

**Syphilis Interventions Towards Elimination (SITE) model:**

**Training Exercises to manipulate inputs and use outputs**

*Version 09 July 2020* 

This document provides a series of training exercises designed to show how to use the SITE model input and output files.  The training exercises are divided into three parts: first (exercises 1-3) the output file, second (exercise 4) the prevalence calibration data, and third (exercise 5) input file.  There is also an annex that reviews typical problems or challenges with the model fit/calibration to country data and suggests input parameters to alter to improve the fit.

**Exercise 1: Plot Syphilis indicator data, in time trend line graphs**

To start this exercise, please open in Excel the training example country output file:

***Syphilis SITE model\_User interface OUTPUT\_29June2020.xlsx.***

The Worksheet ***Trend graphs*** allows you to display model results on a time trend line graph, for the indicator, population group(s) and calendar years of your choice.

* 1. In sheet ***Trend graphs***, generate a graph of *Incident cases*, over years 2000-2030, for 7 groups: Women, Low risk; Men, Low risk; Women, Medium risk; Men, Medium risk; Women, High risk (FSW); Men, High Risk; and MSM. Use the drop-down menus in Cells:
		+ L2-L3, for start and end year
		+ L4, for the indicator
		+ K7-K13, for the population group(s) to display.

Q1A. Which two population groups cover the most incident cases, in 2018?

Answer to Q1A. Women, Medium risk, followed by Men, Medium risk.

Then, change the graph to show just 3 lines: All women, All men, and All women + men.

* 1. Next, change the graph to instead show the *Number clinically treated*.
	2. Generate a graph of the *Prevalence %, RPR+, TPHA+*. For proper display, change the y-axis to be percentages instead of numbers, since prevalence is in the range of 0-1 i.e. 0-100%:

Put the mouse/cursor on the Y-axis, so it shows ‘Vertical (Value) Axis’; click-right with the mouse, then in the drop-down menu select Format Axis. A new pop-up menu will open, under Axis Options, change Maximum and Units (major) such that the results lines are clearly readable. Further down the menu, under Number, are options for changing the number format.

* 1. Finally, change the groups shown, from All women, All men and All women + men, back to the 7 sub-groups: Women, Low risk; Men, Low risk; Women, Medium risk; Men, Medium risk; Women, High risk (FSW); Men, High Risk; and MSM.

Q1B. Now look at the Natural History Flow Chart (screenshot below). Which compartments in the natural history flow chart does ‘*Prevalence %, RPR+ TPHA+’* correspond to? & Which compartments are covered by the indicator ‘Prevalence %, TPHA+?

Answer to Q1B. Prevalence %, RPR+ TPHA+ covers Compartments, 2, 3, 4 and 5.

Prevalence %, TPHA+ covers Compartments, 2, 3, 4, 5 and 6

Q1C. Which two population groups have the highest prevalence, in 2019? Are these the same groups with most incident cases? If not, why not?

Answer to Q1C. MSM in 2019. Up to 2004 Women, High Risk (FSW) had highest prevalence and after 2004 MSM. The number of incident cases in each of these groups is small, however, owing to their relatively small contribution to the total population.

**Natural History Flow Chart, in the syphilis model:**



|  |
| --- |
| **Tip:**The XLS has just 1 single graph, which adapts with the user-selected indicator, population groups and years. If you want to save a graph before producing a next one, copy it (Control + Copy) and paste using *Paste-Special*, selecting format: *Enhanced meta-file, TIFF* or *PDF* into a new Excel sheet, or into Word or Powerpoint. In this format, the graph will be frozen.  |

**Exercise 2: Review and adapt the Summary Results table**

To start this exercise, please open in Excel the training example country output file:

***Syphilis SITE model\_User interface OUTPUT\_29June2020.xlsx.***

The worksheet ***Summary+Cost-Effect*** provides a summary of the results for a specific user-defined year. Please note, that “Number Screened with RPR’ is the number of people screened with RPR only or with a RPR + TPHA combined algorithm as, in the model all compartments positive on RPR (Compartments 2, 3, 4 and 5 in the Natural History Flow Chart) are also TPHA-positive.

2.1 In sheet ***Summary+Cost-Effect***, set the Year to 2019 (Cell C1) and review results.

2.2 Change the Year to 2010 (Cell C1) and review the results again.

Q2A. Which indicators change most markedly, and in what direction?

Answer to Q2A. Contacts traced and diagnosed, in this projection started post-2010.

Population numbers increase, due to population growth; prevalence and incidence (rates and incident case number) decreased from 2010 to 2019, as does the number of cases clinically treated.

2.3 Look at the numbers screened (K3-K16) and number diagnosed positive on screening with RPR (L3-L16). Now change the indicator from ‘***Number Screened with RPR’***(Cell K2) using the drop down menu into *‘****Number Screened with TPHA only*’**. Also change the ‘***Number Screened Positive with RPR’*** (Cell L2) into ‘***Number Screened Positive with TPHA only’.***

Q2B. Do the new numbers in Cells K3-K16 and L3-L16 result make sense? Are results for screening with TPHA only relevant for the projection scenario shown?

Answer to Q2B. In the example projection, all screening was done with both RPR and TPHA tests and as a result there is no coverage of screening with TPHA only. Therefore, both the Number Screened and the Number Screened Positive with TPHA are 0 for all groups. This indicator is not relevant for this country scenario projection.

**Exercise 3: Review and adapt unit cost and program cost**

To start this exercise, please open in Excel the training example country output file:

***Syphilis SITE model\_User interface OUTPUT\_29June2020.xlsx.***

Service delivery unit cost assumptions can be found in Rows 24 to 33 of the worksheet ***Summary+Cost-Effect***. These are combined with model-projected volumes of treatments, screenings and contact tracings (Cells J3 to N16) to generate annual program costs for the country. Annual program costs for each interventions are recorded in Row 19. The total cost of all interventions can be found in Cell Q19.

3.1 In sheet ***Summary+Cost-Effect***, review the service delivery unit costs in Rows 24-33. Are they reasonable for your country?

3.2 In Row 19, review the resulting annual program cost (by intervention, and in Cell Q19 the total of all syphilis interventions).

3.3 Change one unit cost (e.g. for Screening with RPR, e.g. from default US$ 2.1 to US$ 1.0) and see how the program cost changes.

**Exercise 4: Review and update prevalence (calibration) data**

To start this exercise, please open in Excel the training example country prevalence data file:

***Prevalence calibration data template\_20June2020.xlsx***

For the first data point, in Row 2, an ANC sentinel survey in 1989, verify how the observed prevalence gets adjusted for the diagnostic test used.

Rows 4, 6, 7, and 8 present data points that are not included in the fitting, and therefore show in grey font.

Q4A. Why are these data excluded? What is the reason to still show them in the data sheet?

Answer to Q4A. Data were excluded as they duplicate other data (see Column K). The data are shown, to indicate that the user has not overlooked them, but rather excluded them for a specific reason.

Back to the first data point, in Row 2, an ANC sentinel survey in 1989: Column K states that this data point has been interpreted as indicative of Low-Risk and Medium-Risk women. This is how all ANC women have been handled.

**Exercise 5: Review and adapt the country input file**

To start this exercise, please open in Excel the training example country input file:

***Syphilis SITE model\_User interface INPUT\_29June2020.XLSX***

This file has 5 worksheets that users can edit, with their names at the bottom shaded in green. In these 5 worksheets, cells shaded in blue are global parameters and should not be edited. Cells shaded in green are country-specific and can be changed.

The input file also contains 3 worksheets (which, if unhidden, show with the sheet names at the bottom in red-shade). These interact with the R code and should NOT be touched by the user.

Sheet ***Parameters***

The Natural History (biomedical parameters) start in Row 19. Look at Stage durations (Rows 20-24).

Q5A. Why do you think 4 of the 5 are fixed, global parameters and only one, the duration of the Latent phase (Cell C22), is country specific and can be edited by the user)?

Answer to Q5A. Stage durations are biomedical parameters, which in principle do not vary between populations or countries. However, the duration of the Latent stage (Row 22) reflects the natural history interacting with country-specific rates of incidental cure of the syphilis infection, due to antibiotic treatment of other conditions (e.g. pneumonia). Therefore, in countries with high access and usage of (exposure to) antibiotics, this duration will on average be shorter – independent of their coverage of syphilis-specific treatment and screening.

Q5B. Why are the transmission probabilities indicated as user-editable? And what setting-specific determinants of transmission probabilities may influence the values?

Answer to Q5B. Sexual practices, coinfections with other STIs (e.g. HIV/STI acting as cofactor mutually facilitating transmission) and general health status vary between countries and may all have an impact on transmission rates.

Sheet ***Screen+Treat***

Review the coverage Screening with ‘***RPR&TPHA or RPR only’*** (Rows 3-16)

Note: A value of, for example, 0.20 means 20% of that population group is screened that year.

Q5C. Among Low-Risk Women, in which year(s) does screening coverage increase?

Answer to Q5C: From 1996 onwards coverage gradually increases until 2015, then it is stable.

Q5D. From the current/default coverage, e.g. 15% in MSM in 2019, how would one specify a coverage of 100% of MSM each twice yearly?

Answer to Q5D. As 2.0 or as 200%.

Review Clinical treatment coverage, the ***Proportion Treated, of Symptomatic Cases of Primary/Secondary syphilis*** (Rows 35-46).

Q5E. Why is there a second set of clinical treatment coverages below, the ***Proportion of Primary+Secondary cases treated due to symptoms****,* in Cells B48-CD56? Why are these values grey, i.e. not to be edited by the user? Why do the formulas in Cells B51 to CD60 link to sheet ‘Parameters’?

Answer to Q5E. Clinical treatment coverage is specified as the product of:

1. Proportion of symptomatic Primary+Secondary cases treated (sheet ScreeningRates)
2. Proportion of Primary+Secondary Syphilis cases that are symptomatic (sheet Parameters).

Sheet ***ContactTracing***

For each of the 7 risk groups, the effective coverage of contact tracing is the product of:

* ***% of index cases (treated in Prim\_Sec) offered contact tracing*** and
* ***% of partners traced, from index cases offered contact tracing, by group of index case***

Change in Year 2021 the ***% of index cases (treated in Prim\_Sec) offered contact tracing*** from the default value of 20% to 50%, for 1 group of your choice. Verify in sheet ***Screen+Treat,*** Year 2021: Rows 61-69, that the effective coverage for that group has changed accordingly.

Sheet ***PopulationSizes***

Review the adult population distribution over the risk groups, as percentages (Rows 5-16) and in absolute numbers (Rows 21-30).

Q5E. What is the % of adult women who are FSW? Why is this % constant throughout 1970-2050?

Answer to Q5E. 1.6%. This reflects that the country represented had only done 1 population size estimation; the % resulting from that estimation was extrapolated to all earlier and later years.

Note that, besides the 7 risk groups for which syphilis outputs are generated, the population modelled includes 2 additional groups:

* Not Sexually Active Women
* Not Sexually Active Men.

These groups are not exposed to syphilis, so their incidence and prevalence – by definition – is 0.

Q5F. Why are these groups nevertheless included in the input file?

Answer to Q5F. They form part of a national population. So even though they do no not contract syphilis, they should be counted in the denominator when estimating national incidence rates and prevalence.

Sheet ***SexualBehaviour***

Review ***Numbers of partners per year*** for the different groups (Rows 5-13).

Q5G. These parameters are indicated as user-editable (green-shade) for Medium- and High-risk groups and MSM. Why are they indicated unchangeable (grey font) for Low-risk women and men?

Answer to Q5G. The definition of being Low-risk in the model is, having only 1 stable partner in a year.

Review the ***Number of Sex Acts per Partner per year*** (Rows 17 to 20)

Q5H. Why are there 4 rows only, and not men and women separately?

Answer to Q5H. Within any given partnership both partners must have the same number of sexual acts.

Review the ***Marital status/Proportion married*** for the respective groups (Rows 25-33).

Q5I. What proportion of MSM are married, i.e. have a relationship with a Low-risk woman? Does this seem a reasonable value for your country, in view of hidden homosexuality and stigma and bisexuality?

Answer to Q5I. 20%. In most countries, a non-negligible part of MSM are married with a woman (i.e. are bisexual) – thus bridging into the general heterosexual population.

Review ***Condom Use Per Sex Act, % of contacts (random)*** (Rows 52-55).

Review ***Condoms used*** (Rows 65-72).

Verify the calculation / formula linking to the indicators in the rows above.

Q5J. Which 4 underlying parameters determine the number of condoms, by each group?

Answer to Q5J. Population size, Number of partners per year, Number of sex acts per partner per year, and Condom use per sex act.

Having reviewed all input parameters and options, you are ready to use the model. Before running program scale-up scenarios, you may want to evaluate the model’s fit or calibration against the calibration prevalence data, and if needed improve that fit. The Annex provides guidance on how to improve the model’s fit, for a given set of country prevalence calibration data.

**Annex. Improving the baseline calibration/fit**

This annex provides some guidance on how to improve the model’s fit or calibration to a given set of country prevalence calibration data.

Model fit or calibration here means: the prevalence outputted by the model, matches the prevalence in the data – for the different groups and over time.

The table below summarizes typical problems or challenges with the model fit/calibration to country data, in the left-hand column. The right-hand column lists which input parameters are available and recommended for manipulation to improve that fit.

Row numbers of the SITE input file quoted in the right-hand column refer to the input file version also used in the exercises above: ***Syphilis SITE model\_User interface INPUT\_29June2020.xlsx***

| **Problem with model fit/calibration to prevalence data** | **Which input parameters to change** – in order of: most powerful/relevant/appropriate |
| --- | --- |
| Prevalence too high or too low, across most or all groups, and throughout time | * Per-Sex-Act Transmission Probabilities (sheet ***Parameters***, Rows 34-36)
* Size of high-risk populations esp. FSW and clients, relative to Non-active and Low-risk populations (***PopulationSizes***, Rows 8 & 15)
 |
| Prevalence too high or too low, MSM | * Per-Sex-Act Transmission probability (Man to Man) (***Parameters***, Row 36)
 |
| Prevalence declines too much or too little, over 1970-2019 | Time pattern in:* Condom use per sex act (***SexualBehaviour***,Rows 52-55)
* Coverage of syphilis screening
* Coverage of clinical treatment (***Screen+Treat***, Rows 35-43)
* Coverage of Contact tracing (***ContactTracing***, Rows 6-12)
* Number of partners per year (***SexualBehaviour***,Rows 5-13)
 |
| The difference in prevalence between high-risk groups (FSW, HRM and MSM) and medium-risk and esp. low-risk groups is too small or too large | Group-specific sexual behavioural and intervention coverages (see rows above). |
| Incidence too high (or too low) for a given prevalence. Unlike cohort studies, routine surveillance data don’t directly provide incidence. Case reports provide some indication of incident cases, but invariably represent only a part of actual cases, due to <100% symptoms, <100% treatment coverage and <100% reporting completeness | * Decrease (or increase) coverage of clinical treatment (***Screen+Treat***, Rows 35-43)
* Decrease (or increase) coverage of contact tracing (***ContactTracing***, Rows 6-12)
* Prolong (or shorten) duration of the latent stage (***Parameters***, Cell C22)
 |
| The ratio in prevalence, between RPR+ and ‘TPHA+ but RPR-‘ is too high, that is, there are too many people in Compartment 6 (recovered after treatment) | Compartment 6 reflects historic cumulative exposure to syphilis. Its prevalence is relatively high (for a given prevalence of active, RPR+ syphilis, i.e. Compartments 2-5) if syphilis declined much in recent decades. The RPR/TPHA ratio can therefore be manipulated by changing the historic time patterns in intervention coverage and/or in sexual risk behaviours |
| For given behavioural data (from surveys & IBBS) and program coverage data, modelled syphilis prevalence is too low (or too high( compared to data | First, consider the representativeness of the prevalence data. Did the prevalence surveys you are trying to calibrate perhaps under-sample (or over-sample) higher-risk individuals, within the population group it was meant to represent?In the behavioural data used to inform the model’s behaviour parameters, were risk behaviours perhaps under-reported and/or condom usage over-reported? If you think the contrast is not a matter of bias or under-reporting in the data, then consider changing the sexual mixing patterns. * The 3 mixing parameters in sheet ***Parameters*** (Rows 39-45) have complex effects (see Technical Methods report, sensitivity analyses).

More basically:* Higher proportions married in any or all groups will increase epidemic spread (***Sexual Behaviour***, Rows 26-33).
* Duration of Sex Work (***SexualBehaviour***, Row 48): A shorter duration will reduce prevalence in FSW, and increase transmission from the resulting larger (longitudinal/transitional) pool of FSW and ex-FSW, to High-risk male clients and via them, into the overall general population including low-risk women.
 |

Other parameters that users may change, but whose effects are not always straightforward, are:

* **Initial prevalence** (***Parameters***, Rows 48-54): If prevalence at 1970 for the respective groups is not set in line with their subsequent risk/exposure, defined by the behavioural parameters, the model may produce large fluctuations in incidence and prevalence over initial decades – which are not always realistic.
* **Proportion married among MSM** (***SexualBehaviour***, Row 33): Higher marriage rates among MSM will slightly increase prevalence among MSM, but especially among Lower-risk women. Typically, and in the example/default calibration, with low MSM marriage rates, MSM are a quite isolated group.