**READ ME: MISCAN\_SIMULATION**

This document provides an overview of the name, contents, function, location of each file. All files necessary to run the simulation run for my master thesis are denoted by “THESIS\_”, all other files serve as prefaces to these files.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Contents** | **Function** | **Origin** |
| Original files/  Cohorts\_\*.csv | CSV file with name, birth year, start year (year of first participation), number of (estimated) invitees, and status | Original file | [Hilliene Vandermeer – Health Counsel Analyses – Population effect](https://gitlab.com/Hilliene/health-council-analyses/-/tree/main/3.2%20Population%20effect) |
| Original files/  Cohorts\_\*.xlsx | Excel file with name, birth year, start year, number of (estimated) invitees, and status for cohorts 1 to 50 (our simulation period) | Preface to THESIS\_cohort  \_size\_\*.py [includes both table version, comma delimited version, and calculations] | Yoëlle Kilsdonk |
| THESIS  \_cohort\_  size\_\*.py | Python script containing precalculated fraction per cohort | Create cohort sizes. The number of invitees is used to determine the fraction of people that should be simulated per cohort in the final run (later on). | Yoëlle Kilsdonk |
|  | | | |
| Original files/  Strategies.csv | CVS file with strategies per cohort with information on age at screening, type of test, chances of participating; chances of participating given they’ve participated before (defaults to 1 for first invitation), chances of participating given they haven’t participated before (defaults to 1 for first invitation), chances of participating in diagnostic colonoscopy (surveillance) | Original file | [Hilliene Vandermeer – Health Counsel Analyses – Population effect](https://gitlab.com/Hilliene/health-council-analyses/-/tree/main/3.2%20Population%20effect) |
| Strategies.xls | Excel file with strategies per cohort with information on age at screening, type of test, chances of participating; chances of participating given they’ve participated before, chances of participating given they haven’t participated before, chances of participating in diagnostic colonoscopy (surveillance) for cohorts that started and ended between 2014-2020 (our simulation period) | Preface to THESIS  \_strategies.csv [includes both table version, comma delimited version] | Yoëlle Kilsdonk |
| THESIS\_  strategies.csv | CSV file containing the same as above | Input for THESIS\_cohort  \_strategy.py | Yoëlle Kilsdonk |
| THESIS  \_cohort\_  strategy.py | Python script containing methods to create and fill screening strategy based on cohort information | Input for strategy in CRC screening process in THESIS\_MISCAN  \_simulation.py | Yoëlle Kilsdonk[[1]](#footnote-1) |
|  | | | |
| THESIS  \_cohort\_  Birthyear.py | Python script containing year of individuals in each cohort. | Input to create birth table in birth process in THESIS\_MISCAN  \_simulation.py | Yoëlle Kilsdonk (from Cohorts\_\*.csv) |
| THESIS  \_CRC  \_data.py | Python script containing data for the CRC process calibrated on NL observations. | Input for CRC process in THESIS\_MISCAN  \_simulation.py | [Hilliene Vandermeer – Health Counsel Analyses – Population effect](https://gitlab.com/Hilliene/health-council-analyses/-/tree/main/3.2%20Population%20effect) |
| THESIS  \_CRC  \_screening\_  data\_\*.py | Python script containing data for the CRC screening process calibrated on NL observations. | Input for tests in CRC screening process in THESIS\_MISCAN  \_simulation.py | Yoëlle Kilsdonk[[2]](#footnote-2) |
|  |  |  |  |
| THESIS  \_crc\_  screening2.py | Python script as alternative to original panmodel.processes  .crc\_screening file | CRC screening module | Yoëlle Kilsdonk[[3]](#footnote-3) |

To run the simulation, run the “THESIS\_MISCAN\_simulation.py” in Python. Ensure that all files described above are in the proper working directory. After running the population simulation, you are left with two files named “individual\_results\_2014\_2020\_female.csv” and “individual\_results\_2014\_2020\_male”. These files report information on the cohort number, an ID per cohort, the age at event time, and a tag. The variables *universe* and *element* contain no information in this case.

The tags report when an individual was (first) invited, and if they participated in the test, it also reports the current stage of screen diagnosed cancer. It is also reported when clinical cancer is diagnosed (cancer diagnosis as cause of symptoms). In case a test result is positive, additional tags such as “crc\_screening\_triage\_\*” are shown, these are not relevant to our analysis. Finally, the tags also show when someone dies (from other causes or from CRC).

To convert this output to a data set suitable for the Multiple Imputation by Chained Equations (MICE) data set, we wrote the “THESIS\_MISCAN\_simulation\_data\_set” file. This file produces two data sets: “MISCAN\_simulation\_run\_female” and “MISCAN\_simulation\_run\_male”, which contain an ID number, age at event time, conclusion of the FIT, current stage of cancer, and sex. These two data sets are then used in the input for the MICE algorithm.

For more information on running the MICE algorithm, we refer to the enclosed READ ME\_data\_and\_MICE file.

1. Edited from [Danica van den Berg – Master Thesis – MISCAN with Hbsim](https://gitlab.com/danicavdberg/master-thesis/-/tree/master/MISCAN_with_HBsim). Changed all keys (green values) in def create\_screening\_strategy, and changed participation\_surveillance key to participation\_diag for compatibility with THESIS\_strategies.csv file and changes in THESIS\_crc\_screening2 to account for problem with participation for diagnostic colonoscopy (see teams interaction with Hilliene). [↑](#footnote-ref-1)
2. Edited from [Hilliene Vandermeer – Health Counsel Analyses – Population effect](https://gitlab.com/Hilliene/health-council-analyses/-/tree/main/3.2%20Population%20effect). Remove adjusted tests and cutoff sensitivity. [↑](#footnote-ref-2)
3. Edited from [Public Health – Screening –Pan model 2.0 – panmodel – processes – crc\_screening.](https://gitlab.com/erasmusmc-public-health/screening/panmodel2/-/tree/master/panmodel/processes/crc_screening) Changes to log current state and to fix participation of diagnostic colonoscopy. All changes are denoted by \*\*\*EDITED\*\*\* in the code. [↑](#footnote-ref-3)