Estimates of COVID-19 attributable mortality in Aden governorate, Yemen: a geospatial and statistical analysis

Notes on data and R analysis code

6 October 2020

Francesco Checchi
Department of Infectious Disease Epidemiology
Faculty of Epidemiology and Population Health, Keppel St
London School of Hygiene and Tropical Medicine
Francesco.checchi@lshtm.ac.uk

Datasets

All datasets required to replicate the analysis are included in the file <aden_covid_data.xlsx>. The file includes a 'dictionary' tab, whose column 'use' determines which variables are read into R analysis (below). Only variables marked as 'yes' for this column are necessary to replicate the analysis.

R code scripts

The file <aden_covid_control_code.R> loads necessary R packages, sets parameters (these can be modified by the user), loads datasets, calls dependent codes (below) and outputs some tables and figures. Only this code needs to be run to replicate the analysis. The main rate-limiting parameter in terms of computation time is the parameter n_boot, set at 1000: this results in computation times of approximately 5-10 min on a standard laptop.

The dependent code files, called from the above code, are, in order of implementation:

- <aden_covid_bespoke_functions.R>: this code contains all the user-defined functions (not all are in fact used);

- <aden_covid_prepare_data.R>: this code prepares data for analysis;

- <aden_covid_impute_graves.R>: this code fits a model to impute graves from cemetery surface area, and outputs descriptive tables and figures of the data after imputation;

- <aden_covid_case_based.R>: this code implements the case-based approach described in the paper, including spline smoothing and linear interpolation; it outputs tables and figures;

- <aden_covid_gamlss_model.R>: this code implements the approach relying on a generalised additive model for location, size and shape, as described in the paper; it also outputs tables and figures.

- The file <aden_covid_junkyard.R> contains pieces of unused code, and can be ignored.

Requirements to replicate the analysis

Datasets and R code files must be stored on the same folder, where output files will be saved. The directory for reading files is set automatically while the control code is run.

It is recommended to run the code from the open-source RStudio interface (https://rstudio.com/).