Projecting future temperature-related excess mortality from diarrhoeal diseases protocol

Authors: Paul L.C. Chua and Masahiro Hashizume
Affiliation: Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo
Email: paulchua@m.u-tokyo.ac.jp; hashizume@m.u-tokyo.ac.jp

A. Brief Description

We produced global spatio-temporal maps of temperature-related diarrhoeal diseases in ~50 km. The modelling approach is based on the study of Chua et al (2021)[1]. This section describes the modelling process, which comprises three steps (Figure 1):

1. The first step is deriving pathogen-specific risk functions by combining the reported relationships between temperature and pathogen-specific diarrhoeal diseases from published studies in the literature (Chua et al, 2022)[2]. There were 9 pathogens with available literature: non-typhoidal Salmonella, Shigella, Campylobacter, Vibrio cholerae, Escherichia coli, typhoid, rotavirus, norovirus, and Cryptosporidium.
2. The second step is projecting the diarrhoeal disease mortality rate not related to climate change, which is referred to as the "baseline mortality rate." This was based on socio-demographic index, time, and risk factor scalar.
3. The third step estimates the excess temperature-related diarrhoeal diseases based on the projected annual temperature anomaly, risk function by pathogen, projected baseline mortality rate, and projected population.

[1] Chua et al. Global projections of temperature-attributable mortality due to enteric infections: a modelling study. The Lancet Planetary Health. 2021;5(7):e436-45. doi:10.1016/S2542-5196(21)00152-2.
[2] Chua et al. Associations between ambient temperature and enteric infections by pathogen: a systematic review and meta-analysis. The Lancet Planetary Health. 2022;6(3):e202-18. doi:10.1016/S2542-5196(22)00003-1.

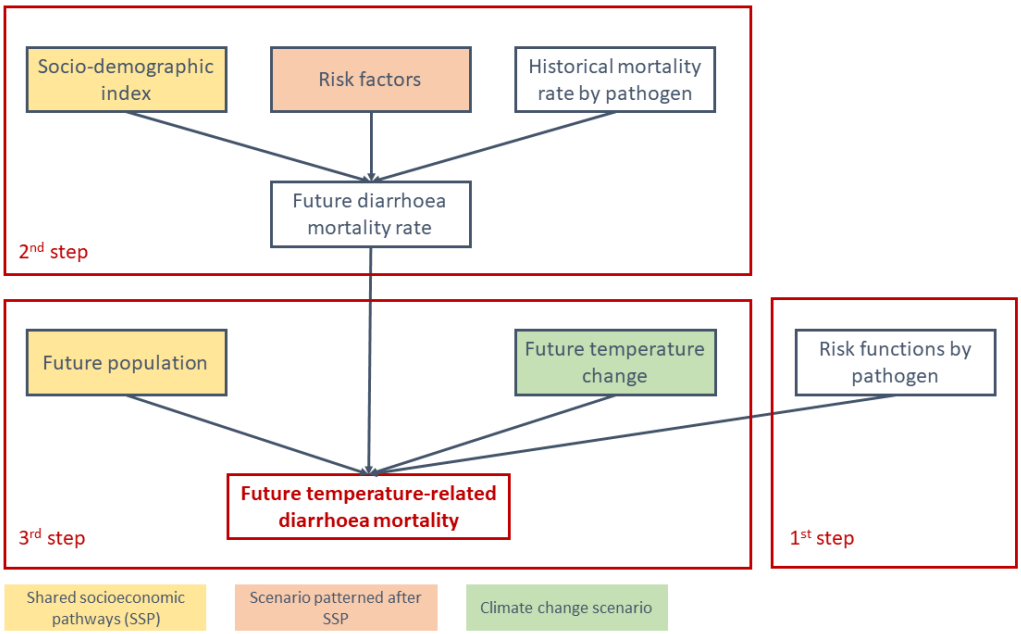

Figure 1. Modelling approach for diarrhoea projections

The following paragraphs expound the (1) dataset used, (2) processing required in cleaning and preparing the dataset, (3) statistical modelling, (4) current data output, and (5) future work including production of adaptation scenarios and meeting challenges.

B. Dataset

The dataset listed below were selected or narrowed down based on the conceptual framework in Figure 1. All data were freely collected online from existing databases.

Country shapefiles

We downloaded country shapefiles from Database of Global Administrative Areas (GADM) through their website: https://gadm.org/.

Projected Temperatures

Projected temperatures are required in the modelling because they are the basis of the future diarrhoea cases as defined by temperature-diarrhoea risk functions. The daily historical and future temperatures were downloaded from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP): https://data.isimip.org/. The ISIMIP climate data includes outputs from several general circulation models (GCM) which were bias-corrected or updated in the most recent year. The Chua et al (2021)[1] study used the old version of projected temperatures from the Coupled Model Intercomparison Project (CMIP) which was CMIP5 with the baseline year of 2005 and projected years

in 2006—2100 from five GCMs: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, and MIROC5. Currently, there are updated projected temperatures from CMIP6 with baseline year of 2014 and projected years from 2015-2100. The projected temperatures are based on climate change scenarios based on the rise of global temperatures from greenhouse gas emissions that was the combination of Shared Socioeconomic Pathways (SSP) and Representative Concentration Pathways (RCP).[3] The global temperatures were available by 50 km grids (0.5°×0.5°) in the format of a 3-dimension array (X=longitude, Y=latitude, Z=time) saved in NetCDF (network Common Data Form).

Projected population

Historical and future population are required in the modelling to derive the baseline number of deaths related to diarrhoeal diseases. Several studies performed population projections by SSPs by country (data available from International Institute for Applied Systems Analysis) and downscaled the outputs into small grids. For Chua et al (2021)[1] study, the projected population was collected from https://data.isimip.org/ derived from Samir and Lutz (2017)[4]. This data was in the format of a 3-dimension array (X=longitude, Y=latitude, Z=time) saved in NetCDF.

Temperature-diarrhoea risk functions by pathogen

Risk functions depicting the relationship between temperature and diarrhoea is required in the modelling. We used pooled relative risks representing the percentage change of diarrhoea incidence per 1°C temperature rise according to pathogen derived from our recent systematic review and meta-analysis from published research articles from 2000-2020.[5] Table 1 enumerates the percentage changes per 1°C temperature rise per pathogen.

Table 1. Percentage change in diarrhoea incidence per 1°C temperature rise by pathogen

Diarrhoeal disease pathogen	**Percentage change in incidence (95% confidence intervals)**
Non-typhoidal *Salmonella*	5.1 (3.6 to 6.7)
Shigella	7.0 (4.4 to 9.6)
Campylobacter	2.3 (0.7 to 4.0)
Vibrio cholerae	5.4 (4.2 to 6.6)
Escherichia coli	4.3 (1.2 to 7.4)
Typhoid (*Salmonella typhi*)	15.1 (7.1 to 23.6)
Rotavirus	-4.4 (-10.5 to 2.1)
Norovirus	-10.5 (-19.3 to -0.9)

[3] O'Neill BC et al. The scenario model intercomparison project (ScenarioMIP) for CMIP6. Geoscientific Model Development. 2016 Sep 28;9(9):3461-82. doi: 10.5194/gmd-9-3461-2016.
[4] Samir KC, Lutz W. The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. Global Environmental Change. 2017;42:181-92. doi: 10.1016/j.gloenvcha.2014.06.004.
[5] Chua et al. Associations between ambient temperature and enteric infections by pathogen: a systematic review and meta-analysis. The Lancet Planetary Health. 2022;6(3):e202-18. doi: 10.1016/S2542-5196(22)00003-1.

Diarrhoeal disease pathogen	Percentage change in incidence (95% confidence intervals)
Cryptosporidium	17.0 (8.0 to 26.0)

Historical mortality rates of diarrhoeal diseases

Mortality rates of diarrhoeal diseases are required to derive the baseline diarrhoea mortality (together with the population). Country-level mortality rate by pathogen modelled from the country reports and other available information are produced by Institute of Health Metrics and Evaluation-Global Burden of Disease (IHME-GBD) and are available at https://vizhub.healthdata.org/gbd-results/ starting from 1990. There are 15 pathogens available but only 10 pathogens with temperature-diarrhoea risk functions were selected namely: non-typhoidal *Salmonella*, *Shigella*, *Campylobacter*, cholera, enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), typhoid, rotavirus, norovirus, and *Cryptosporidium*. The EPEC and ETEC used the same risk functions. For Chua et al (2021)[1] study, the mortality rates in 1990 to 2017 by age groups were downloaded. Five age groups included were: less than 5 years, 5-14 years, 15-49 years, 50-69 years, and 70+ years. Currently, the mortality rates are available up to 2019 in 205 countries. This data was in a CSV file broken down by columns of country, age group, year, pathogen, and mortality rate. The 10 pathogens are the majority of the diarrhoeal diseases as shown in Table and Figure 2.

Table 2. Mortality rates (deaths per 100,000) per aetiology/pathogen in 1990 and 2019 from IHME

Label	Aetiology / pathogen	1990		2019	
		Rate	%	Rate	%
1	Non-typhoidal Salmonella	2.55	5.02	0.8	4.26
2	Shigella	5.62	11.05	1.92	10.25
3	Campylobacter	4.72	9.28	1.8	9.62
4	Cholera	1.55	3.06	1.52	8.11
5	Enteropathogenic E coli	0.91	1.79	0.27	1.43
6	Enterotoxigenic E coli	1.32	2.61	0.51	2.75
7	Typhoid fever	3.41	6.72	1.42	7.61
8	Rotavirus	12.32	24.25	3.04	16.28
9	Norovirus	4.54	8.94	1.76	9.39
10	Cryptosporidium	4.87	9.58	1.72	9.23
11	Adenovirus	4.48	8.81	1.38	7.41
12	Aeromonas	1.11	2.18	0.36	1.94
13	Clostridium difficile	0.16	0.31	0.42	2.22
14	Entamoeba	1.61	3.17	0.43	2.31
15	Paratyphoid fever	0.63	1.24	0.3	1.61

16	Invasive Non-typhoidal Salmonella	0.94	1.86	1.02	5.47
17	Other intestinal infectious diseases	0.07	0.14	0.02	0.1

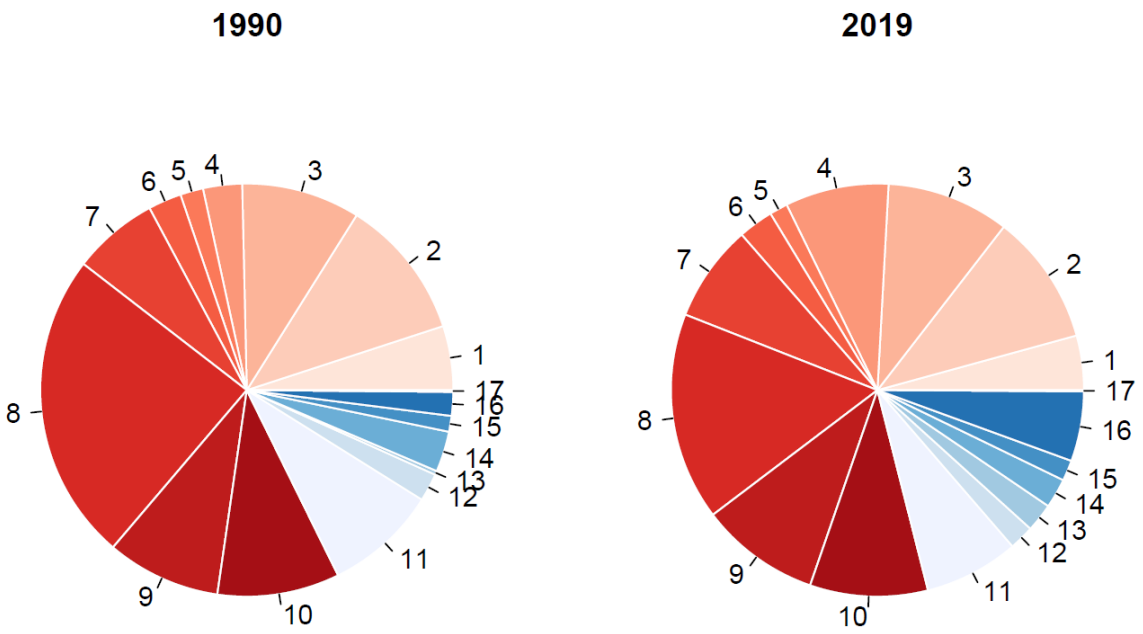

Figure 2. Proportion of pathogens for diarrhoeal diseases.

Labels are in Table 2. Red colours are the 10 pathogens included in the dataset and blue colours are aetiologies/pathogens not included.

Sociodemographic index (SDI) related data

To model the future diarrhoea mortality rates, SDI is the major variable required. The country-level SDI was the combination of projected gross domestic product per capita (GDP), mean years of schooling per capita in 25+ years old (MYS), and total fertility rate in 15-25 years old (TFR) downloaded from International Institute for Applied Systems Analysis the Wittgenstein Centre for Demography and Global Human Capital database by SSPs. Only data for 179 countries were available and used for Chua et al (2021) study. The GDP, MYS, and TFR data were in a CSV file format broken down by columns by country and year.

Risk factors related to diarrhoeal diseases

To model the future diarrhoea mortality rates, the risk factors of diarrhoeal diseases were considered. Overall population attributable fractions (PAFs) of seven risk factors were produced namely: (1) unsafe sanitation, (2) no access to handwashing facility, (3) child underweight, (4) child wasting, (5) child stunting, (6) vitamin A deficiency, and (7) zinc deficiency. Historical summary exposure values for each risk factors by age group were downloaded from https://vizhub.healthdata.org/gbd-results/. Each

data was expressed in summary exposure values (SEV) in a CSV format with columns of country and age groups expressed in 0-1 value (continuous data).

C. Data Preparation and Processing

This section enumerates the required processing of the raw data to prepare for the modelling.

Country-assignment of grids

We used the mesh or grids from the longitudes and latitudes of the temperature data. Each grid was allotted to a country if the grid centroids are within the country-level shapefiles.

Future temperature anomaly

Reference annual temperatures were calculated by taking the mean of daily temperatures from 1976-2005. At least 30 years are required as reference and the limit of year 2005 was selected because of the baseline year of CMIP5. We took each ~50 km grids reference temperatures. The annual temperature anomalies were the difference between reference temperatures and annual future temperatures from 2006-2099 by ~50 km grids.

Building SDI

The GDP and MYS were in 5-year intervals, while the TFR was expressed as 5-year averages. To get annual data, GDP and MYS were linearly interpolated. The TFR annual data, on the other hand, were derived by simply replicating the 5-year average values into their respective years. To create historical GDP before 2000, we used the historical GDP patterns from World Bank expressed in constant 2005 International Dollars from James et al. (2012)[6] and linearly interpolated.

Each value of GDP, MYS, and TFR per year was converted into a scale of 0 to 1 by scaling with best and worst values:

1. Worst and best GDP are Int$ 250 and Int$ 60000, respectively.
2. Worst and best MYS are 0 and 17, respectively.
3. Worst and best TFR are 3 and 0, respectively.

We cut the 2.5% and 97.5% GDP values since there were values lower than the worst value. Geometric mean of three indicators were computed to derive SDI. We were able to calculate SDI for SSP 1, 2 and 3.

[6] James et al. Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. Population Health Metrics. 2012. doi:10.1186/1478-7954-10-12.

Scalar of risk factors

Annualised rate of change (AROC) δ each country l, age a, and risk r from historical SEVs for each risk factor were computed using the following calculations:

$d_{l,a,r,t} = logit\left(SEV_{l,a,r,t}\right) - logit\left(SEV_{l,a,r,t-1}\right)$,

$\delta_{l,a,r} = mean\left(d_{l,a,r,t}, w_{r,t}\right)$,

$w_{r,t} = (t - 1990 + 1)^{\omega_r}$

Weights w_t were determined by a recency-weighting parameter from ω from Foreman et al. (2018). We used 15th,50th, and 85th percentiles of AROC and projected from 2006-2099 to represent the low health investments (LHI), baseline health investments (BHI), and additional health investments (AHI) which depict worse, middle-of-the-road, and best future scenarios. Each risk factor SEV were converted into population attributable fractions (PAFs) and all PAFs were combined into a single overall PAF representing all risk factors.

D. Methods

We explain here the future scenarios selected to simplify presentation of the projected values, as well as the explanation of the modelling done for the second step and third steps shown in Figure 1.

Future scenarios

There are numerous ways and options to select future scenarios but for simplicity to present, the Chua et al (2021)[1] study chose three scenarios: (1) optimistic, (2) intermediate, and (3) pessimistic. The optimistic scenario utilizes SSP 1, which reflects high socio-demographic development and low global warming (RCP 2.6). To align the risk factors, best reductions to the seven risk factors were selected which is the 85th percentile of AROC. The intermediate scenario applies middle-of-the-road scenarios like SSP 2 (which generally follows historical patterns for socio-demographic development), RCP 4.5, and 50th percentile of AROC for risk factors. The pessimistic scenario follows the worst-case that were the combination of SSP 3, RCP 6.0. and 15th percentile of AROC.

Future diarrhoea mortality rates

The baseline diarrhoea mortality rate (without climate change) was derived from a linear mixed model with the SDI and trend (or year) as fixed effects, and the country, age, and pathogen as random effects[7]:

[7] Vollset et al. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. The Lancet. 2020;396(10258):1285-306. doi: 10.1016/S0140-6736(20)30677-2.

$$ln\left[E\left(D_{r,l,a,t}\right)\right] = \alpha_{r,l,a} + \beta_0 SDI_{<0.8,l,t} + \beta_1 SDI_{\geq 0.8,l,t} + \theta_a t + ln(S_{l,a,t}),$$

where the logarithm of projected annual enteric infection mortality rate D is expressed as an additive relationship between the intercept α, the SDI (as a piecewise-linear function with a cutoff value of 0.8), the year t, and a scalar of risk factors S (as an offset). The symbol r denotes the pathogen, l the country, and a the age group, and β_0, β_1, and θ denote the parameters of estimation. The years included to train the model are from 1990 to 2005 (because the temperature projections are based on 2005 as baseline). Projections were produced by scenario: optimistic, intermediate, and pessimistic. The modelling is done separately by pathogen and age-group. To validate the model, mean and median root mean squared error from actual and forecasted mortality rates (using intermediate scenario because they depict the continuation of patterns) for 2015–2017. The outputs are produced in CSV files of diarrhoea rates by country and year. Figure 3 shows the optimistic and pessimistic projection outputs of the mortality rate by pathogen.

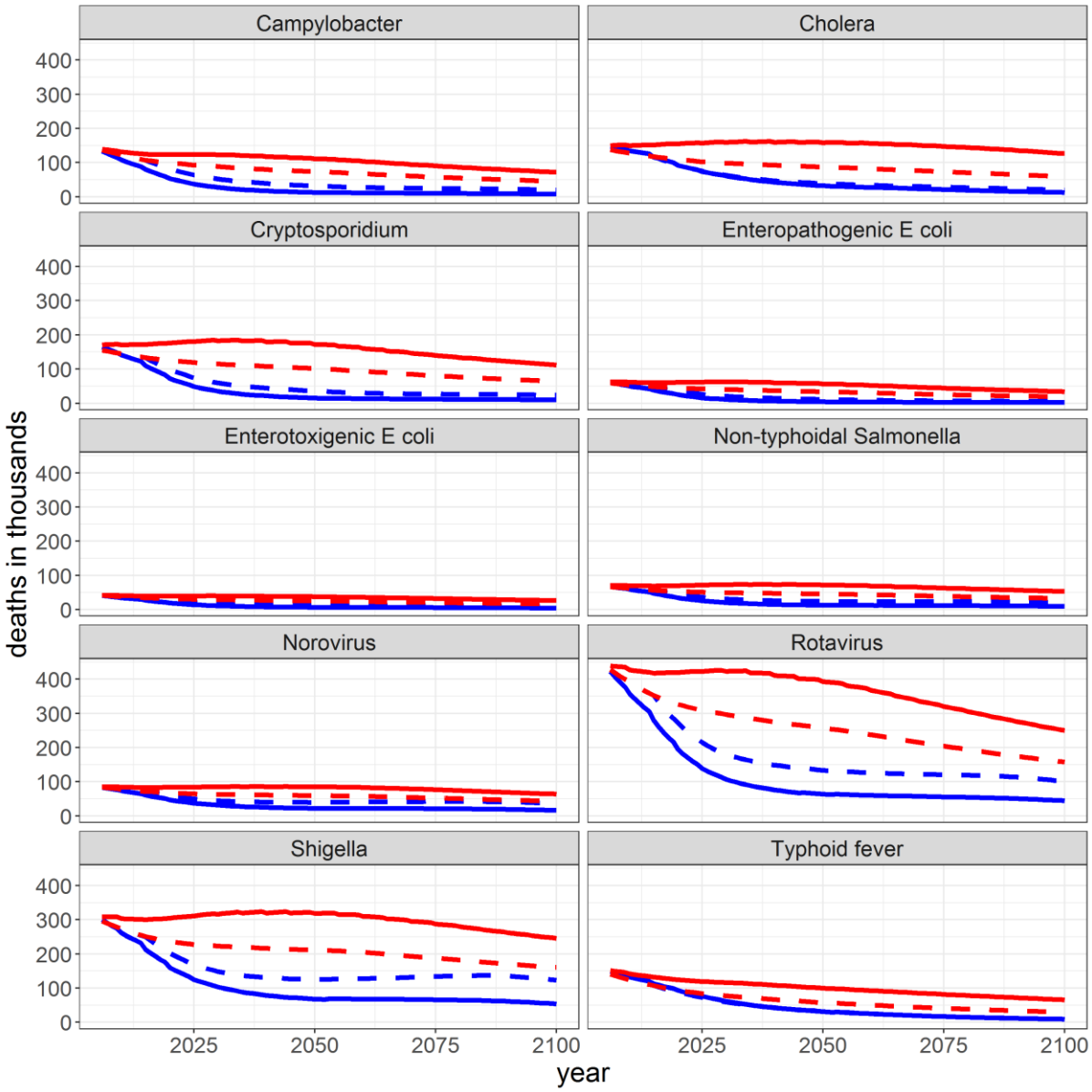

Figure 3. Projected mortality rates by optimistic (blue) and pessimistic (red) scenarios without climate change.

Solid lines refer to outputs from IHME model (explained in this report) and dashed lines are outputs from Mathers-Loncar model.

Projection modelling

The annual number of temperature-related diarrhoea mortality from the baseline enteric infection mortality rates using the product of temperature anomalies and pathogen-specific temperature sensitivity as follows:

$$d_{r,l,t} = D_{r,l,a,t} \times P_{l,a,t} \times \frac{e^{\left[\beta_r \times \left(T_{projected} - T_{baseline}\right)\right]-1}}{e^{\left[\beta_r \times \left(T_{projected} - T_{baseline}\right)\right]}},$$

where d is the temperature-attributable enteric infection mortality, D is the baseline enteric infection mortality rates derived from the first stage, P is the population, β is the pathogen-specific mortality change per 1°C increase in temperature (Table 1), and T is the annual mean temperature.

We performed the calculation by grid. To calculate uncertainty, a bootstrapping approach to account for the uncertainty related to different general circulation models and β values can be utilized. The outputs are produced by 50 km grid cells and saved in a 3-dimension array (X=longitude, Y=latitude, Z=time) saved in NetCDF.

E. Building adaptation scenarios

The current outputs are produced in simplified future scenarios generalising the adaptation. To make the maps or visualisations more informative for the general audience like policy- and decision-makers, impact of individual adaptation strategies should be shown. Adaptation strategies that can be considered for diarrhoeal diseases in view of climate change impacts may include improvements in: (1) rotavirus vaccination coverage, (2) oral rehydration salts, (3) child nutrition, and (4) access to sanitation and hygiene facilities. These strategies will be modelled separately to produce projected values. Targets followed internationally or nationally could be produced as realistic options for adaptation but national or local level targets should also be taken into consideration.

The following are descriptions of the modelling for the possible adaptation strategies:

1. *Rotavirus vaccination*
   The coverage rates could be modelled after the SDI and coverage rates of important vaccination like DPT (diphtheria, pertussis, and tetanus) since the rotavirus vaccine would be included in those schedules. By using the coverage rates of DPT, countries without rotavirus vaccination could be projected with an assumption of introduction in the future. Then the projected rotavirus vaccination coverage rates can then be inputted in the diarrhoea mortality model as a replacement of SDI.
2. *Oral rehydration salts*
   Similar with rotavirus vaccination, ORS coverage can be modelled using SDI with the assumption that better socio-demographic conditions lead to better response in primary healthcare services including ORS. This becomes a good prevention of mortality against diarrhoeal diseases. This item can be inputted in all modelling of pathogens.
3. *Child nutrition*
   Projections in the reductions of child undernutrition that will be represented by wasting, underweight, and stunting can be based on targets laid out by World Health Organization in relation with Sustainable Development Goal 2 that is zero prevalence of malnutrition by 2030. Modifications may be considered on the decade the target could be achieved given the possibility of much later realization/achievements of these targets. These malnutrition proportions could be translated into population attributable fractions and modelled according

to the considered timeline of zero malnutrition proportion. Final outputs will be combined with the risk factor scalar.

4. *Access to sanitation and hygiene facilities*
   Household coverage of sanitation and hygiene facilities could be based on targets mentioned in SDG 6 that is access for all and zero open defecation by 2030. The 100% coverage of sanitation and hygiene could be translated to population attributable fractions and model the projected trends overtime. Outputs will be combined with the risk factor scalar.

F. Challenges

There are several challenges in producing outputs for adaptation to the future diarrhoeal diseases in the perspective of modelling:

1. The current approach in including the adaptation strategies is through the parameter risk factor scalar. The scalar are pooled population attributable fractions of assumed exposure to risk factors and do not show the direct impacts of improved coverage rates of adaptation strategies. Assumptions may need to be applied on the exposed population on risk factors given better adaptation strategies. Another way is to explore different baseline modelling approach to use the actual reported coverage of diarrhoea-related strategies. Thus, current modelling approach can be compared to another modelling of strategies.
2. The current ~50 km grids are too course and cannot be used to compute for smaller administrative level boundaries sub-nationally. There is a need to downscale to smaller grids to allow calculations of sub-national counts of temperature-related diarrhoea deaths. This is further discussed in the next section about downscaling approach of projected diarrhoea mortality.
3. The outputs are for mortality but do not show the morbidity like incidence or prevalence of diarrhoea cases. Modelling for the morbidity could be a good addition but may technically be difficult and may need to explore the most appropriate baseline modelling.
4. Apart from temperature, there are a number of climate variables that affect the transmission of diarrhoeal diseases. It should be explored the addition of rainfall and other climate-related events like tropical cyclones and flooding on top of the temperature modelling.
5. Updating the risk functions of temperature with current evidence should be considered in the future. More studies are being produced that are pathogen-specific. Moreover, multi-country outputs for pathogen-specific diarrhoeal diseases should be explored as more data are becoming available with better diagnostics.
6. There are more waterborne diseases that can be explored apart from diarrhoeal diseases in the future.
7. Involvement of actual policy- and decision-makers should be ensured to provide realistic feedback on the outputs. There may be multiple rounds of updating the information to suit the actual needs of the user and reflect the adaptation options they need to visualise.

grids.