations on the Enter Data tab

The Project Tab – Fitting your data

The Projection Tab is used to fit epidemic curves to your sub-population's  surveillance data. When you first come here from the Enter Data Tab, it appears as in Figure 25 below.

On this tab, you fit the adult HIV epidemic curve using the four parameters of the UNAIDS Reference Group model. This allows EPP to develop trends for HIV prevalence and HIV incidence in the sub-populations that compose your national epidemic. Later in the process, these individual curves will be combined to give similar prevalence and incidence trends and estimates on a national level.

The estimates produced by EPP are to be exported to Spectrum and used to develop further estimates of the impact of the HIV epidemic in your country. You should use the actual prevalence and incidence values resulting from Spectrum for your national estimates as Spectrum uses a more correct demographic model than EPP.

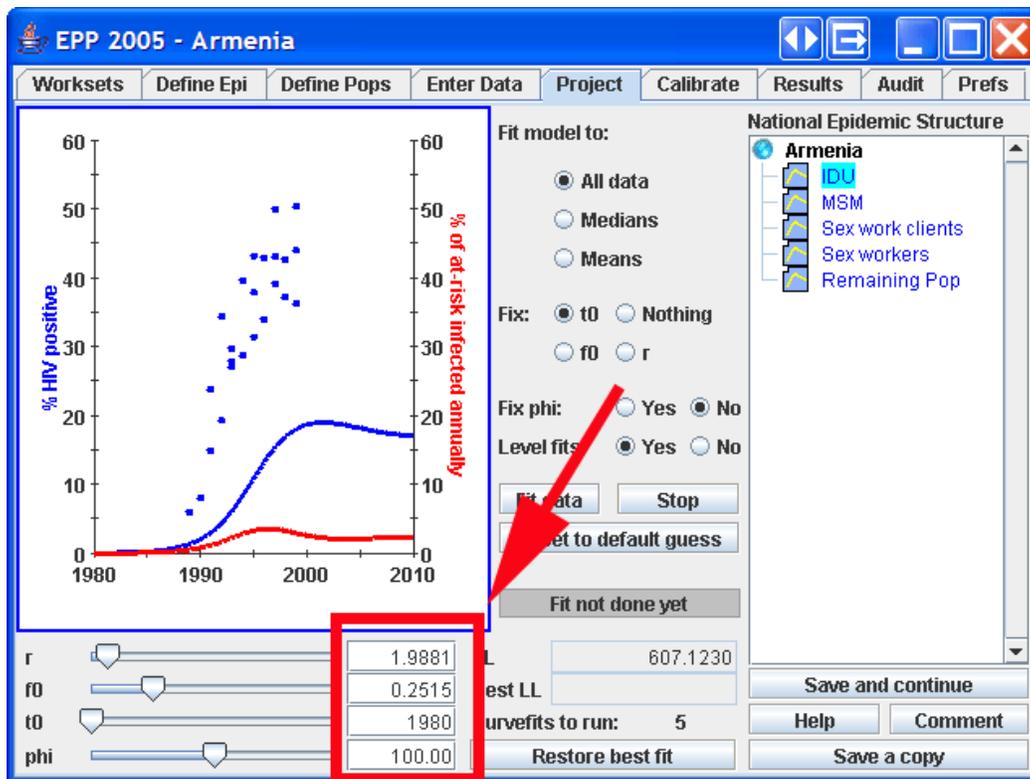


Figure 25 – The Projection tab after entering HIV prevalence data, but before fitting

Some basic components of this page and their function

The fitting parameters are shown in the red box in Figure 25. When you arrive at this page, the four parameters for model fitting (see arrow) will already have some initial guesses set. These are initial values that have been found to give reasonable fits in many situations. The four parameters, contained in the red box in Figure 25, are r , f_0 , t_0 and ϕ . These were described earlier in this document. The “Help” button on this tab also has some graphs illustrating the impact of each of these parameters if you would like more information about their effects.

The graph contains two lines and the data points (or means and medians, if that is what you are fitting) for the sub-population highlighted in light blue in the epidemic structure tree. The two lines are clearly labeled in the axes of the graph. The blue line is the HIV prevalence (percent of the sub-population living with HIV in the year on the x-axis). The red line is the incidence, defined as the percent of the susceptible (uninfected) members of the sub-population who became infected in the last year.

Sliders for interactive fitting. The area below the curve has ‘sliders’. These allow the user to interactively change the parameters and see the effects of making those changes in the graph. To adjust a slider, just click on the marker on the line and move it to the right or left. As you move it, the values of that parameter will change. This will be reflected in the text box showing the value of that parameter to the right of the sliders, and it will be reflected in the graph.

The log likelihood and best fit log likelihood. To the right of the parameters there are two fields labeled “LL” and “Best LL”. EPP 2005 uses a maximum likelihood fitting approach. EPP 2005 fits your data by iteratively changing the parameters and calculating the log likelihood, readjusting the parameters and recalculating the log likelihood, and continuing with this process until it finds the maximum value of the likelihood. This corresponds to the smallest possible value of LL.

Thus, the log likelihood is really a measure of how good your fit is and your objective is to minimize it – that is, make it as small as possible. EPP automatically fits and finds what it believes to be the set of parameters making LL as small as possible. However, many different solutions are possible and by changing the parameters yourself or starting from a different initial guess you may be able to achieve a lower log likelihood than the best fit that results from the initial guess. However, usually the initial attempt is pretty good and you’ll have to try a bit to bring it down further.

FITTING THE EPIDEMIC CURVE

Between the graph and the National Epidemic Structure panel is an area that controls the fitting of the model. Before leaving this page you will need to fit an epidemic curve to each of the sub-populations  shown in the epidemic structure. To fit the curve, follow these steps:

1. ***Decide what to fit to by clicking an option under “Fit model to:”.*** You can fit the curve to all the surveillance data points or to just the median or mean for each year. **Usually you will use all the data, and this is the default of the program.** You might want to use the median or mean if you have some data points that appear to be extraneous outliers. You might also want to try the mean or median if you have so many points that it is hard to see the trend in the data.
2. ***Decide what will be kept fixed when fitting, by selecting one of the radio buttons next to “Fix”.*** This gives the program guidance on what parameters you would like to be adjusted in fitting the curve. Fixing t_0 , the time at the start of the epidemic, is most common, though occasionally you will also decide to fix some of the other parameters ahead of time. Most of the time, you will choose to fix t_0 , simply because you have a good idea when the epidemic really started in your country. ***Fixing t_0 is the default option.***

RECOMMENDATION

For your FIRST FIT, it is recommended that you fix t_0 and phi. This provides the quickest fit.

In the most general use of the model, you will search for the best values for all four parameters: t_0 , r , f_0 and ϕ by selecting “Fix nothing”. However, this can be very time consuming and you generally know the start year of the epidemic. This is often estimated as two years before the first AIDS case report or as the first year with non-zero prevalence in ANC reports. If you do know the start year of the epidemic then it is best to select the option to “Fix t_0 .” This will allow you to find the best fitting HIV prevalence curve more quickly.

FOR ADVANCED USERS

You can also fix r or f_0 and have the model fit the other parameters. You might do this if you were fitting several similar epidemics and wanted them to all use the same values of r or f_0 . You can also choose to “Fix nothing” but the model may take a long time to find the best fit.

3. ***Decide if you want to fix phi.*** You should generally select the option to “Fix phi” initially. This will allow you to fit the model first with the default assumption that prevalence is not declining, which is generally the case with national epidemics. If you have selected the YES option for “*Fix phi*” then you should make sure that phi is set to a large value, such as 100. Currently, EPP defaults to this value. Later you can adjust phi and refit to see whether a curve with declining prevalence provides a better description.

RECOMMENDATION

DO NOT PLAY WITH ‘SLIDERS’ UNTIL AFTER A FIRST CURVE IS FIT WITH A FIXED t_0 AND phi!!

4. ***Decide whether to use level fits.*** Next, you must make a decision about ‘Level fits’. Level fits help to adjust the curve to deal with expansion of the surveillance system to include lower prevalence sites. This is described in more detail in the help on this tab. If you decide to do a level fit, EPP will first run a non-level fit to set the initial guess for the parameters to be used in the level fit.
5. ***Fit the model by clicking on the “Fit data” button.*** The program will now search for the best fitting curve for the surveillance data, displaying intermediate steps in the graph as it proceeds. When it is done, the result will be shown in the blue line in the graph and the best fit parameters will be in the parameter text fields in the red box in Figure 25.

Figure 26 shows the results of a fit with levels turned off. Three things to notice here:

- The level fits selector, circled in red, is turned off. This tells EPP to do a non-levels fit.
- The level fit indicator, circled in blue, shows that levels were not used in doing this fit.
- The best fit values of the parameters are shown in the green box.

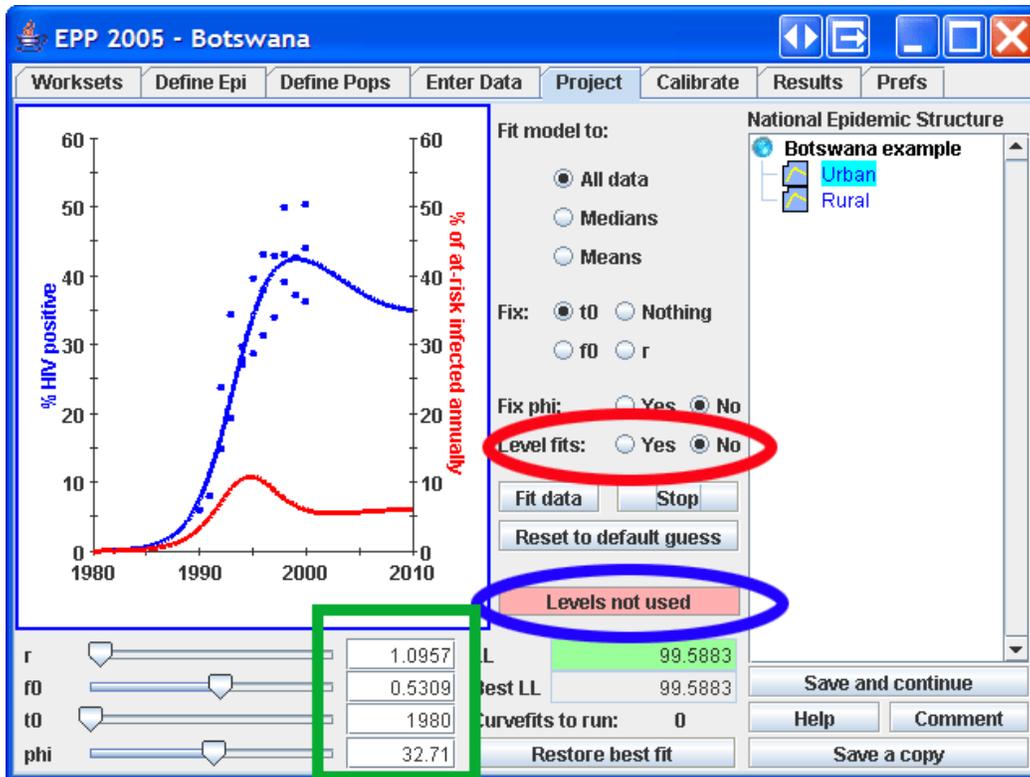


Figure 26 – Results of a fit with levels turned off.

You should note the value of LL, the log likelihood, and if you decide to alter any of the parameters, either by moving the sliders or typing new values in the green box in Figure 26, you should compare the new LL to the original. The smaller of the LLs represents the better fit.

The best LL fit will be indicated by a GREEN background in the cell for LL (as in Figure 26). If altering the parameters yields a LL that is significantly different from the original LL, the background for the LL cell will be RED. Other less important alterations to the parameters will result in a LL cell that has a YELLOW background.

Figure 27 shows the result of a level fit. Note that the level fit selector, circled in red, is set to yes and the level fit indicator, circled in blue, shows that levels were used in this fit.

You may choose to restore the best fit by hitting the “Restore best fit” button at any point.. This will bring you to the best fit value the LAST TIME you chose to FIT DATA.

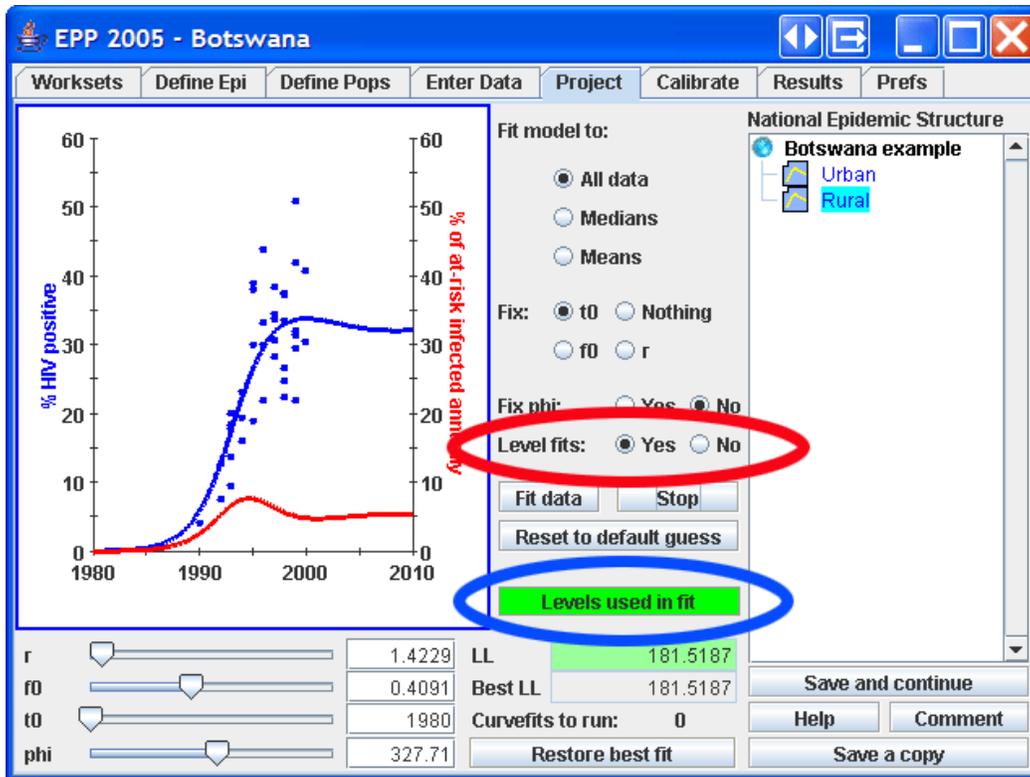


Figure 27 – Results of a fit done with levels turned on.

NOTE: With all other choices on this page you may click on ‘return to original fit’ to retrieve your initial parameter values. WITH THIS LEVEL ADJUSTMENT YOU CANNOT. Therefore, it is recommended that you save and name the first result (non-level fit) before you proceed to the LEVEL FIT option.

If any sub-populations  have data from multiple sites, setting the level fits option to YES is a reasonable decision. By doing this, the program assumes the data represents consistent reporting from individual sites over time (usually 5 years or more). This is a common assumption if you have serial data from the same site. If you have data that is not consistently reported, then set this to NO.

For more discussion on LEVEL FITS, see Appendix F.

6. **Click on “Save and continue” when satisfied with your fit.** This will save your parameters for this sub-population and move you to the next sub-population. Once you have fit all sub-population, “Save and continue” will take you to the calibration page.

Note that the program tracks the numbers of curves you have yet to ‘run’.

Exploring the effects of changing the parameters

Having run or completed a fit, it is sometimes useful to be able to see the effect of changing the model parameters by looking at a diagnostic measure. In this case, we use the Log Likelihood (LL). This can be done by entering a new value for any of the parameters to the lower right of the graph and pressing ENTER, or by using the sliding scales on the lower left hand side of the page. The model will then be recalculated using that new value and the result displayed in the graph, along with a new LL value. Generally, the smaller the LL, the better the model is a ‘fit’.

See Appendix D ‘Fitting the Model’ for more details on how to modify the fit to produce the most reasonable result.

Once you have an acceptable fit, click the “Save” button and these values will be stored in the projection file along with the best-fit values.

You should then repeat this procedure for each population in your workset. Once you have completed all the populations you can advance to the Calibrate tab.

The Calibrate Tab – Making final adjustments

In most generalized epidemics, the prevalence trends are based on data from pregnant women. While this is believed to be a proxy for prevalence in the general population, it may overestimate or underestimate prevalence depending on where the surveillance sites are actually located. If they are mostly in urban areas, as early sites often were, they are likely to overestimate national prevalence. Even “rural” sites are often peri-urban and are rarely placed in the most remote areas where a large portion of the population lives. With increasing frequency in the last few years, additional data about HIV prevalence are being collected in countries that may need to be considered before finalizing national EPP projections. These additional sources may be national surveys, regional surveys, or other sources of information on prevalence in the population at large that may be used to adjust the level of the curve fit to surveillance trend data.

On the Calibration tab EPP 2005 allows you to make adjustments based on these additional data sources. In generalized epidemics, there are four options:

- Make no adjustments.
- Adjust the urban and rural prevalence on a national scale to specified levels.
- Adjust all urban and rural prevalences in the country up or down by a scale factor.
- Adjust each sub-population’s fit to the data individually.

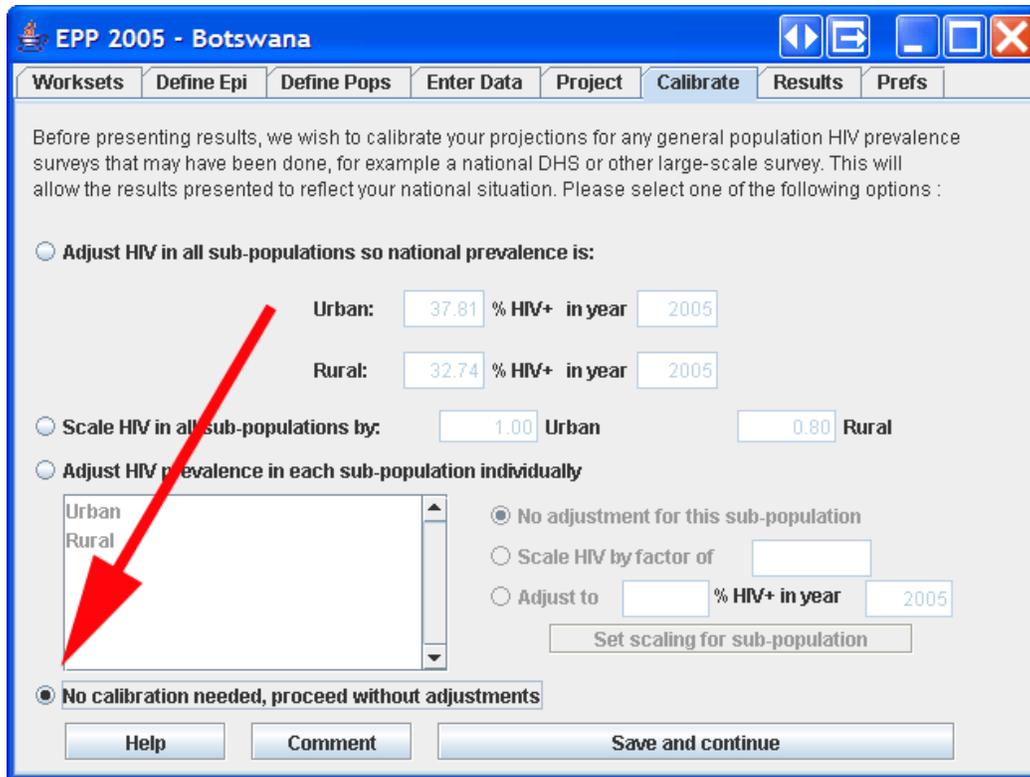


Figure 28 – The simplest choice, don’t make any adjustments

The first option is the simplest. If you have NO ADDITIONAL SOURCES OF DATA, select the option at the bottom of the Calibration tab, as shown in Figure 28. This will allow you to use your projections exactly as they appeared on the Projection tab. Then just hit “Save and continue” to move on and examine your national epidemic on the Results tab.

IF YOU HAVE ADDITIONAL INFORMATION ABOUT HIV PREVALENCE IN YOUR COUNTRY IN A GENERALIZED EPIDEMIC:

If you HAVE ADDITIONAL SOURCES OF PREVALENCE INFORMATION, e.g., representative national surveys, you may want to adjust your results before continuing analysis. This will produce estimates that better represent your national situation.

FOR GENERALIZED EPIDEMICS (URBAN/RURAL)

- The first option on the Calibration tab (figure 29) is to apply *global scaling* to the prevalence curves so that the findings of national prevalence surveys are taken into account. To use this option, you must have national prevalence values for urban and rural locations, and you must specify the year those values were collected. These data might come from national population based surveys, such as the Demographic Household Surveys (DHS). When you initially enter this page, the urban and rural prevalences are set to the urban and rural prevalence that comes from your data fits on the Projection tab.

Figure 29 – Scaling your national epidemic to match measured urban and rural prevalence levels

CAUTION: These surveys may have their own biases associated with the results, for example, large non-response biases. In that case, the results from the survey must be adjusted to account for that bias. For more information on making such adjustments, see the UNAIDS document “*Guidelines for measuring national HIV prevalence in population based surveys*”.

- The second option (figure 30) allows you to scale all the curves by adjusting the data using a scaling ratio. For example, some countries know that “rural” sites overestimate the true prevalence in rural areas by 20%. Thus they might want to adjust their rural ANC data by a factor of 0.8. This value is derived by assuming that ANC data underrepresents rural prevalence by 20%. Subtracting that 20% (or 0.2) from 100% (1.0) yields 0.8. ***This value is a recommended default for most countries with generalized epidemics.***
- A third option (Figure 31) allows you to apply the adjustment to each data fit individually. Using this option presumes that you have a great deal of data about the epidemic in your country and that the data that you entered on the ‘Enter Data’ tab requires this adjustment. Choosing “Adjust scaling of HIV to each sub-population individually” allows you to access the sub-population list in the window below. For each of the sub-epidemics in the window you then choose one of the options to the right of the window, and then select “Set scaling for sub-population”. If you do NOT select “Set scaling for sub-population”, your decision will not be recorded.

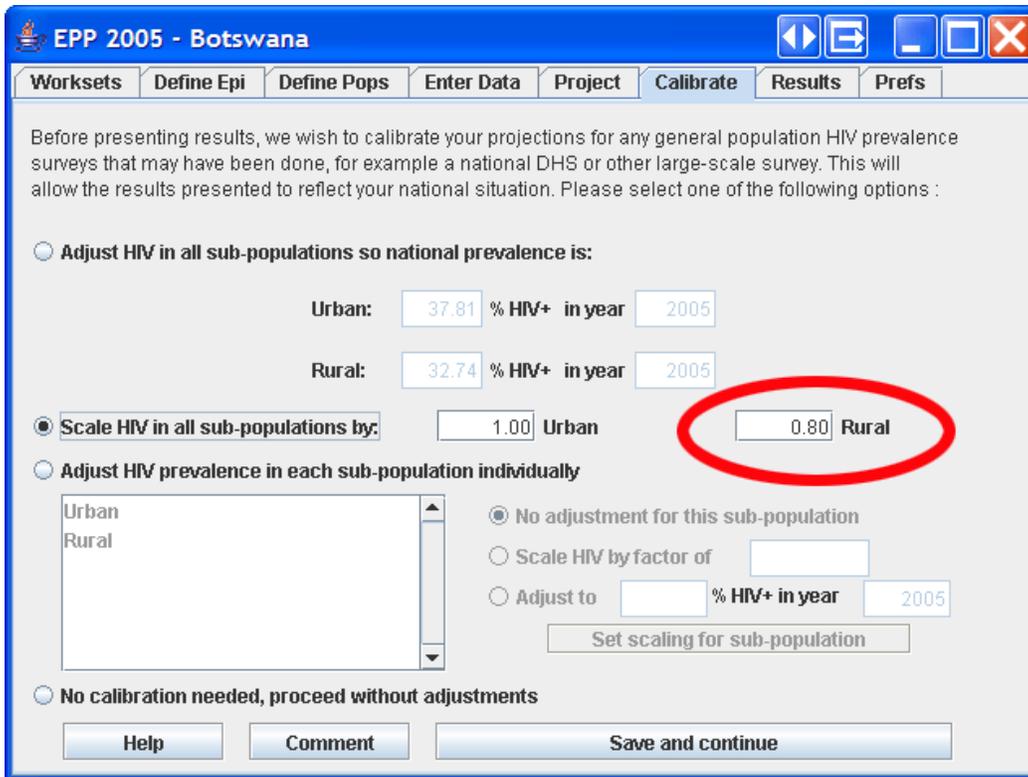


Figure 30 – Scaling all your urban and rural projections by a fixed factor

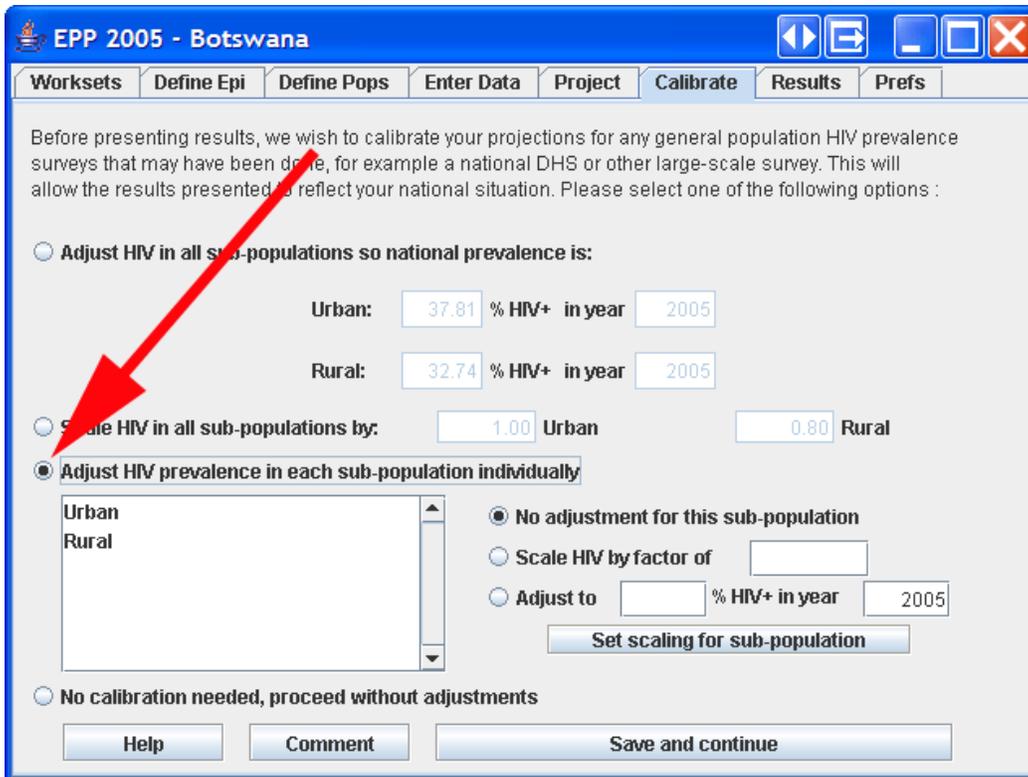


Figure 31 – Choosing to adjust each sub-population's data fit individually

Once you have made your decision on how to calibrate your data by making one of the above four choices, click “Save and continue” and you’ll be taken to the Results tab.

The Results Tab – Reviewing & saving your results

This screen displays the results of all the projections you have made. An example is provided below (Figure 32).

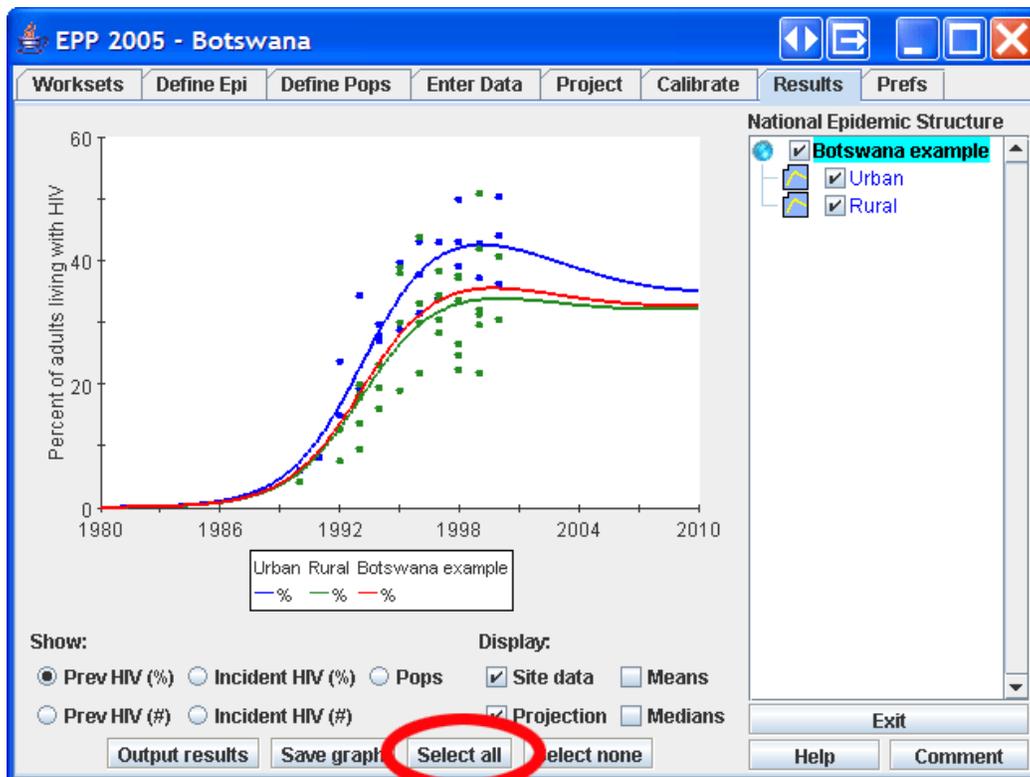


Figure 32 – A display of the national epidemic prevalence for Botswana

If you click the “Select all” button at the bottom of the page the data and projections for all sub-populations will be shown in the graph. You can then add or delete check marks next to the sub-population names  in the national epidemic structure tree to change which sub-populations are displayed in the graph.

Using the radio buttons next to “Show:” you can set the graph to show five variables:

- *Prev HIV(%)* - this gives the prevalence (in percent of the sub-population HIV+) for each selected sub-population, sub-epidemic or national epidemic,
- *Prev HIV(#)* - graphs the estimated number living with HIV for each selected sub-population, sub-epidemic, or national epidemic,
- *Incident HIV(%)* - gives the incidence, in percent of susceptibles infected each year for this sub-population or sub-epidemic
- *Incident HIV(#)* - graphs the number of new infections in a year in this sub-population or sub-epidemic

- *Population size* - shows an estimate of the absolute size of this population (number of people over age 15).

Using the radio check boxes next to “Display:” you can display or hide the “Site data”, “Medians”, “Means” and “Projection” (the epidemic curves).

To examine the results as a table, click the “Output results” button below the graph. The arrow in Figure 33 shows you where this button is. You will see an additional screen pop up with results as displayed in the figure.

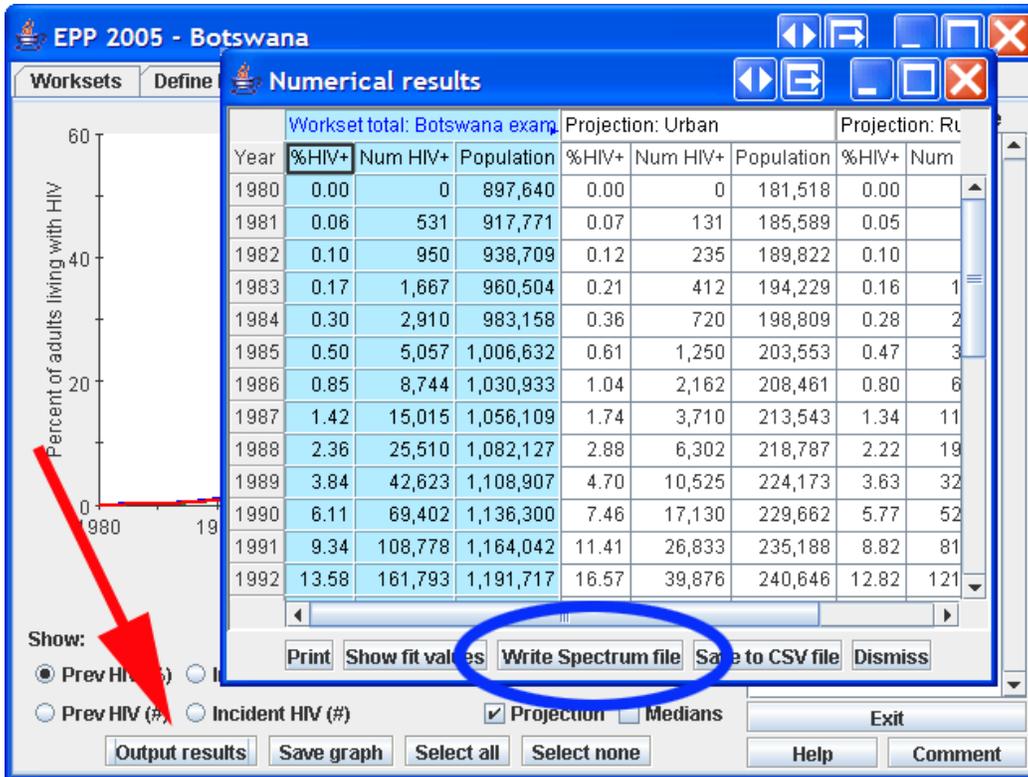


Figure 33 – The table generated by “Output results” and the button to generate the Spectrum file

To write the results to a file that can be read by Spectrum, click the “Write Spectrum File” button that is visible when you are examining the numerical results. Unless you changed this on your Workset page, by default, the program will store your output in the ‘eppout’ directory under the EPP directory. You will be requested to provide the name of the file, which will end in an “.spt” file extension.

To print the table, click the “Print” button.

You may also save a copy of the generated graph on the Results tab to a *.jpg file by clicking on ‘Save graph’ on the Results tab.

The Prefs Tab

The Prefs tab or User Preferences tab (Figure 34) is intended for users with more extensive data about the epidemic and good knowledge of modeling. There are opportunities on this page to change a number of parameters, however, one should be cautious about changing any of these as they all affect your fits.

The values that can be changed here fall into 4 categories:

- Language of the EPP Interface (currently only English is working)
- HIV modeling parameters. These relate to the survival of those living with HIV after infection, mother to child transmission and fertility reductions among HIV positive women
- Population parameters. If you change the values here, the changes will apply to all sub-populations in the workset.
- Default sample sizes by sub-population. These can be changed to reflect your local surveillance system if they use sample sizes other than 300.

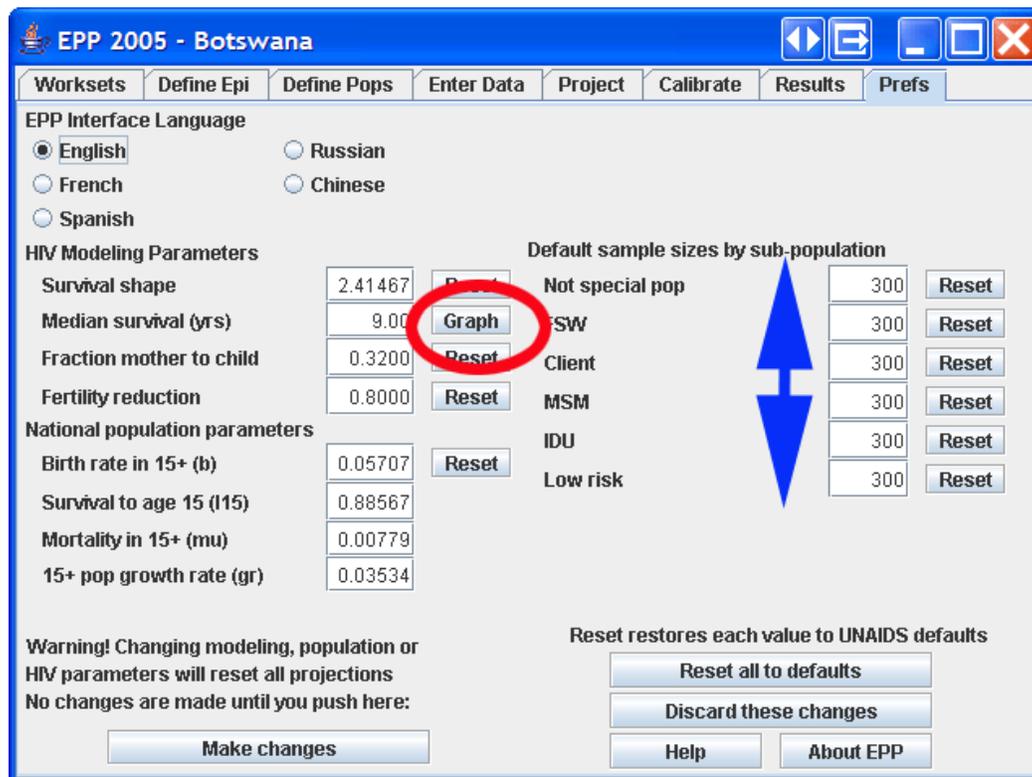


Figure 34 – The User Preferences Tab in EPP 2005

WARNING – CHANGING ANY OF THE VALUES HERE WILL INVALIDATE ALL FITS DONE TO DATE.

If you are going to change any of these, we recommend you switch here and CHANGE THEM IMMEDIATELY AFTER YOU CREATE YOUR WORKSET. Otherwise, if you enter data and make fits and then come to this page and change

them, all of your fits will be reset. This is necessary because your fits depend on each of the variables on this page (other than language, of course).

NOTE: NONE OF THE CHANGES YOU ENTER HERE WILL ACTUALLY BE MADE IN YOUR WORKSET UNTIL YOU HIT “MAKE CHANGES”. You can abort these changes by clicking on the button labeled “Discard these changes” or by hitting “Reset all to defaults”. The individual “Reset” buttons next to the various entries will reset only that entry or set of entries.

FOR ADVANCED USERS

Selecting the ‘GRAPH’ button next to ‘MEDIAN SURVIVAL TIME’ will provide a pop up window showing a graph of progression of HIV to death (see Figure 35). You may alter the values of ALPHA and BETA of the Weibull curve describing survival over time.

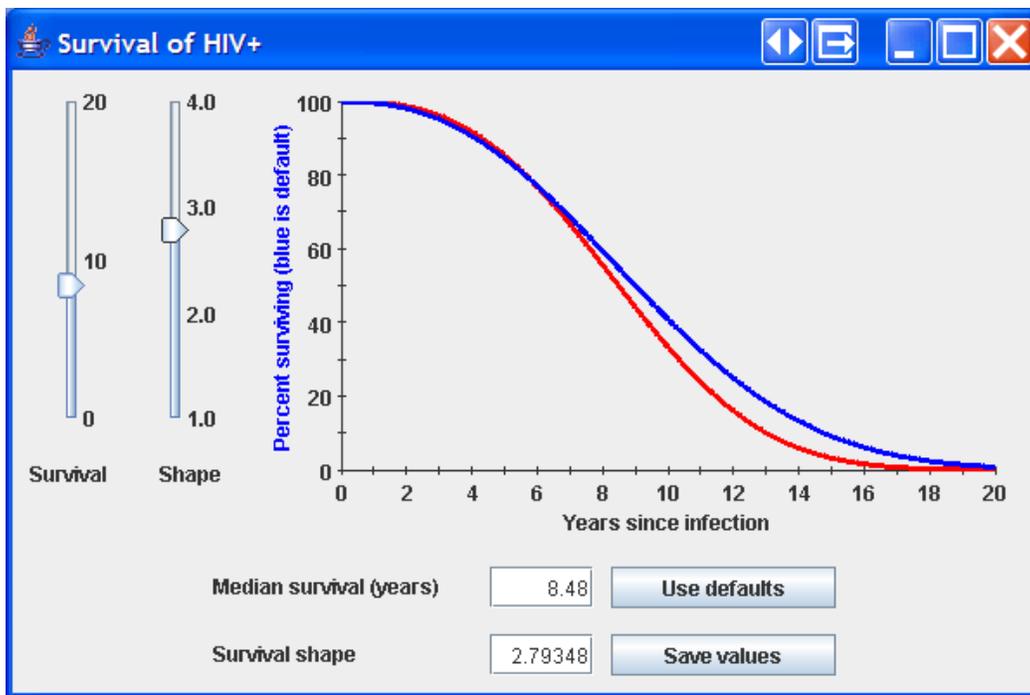


Figure 35 – The graph for adjusting survival shape and median survival

Again it is strongly recommended that only advanced users change the default values supplied by UNAIDS in EPP 2005.

Appendix A, EPP References and Further Reading A

Walker N, Stanecki KA, Brown T, Stover J, Lazzari S, Garcia-Calleja JM, Schwartlander B, Ghys PD. Methods and procedures for estimating HIV/AIDS and its impact: the UNAIDS/WHO estimates for the end of 2001AIDS. 2003 Oct 17;17(15):2215-25

Appendix B, EPP Glossary and Acronyms

Appendix B, Glossary and Acronyms

Acquired Immunodeficiency Syndrome (AIDS): The late stage of HIV infection which includes development of one or more opportunistic illnesses (illnesses that occur because of low levels of CD4 lymphocytes).

AIDS: See ‘Acquired Immunodeficiency Syndrome.’

Algorithm: Step-by-step procedure.

ART: anti-retroviral therapy

Case: An individual in the population or sample with a particular disease of interest.

CDC: See ‘Centers for Disease Control and Prevention.’

Centers for Disease Control and Prevention (CDC): The US Department for Health and Human Services agency with the mission to promote health and quality of life by preventing and controlling disease, injury, and disability.

Concentrated HIV epidemic: The epidemic state in which HIV has spread to a high level in a defined subpopulation but is not well established in the general population. (HIV prevalence is consistently >5% in at least one defined subpopulation and is <1% in pregnant women in urban areas.)

Confidence interval: The compound interval with a given probability, for example, 95% that the true value of a variable such as mean, proportion, or rate is contained within the limits.

Double Aids orphans: orphans who have lost both parents to AIDS

Epidemic: The occurrence of a disease (or other health-related event) at a level of increase to a baseline. For example, the high prevalence of HIV found in many parts of the world today, including sub-Saharan Africa, Latin America and South and Southeast Asia.

Epidemic pattern: a description of the progression from infection to death, the distribution of infection by age and sex, transmission of HIV from mother-to-child, the effect of HIV infection on fertility, and the effects of anti-retroviral therapy.

Epidemic structure: the structure (usually geographic or sub-population specific) by which an epidemic is described, such as urban/rural or low-level high risk groups.

Epidemiologic model: A formal framework to convey ideas about infectious disease transmission. Construction requires three major types of information: (a) a clear understanding of the disease life-cycle, (b) the mode and rate of transmission between individuals, and factors affecting this; and (c) population characteristics such as

demographic structure and patterns of risk behaviour. For more information, see <http://www.iaen.org/models/glossary.php>

Epidemiology: The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems.

Generalised HIV epidemic: The epidemic state in which HIV is firmly established in the general population. (HIV prevalence is consistently >1% in pregnant women.)

High-risk group: A group in the community with an elevated risk of disease, often because group members engage in some form of risky behaviour.

HIV: Acronym for ‘Human Immunodeficiency Virus.’

HIV-1: A type of HIV with slight genetic variations from HIV-2. More easily transmitted than HIV-2.

HIV-2: A type of HIV with slight genetic variations from HIV-1. Less easily transmitted than HIV-1.

HIV-negative: Showing no evidence of infection with HIV (for example, absence of antibodies against HIV) in a blood or tissue test.

HIV-positive: Showing indications of infection with HIV (for example, presence of antibodies against HIV) based on a test of blood or tissue.

HIV subtypes: Distinct lineages of HIV that contain genetic differences.

Human Immunodeficiency Virus: A retrovirus that causes AIDS by infecting T-cells of the immune system.

IDS: See ‘Integrated Disease Surveillance.’

IDU: Acronym for ‘injecting drug user.’

Incidence: A measure of the frequency with which an event, such as a new case of illness, occurs in a population over a period of time. The denominator is the population at risk; the numerator is the number of new cases occurring during a given time period.

Low-level HIV epidemic: The epidemic state in which HIV has never spread to significant levels in any subpopulation, although HIV infection may have existed for many years. (HIV prevalence has not consistently exceeded 5% in any defined subpopulation or in the general population.)

Maximum likelihood; the simple principle of maximum likelihood parameter estimation is to find the parameter values that make the observed data most likely. For more information see http://statgen.iop.kcl.ac.uk/bgim/mle/sslike_3.html

Mean: The measure of central location commonly called the average. It is calculated by adding together all the individual values in a group of measurements and dividing by the number of values in the group.

Mortality rate: A measure of the frequency of occurrence of death in a defined population during a specified interval of time.

Mortality rate, infant: A ratio expressing the number of deaths among children under one year of age reported during a given time period divided by the number of births reported during the same time period.

MSM: Acronym for ‘men who have sex with men.’

Parameter: One of a set of independent variables that express the coordinates of a point

PMTCT: Acronym for ‘prevention of mother-to-child transmission.’

Point prevalence: The amount of a particular disease present in a population at a single point in time.

Population: The total number of inhabitants of a given area or country. In sampling, the population may refer to the units from which the sample is drawn, not necessarily the total population of people.

Population sub-group: A group within a population that share certain characteristics or behaviours.

Prevalence: The proportion of persons in a given population with a disease or condition at a given point in time.

Rate: An expression of the frequency with which an event occurs in a defined population.

Risk: The probability that an event will occur; for example, that an individual will become ill within a stated period of time.

Risk factor: An aspect of personal behaviour or lifestyle; an environmental exposure; an inborn, inherited, or demographic characteristic. Associated with an increased occurrence of disease or other health-related event or condition. For example, injection drug use is a risk factor for acquiring HIV.

Sexual transmission: Transmission of an infectious agent, such as HIV, that occurs predominately through unprotected vaginal or anal intercourse, and less frequently through oral intercourse.

Sub-population: See ‘population sub-group.’

Surveillance: The systematic collection, analysis, interpretation, and dissemination of health data on an ongoing basis, to gain knowledge of the pattern of disease occurrence and potential in a community, in order to control and prevent disease in the community.

Time trend: changes over time

Transmission: Any mode or mechanism by which an infectious agent is spread through the environment or to another person.

Trend: A long-term movement or change in frequency, usually upwards or downwards.

Variable: Any characteristic or attribute that can be measured.

Appendix C, Useful Links

<http://www.unaids.org/en/default.asp>

<http://www.unaids.org/en/resources/epidemiology.asp>

<http://www.unaids.org/en/resources/publications.asp>

<http://www.unaids.org/en/resources/terminology.asp>

<http://www.unpopulation.org>

Appendix D. Additional Details on the EPP Model..... D
 FITTING THE MODEL

E. Fitting the model

When you fit a model in the “Project Tab” you need to examine the results to see if they are reasonable. In some cases the program may show a curve that does not fit the data well at all. This is most likely to happen when there are few data points. In this case, you will need to customize the fit. You can do this by specifying different initial values or by fixing certain parameter values. The most appropriate action to take will depend on the problem encountered. Table 1 shows common problems and the most likely solutions.

Table 1. Common problems and likely solution

Problem	Possible solutions
Prevalence curve is too high early in the epidemic (it passes well above the surveillance data points)	<ul style="list-style-type: none"> • Use a later start year, to • Use a lower value of r
Prevalence curve is too low early in the epidemic year, to	<ul style="list-style-type: none"> • Use an earlier start • Use a higher value of r
Peak prevalence is too low	<ul style="list-style-type: none"> • Use a higher value of r • Use a higher value of fo
Peak prevalence is too high	<ul style="list-style-type: none"> • Use a lower value of r • Use a lower value of fo
Prevalence declines too rapidly after the peak phi	<ul style="list-style-type: none"> • Use a higher value of
Prevalence does not decline enough after the peak	<ul style="list-style-type: none"> • Use a lower value of phi

The program can be used to customize the curve to provide a better fit to the surveillance data. Two different approaches can be used.

- **Change the initial values.** You change the initial values of any of the

parameters (in the boxes under “Initial guesses for model parameters”) and then re-fit the curve. This causes the model to start from a different point that may allow it to find a better fit. This is not always the case, however. Many times the model will find the same fit no matter what the starting conditions are.

- **Fix the values of model parameters.** Instead of allowing the model to vary all the parameters you can specify values for some of them. There are two ways to do this.

- o **Refit.** You can specify that either **to**, **fo**, or **r** should remain fixed at the initial value by using the “*Fix r*, *Fix to* or *Fix fo*” radio buttons. Select the parameter you want to remain fixed. Be sure that the initial value for that parameter is set as you want it. Then click the “Fit Model” button. The model will try to find the best fit by varying the other parameters.

Examining multiple start years. The default setting of EPP is to have the start year fixed. If you are not sure about the start year of the epidemic, you can let EPP search for the best start year. To do this, change the selection of the radio button from “Fix to” to “Fix nothing”. This allows all four parameters to vary. The values tried for start year of the epidemic will vary from the start year of the projection to the first year before there is non-zero prevalence in the data file.

Refitting with a different start year. If the epidemic curve seems to start too early or too late, you may want to change the start year of the epidemic. To change the start year type a new value in the box for to and hit Enter. Click the “*Fix to*” box and click the “Fit Model” button. This will use your new set of initial guesses and rerun the fitting algorithm.

Improving the fit by modifying r, fo or phi. Depending on the problem you are facing you may want to try difference values of **r**, **fo**, or **phi**.

- A lower value of **r** will produce a slower epidemic while a higher value will produce a faster one. Typical values range between 0 and 10 with most around 0.5 to 2.5.

- A lower value of **fo** will reduce the peak of the curve and a higher value will raise it. Typical values are between 0 and 0.40.

- A lower value of **phi** will cause the curve to decline more rapidly from its peak while a higher value will cause it to decline less. Typical values are between -2 and 100.

To make any of these changes you can try changing the initial values and then refitting the curve to see if that makes any difference. If that does not work, you can change any of these values directly in the boxes to the right of the graph.

Remember to press “Enter” after making a change. The model will recalculate the curves with these new parameter values. In many cases you will be able to tell if the new curves is better or worse just by looking at it. A more precise measure is to look at the figures displayed next to the “LL” label. This value is the Log Likelihood. (For more discussion of log likelihood models, see http://statgen.iop.kcl.ac.uk/bgim/mle/sslike_3.html .) The best fitting curve is the one with the lowest log likelihood. You can compare two different values for a parameter by seeing which one produces a lower log likelihood.

Working with limited data. The EPP model will generally do a very good job of finding an epidemic curve to fit your surveillance data if lots of data are available for all parts of the epidemic. However, the model cannot determine the shape of the curve when no data are available. For example, if the surveillance data do not show any downward trend after the peak, then the model will have no way to determine the value of **phi**, the parameter that governs the shape of the curve after the epidemic. Similarly, if the surveillance data indicate that prevalence has not yet peaked, it will be difficult for the model to determine the best value of **fo**, the parameter that determines the peak of the curve. In these cases, almost any value of **phi** or **fo** will produce an equally good fit (e.g., a similar value for the sum of squared errors).

In these cases human judgment must be used. You will have to try different values of the parameters, look at the shape of the epidemic curve and select the one that seems best to you.

In most surveillance sites in mature epidemics prevalence rises to some peak value and then either stabilizes at that value or declines slightly. As a result, if you have no surveillance data past the time of peak prevalence, it is generally best to use a high value of **phi** (50 or more) so that prevalence will remain relatively stable after the peak.

If prevalence has not yet peaked, it will be difficult to determine the best value for **fo**. Today, national prevalence ranges from 0 to around 35 percent. Values of **fo** above 0.35 would be very extreme.

Determining whether prevalence is declining. The epidemic model underlying EPP uses four parameters to fit an epidemic curve to the surveillance data. The phi parameter determines how rapidly prevalence declines from its peak. In most cases, the best fitting curve will be one in which prevalence declines significantly. This is because the extra degree of freedom will allow EPP to find a better mathematical fit when the epidemic curve changes direction than when it stays constant at peak prevalence. This does not necessarily mean that prevalence is, in fact, declining. You need to use your own judgment and consider additional factors before deciding whether prevalence is actually declining or not.

Appendix F. Additional Details using Level Fits..... F

Why LEVEL FITS?

For most countries, early surveillance for HIV developed around locations where HIV prevalence was highest. Over time as infrastructure developed, more and more sites for surveillance for HIV prevalence have been added, and consequently, surveillance has moved to lower and lower prevalence sites.

Level fits is a method to attempt to adjust for that change in surveillance. Without level fits, newer lower prevalence sites may falsely drive your epidemic curve down in later years of the epidemic. Using level fits takes this in to account by equalizing the impact of new information.

Level fits assumes that the epidemic curves from different sites retains similar shapes over time. The curve may grow at a different rate, but the curves peak at the same time. All curves from all sites are summed to get a TOTAL prevalence.

By using level fits in EPP, EPP creates a more realistic curve that accounts for new information. It is a method for dealing with new prevalence in surveillance with low levels of HIV prevalence.

In many situations, it makes no real difference.