



2018 SKILLS BUILDING PROGRAM

BIG DATA, ARTIFICIAL INTELLIGENCE AND DECISION SCIENCE IN HEALTH AND NUTRITION

Collating data and populating the Optima TB databook

In partnership with





- Populations can be further broken down into smaller groups to better reflect the epidemic.
- Important to consider the availability of data for subpopulations before including them in the analysis

Population groups, an example

Children aged 0-4 years

Children aged 5-14 years

Adults aged 15-64 years

Adults aged 65 years and older

Prisoners

Coinfected and comorbidities

People living with HIV (PLHIV)

Diabetics

Exercise: Creating a new project



From the Projects page select Create new project

Years for data entry should reflect

- The first year from which you want to model the epidemic (default 2000)
- The last year for which you may wish to enter data (we recommend allowing for future years for easy updates)

For this training, it is recommended to pick just two populations such as “General population” and “PLHIV” or a different most significant key population.

Optima TB databook indicators



Data are entered into the databook, an Microsoft Excel spreadsheet, which is then uploaded into the Optima TB model.

Enter data in the following sheets:

1. Population definitions
2. Demographics
3. Notifications
4. Treatment outcomes
5. Latent treatment
6. Initialization estimates
7. New infections proportions
8. Optional data

Additional sheets (shaded in grey) have default values and usually do not need to be modified:

- Infection Susceptibility
- Untreated TB Progression Rates
- Interactions
- Transfers

Entering data in the Optima TB databook



**Blue cells =
input data
required**

**Red cells =
ignored**

Number of vaccinations administered		Units	Constant		2000	2001	2002
0-4	Number		100000	OR	90000	95000	99000
5-14	Number			OR			
15-64	Number			OR			
65+	Number			OR			
Prisoners	Number			OR			
LTBI treatment initiations total		Units	Constant		2000	2001	2002
0-4	Number		0	OR			
5-14	Number		0	OR			
15-64	Number		0	OR			
65+	Number		0	OR			
Prisoners	Number		0	OR			

**Non-shaded
cells= structural
(do not edit)**

**Grey cells =
default values**

“Constant” value vs annual data



- Data can be entered either with a single value in the “constant” column

OR, alternatively,

- One or more values in the annual data columns

Every row needs to have *at least* one value either in the constant column or in the annual data columns

A constant value is the same as entering that value for all individual years, and should generally be used if you need to make an assumption with limited data.

Considerations when entering data



- Data cannot be entered into the white cells
 - Unless a tag '#ignore' is entered into the first column for a row in which additional data is entered.
- Input data for the model versus calibration data
- Optima will **automatically interpolate missing data**
 - First entered data point is used for all previous years
 - Data values are interpolated linearly between every two entered data points
 - Last entered data point is used for all years thereafter
- “Constants” are ignored if any annual data points are entered for that parameter



- **Proportions:** these values should be interpreted as a proportion of outcomes that are grouped together and should sum to 1
 - such as the proportions of people who have different treatment outcomes.
- **Probability:** this refers to an annual probability of an outcome occurring at least once during that year,
 - such as the annual probability that latent TB progresses to active TB.
- **Number:** input values will be used directly for annual totals
 - such as the number of people initiating treatment in a given year.
- **N.A.:** relative multipliers that don't have any units
 - such as the relative risk of being infected with TB for someone with vaccination compared to someone without vaccination

2. Demographics



- Demographics data include:
 - Population size
 - avoid double-counting across population groups, verify total population size
 - Number of births
 - Non-TB deaths
 - Data on migration
- Data input units are specified for each, and may include options in the dropdown menu
- Exclude TB-related deaths from non-TB death rate



- Death rates may be entered as a probability instead of a number
- Non-TB death rate = crude death rate – estimated TB death rate

A “Non-TB death rate” of 0.04 would mean that each person in that population has a 4% chance of death due to non-TB related causes each year.

3. Notifications



- Clinical numbers of TB diagnoses and new treatment initiations
 - Should include best estimates of actual diagnoses and treatments if reporting is not comprehensive
- Notification data must be broken down by smear status, drug resistance and population
 - Where data is not available in this format, some judgements will have to be made
- Treatment initiation data must be broken down by drug resistance strain and population
 - If outcomes differ by smear status, then this can be calibrated later via the website



4. Treatment outcomes

- Should be entered using cohort data if available (most accurate to least):
 - Cohort data for people **concluding** treatment in the specified year
 - Cohort data for people **initiating** treatment in the specified year
 - Annual reported number of each outcome, adjusted so that the proportions add to 1
 - Best estimates using other sources
 - Treatment modality reported efficacy
 - Local expertise
 - Regional values

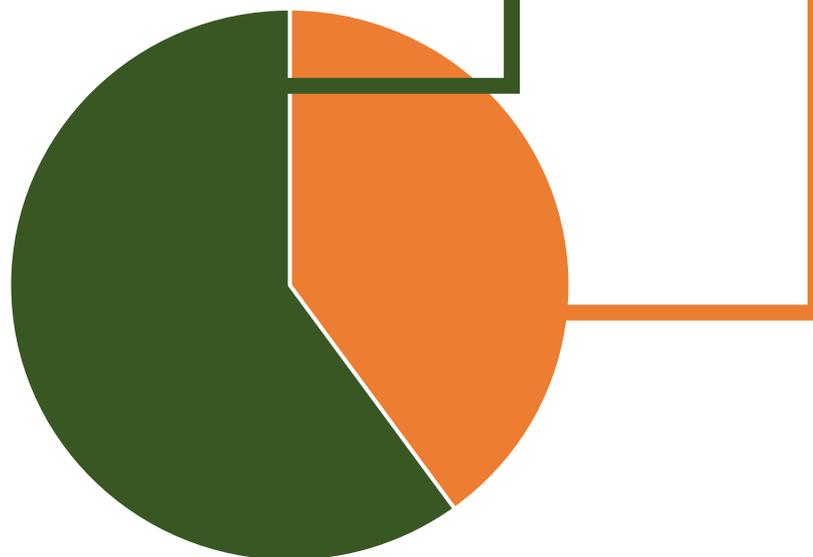
Proportions example: Population to Smear status



New active infections: proportion of population that are SP		Units	Constant	2000	2001
General population		Proportion	OR	0.6	0.55
Prisoners		Proportion	OR	0.7	0.65
New active infections: proportion of population that are SN		Units	Constant	2000	2001
General population		Proportion	OR	0.4	0.45
Prisoners		Proportion	OR	0.3	0.35

Proportion of the general population that are smear positive/smear negative

$$\text{Total} = 0.6 + 0.4 = 1$$



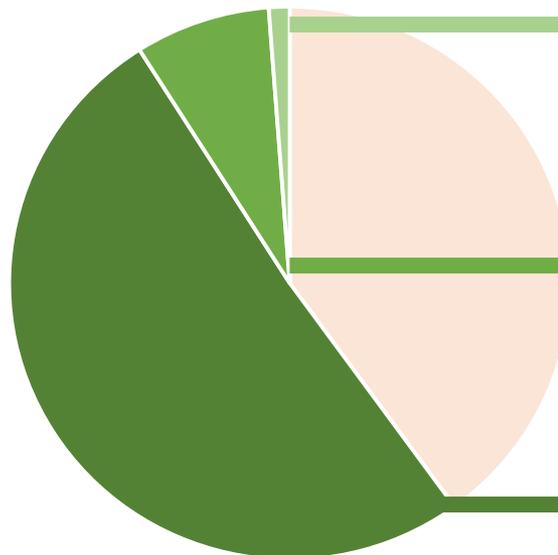
Proportions example: Smear status to drug resistance



New SP infections: proportion of population that are SP-DS		Units	Constant	OR	2000
General population		Proportion		OR	0.85
Prisoners		Proportion		OR	0.8
New SP infections: proportion of population that are SP-MDR		Units	Constant	OR	2000
General population		Proportion		OR	0.13
Prisoners		Proportion		OR	0.18
New SP infections: proportion of population that are SP-XDR		Units	Constant	OR	2000
General population		Proportion		OR	0.02
Prisoners		Proportion		OR	0.02

Proportion of the SP general population that are DS/MDR/XDR

$$\text{Total} = 0.85 + 0.13 + 0.02 = 1$$



5. Treatment of latent TB infections (LTBI)



- The number of annual BCG vaccinations for each year should be input here
- Default values of zero for LTBI treatment can be updated if LTBI treatment program exists

6. Initialization estimates



- Initialization (start point) estimates are required to initiate model (e.g. in year 2000)
- At least one value is mandatory for the latent and active prevalence
 - other estimates will improve the initialization
- Population sizes is duplicated here, to enable initialization
 - Can be adjusted during calibration, without changing actual data

7. New infections proportions

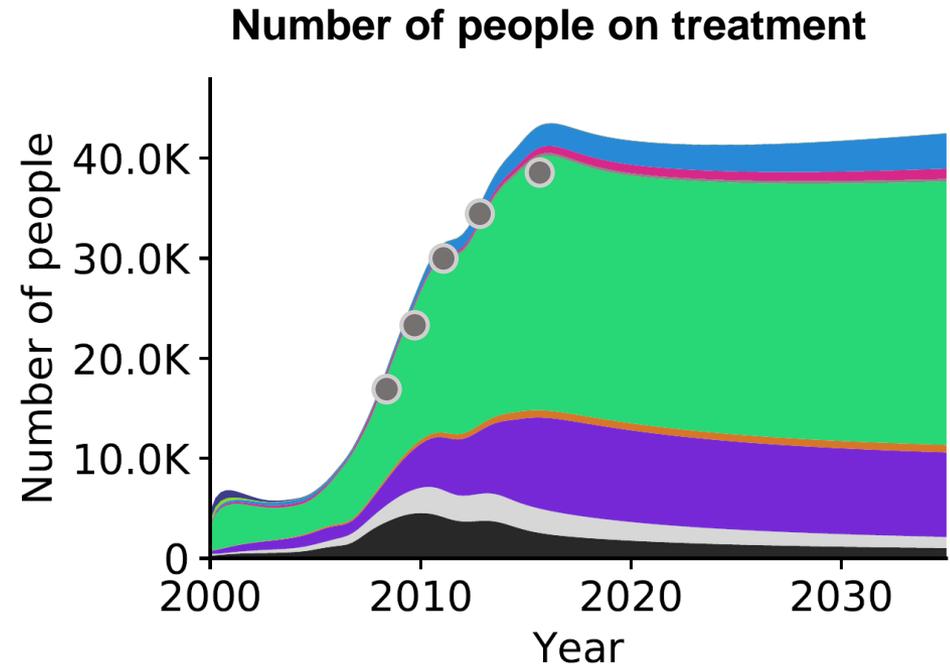


- Proportional breakdowns by smear status and drug resistance input here
 - calculated using notification data
- Example: In 2017 for population ages 5-14, there were **1500 Smear positive (SP)** notifications and **750 smear negative (SN)** notifications.
 - SP proportion of new active infections: $1500/(1500+750)$
 - **SP proportion = 0.67**
 - SN proportion of new active infections: $750/(1500+750)$
 - **SN proportion = 0.33**
- In contexts where there is incomplete notification data, estimates here can be used to disaggregate notification data
 - Some smoothing may be necessary to ensure data is consistent and reasonable.

8. Optional data



- Optional, are **not model inputs**
 - Used to plot known values against model outputs
- Values here are point estimates e.g. the number on treatment as of January 1 each year, rather than the total over the entire year





- Documenting data sources is important. Comments should be added to cells, outlining:
 - **Source:** a reference for the data
 - **Notes:** highlighting the quality of the data (such as sample size or confidence bounds), and any assumptions and/or calculations.
- Example
 - **Source:** Paper, Author, Year: value A
Report, Author, Year: value B
 - **Notes:** Paper was a very small study in one town of the country, used Value B from Report as this was a national study on a large cohort.



QUESTIONS?

Considerations when entering data



- Data cannot be entered into the white cells
 - Unless a tag '#ignore' is entered into the first column for a row in which additional data is entered.
- Input data for the model versus calibration data
- Optima will **automatically interpolate missing data**
 - First entered data point is used for all previous years
 - Data values are interpolated linearly between every two entered data points
 - Last entered data point is used for all years thereafter
- “Constants” are ignored if any annual data points are entered for that parameter

To review: the Optima TB model schema

