

Supplementary Materials and Methods

1. Usage of the R Library Provided in SUMRAY

The package built in R provides a simple hands-on tool to compute cumulative excess risks (CERs) of radiation-induced cancer under a specified exposure scenario. The main function of the package, CER, can compute the risk in the following input data: (1) exposure information (age, dose in Gy, sex), (2) risk model, (3) reference data, and (4) other options specified in the argument as follows.

```
CER( agex, doseGy, sex, riskmodel=LSS_allsolid_mortality, wgt=c(1,0),
     baseline=allsolid_mortality_Japan2018,
     mortality=mortality_Japan2018,
     maxage=max(mortality$age),alpha=0.05, n_mcsamp=10000, seed=NULL )
```

1) Specification of population reference data sets

An exposure scenario is specified by *agex* (age at exposure), *doseGy* (dose in Gy), *sex* (1 for male or 2 for female), and *maxage* (the maximum age to follow-up). For each of *agex* and *doseGy*, a single value may be specified for a single exposure, while a vector is specified for a series of exposures.

2) Specification of population reference data sets

A risk model is specified as *riskmodel*, which must be a list object that contains two list objects named *err* and *ear*, each of which must contain a vector (named *para*) of the model parameter estimates, a matrix (named *var*) for the variance-covariance matrix of the parameter estimates, and a function (named *f*) to calculate the age-specific excess risk given a parameter set in the first argument followed by a *data.frame* object named *data* to specify the age and exposure scenario, with *age*, *dose*, *agex* (age at exposure), and *sex*. The user can specify any risk model following this data structure. The default object for *riskmodel* in the CER function is *LSS_allsolid_mortality*, which is a risk model derived from analysis (LSS report 14) for all-solid-cancer mortality in the Life Span Study cohort of Japanese atomic-bomb survivors [18]. As shown below, the functions *err* and *ear* under the risk model *LSS_allsolid_mortality* compute the excess relative and absolute risks multiplied with the latency adjustment function (4) with the value of 2.5 for the location parameter (which is the years since exposure at which the adjustment factor is 0.5). By modifying these functions, any other latency adjustment can be considered.

```
> LSS_allsolid_mortality
$err
$err$para
[1] 0.422650 -0.345890 -0.857428 0.344079

$err$var
      [,1]      [,2]      [,3]      [,4]
[1,] 0.00253235 0.00140692 0.00692967 -0.00028615
[2,] 0.00140692 0.00661873 -0.01588490 -0.00101646
[3,] 0.00692967 -0.01588490 0.17912000 0.00354537
[4,] -0.00028615 -0.00101646 0.00354537 0.00771445

$err$f
function (beta, data) {
  beta[1] * data$dose *
  exp(beta[2] * (data$age - 30)/10 + beta[3] * log(data$age/70)) *
  (1 + c(-1, 1)[data$sex] * beta[4]) /
  (1 + exp(-(data$age - data$agex - 2.5)))
}
```

```

$ear
$ear$para
[1] 0.00263965 -0.21346100 3.38439000 0.06769100

$ear$var
      [,1]      [,2]      [,3]      [,4]
[1,] 9.87492e-08 6.41078e-06 4.17608e-05 -1.25134e-05
[2,] 6.41078e-06 5.12184e-03 -1.22614e-02 -2.25000e-05
[3,] 4.17608e-05 -1.22614e-02 1.33872e-01 -5.84032e-03
[4,] -1.25134e-05 -2.25000e-05 -5.84032e-03 9.31031e-03

$ear$f
function (beta, data) {
  beta[1] * data$dose *
  exp(beta[2] * (data$age - 30)/10 + beta[3] * log(data$age/70)) *
  (1 + c(-1, 1)[data$sex] * beta[4]) /
  (1 + exp(-(data$age - data$age - 2.5)))
}

```

3) Specifications of population reference data sets

Population reference data are specified by the two arguments: baseline (the baseline rate of the target endpoint) and mortality (the all-cause mortality rate), both of which contain age- and sex-specific rates in the target population. Each of the reference data sets is specified in a `data.frame` object containing the `age`, `male`, and `female` columns. The `age` column can contain ages in integers starting from 1 to a number as large as `maxage`, or midpoints of intervals of equal periods, e.g., 2.5, 7.5, 12.5, ..., which are midpoints of the five year intervals (0–4, 5–9, 10–14, ...). The `male` and `female` columns should contain sex-specific rates corresponding to the age (category) in each row. The default data sets of `allsolid_mortality_Japan2018` and `mortality_Japan2018` (mortality rates from all-solid cancers and all causes in Japan in 2018) are provided by the package (shown below).

```

> mortality_Japan2018
  age    male    female
1   1 0.0001038 0.0000970
2   2 0.0003114 0.0002910
3   3 0.0005190 0.0004850
:
98  98 0.3129080 0.2359890
99  99 0.3397930 0.2697846
100 100 0.3666780 0.3035802

> allsolid_mortality_Japan2018
  age    male    female
1   1 0.0000020 0.0000032
2   2 0.0000060 0.0000096
3   3 0.0000100 0.0000160
:
98  98 0.0269670 0.0129870
99  99 0.0269670 0.0129870
100 100 0.0269670 0.0129870

```

4) Other options

The package provides some optional arguments to control the implementation of risk calculation. The `wgt` argument specifies weights for the excess relative risk (ERR) and excess absolute risk (EAR) transfers (e.g., `c(1,0)` [default], which indicates ERR

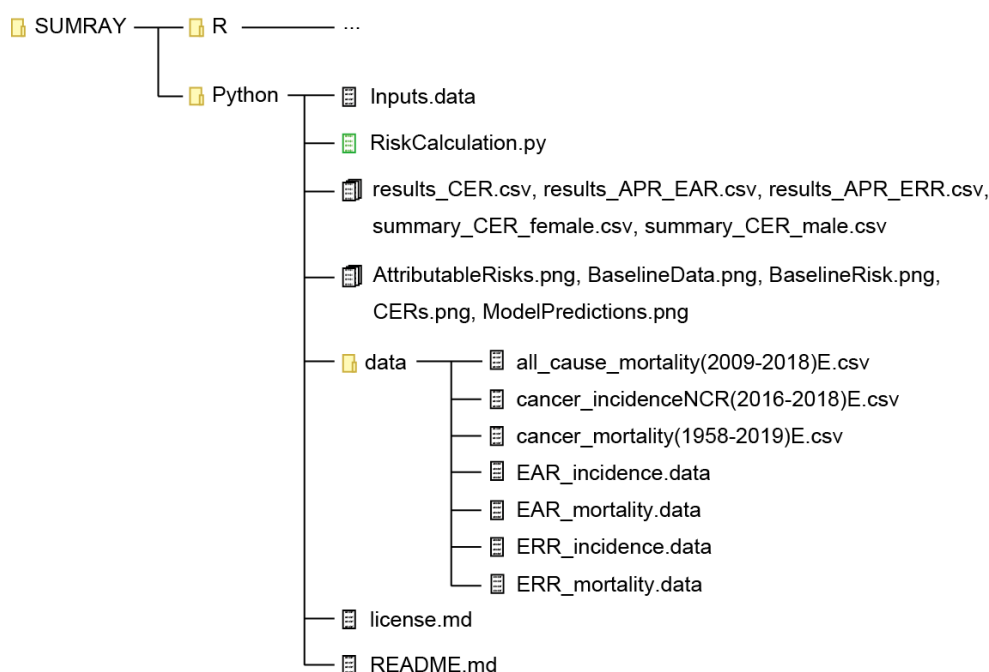
transfer, or $c(0.7, 0.3)$ to weight 70%:30% for ERR:EAR). The argument `alpha` specifies the significance level for the confidence interval, and the argument `n_mcsamp` specifies the number of Monte Carlo samples (with a default of 10,000). The `seed` argument can fix the random number seed, which may be useful for replicating exactly the same results from Monte Carlo approximations.

2. Usage of the Python Script Provided in SUMRAY

The Python script was written in Python 3, which is a high-level, interpreted, general-purpose programming language, and was tested on a virtual environment consisting of Python 3.8.5 and extra-modules of *os*, *sys*, *csv*, *pandas*, *numpy*, *matplotlib*, and *scipy*.

Supplementary Fig. S1 illustrates the directory-and-file structure of the Python-version SUMRAY code. The Python script named “RiskCalculation.py” requires a directory named “data” and an input file named “Inputs.data” in the same directory. The “data” directory contains the comma-separated values (.csv) files of the baseline risk of the Japanese population and the American Standard Code for Information Interchange (ASCII) files for the risk-model parameters. The details of the baseline risk and the risk-model parameters used in the default setting are explained in the main body of this article. If the users prepared their own baseline and risk-model files with different names from the default ones, they must amend the Python script to update the file-name definitions given at the first part of the script.

The file “Inputs.data” contains seven parameters and information regarding exposure events to configure a risk-assessment problem. In the file, lines starting with the character “#” are comments and are skipped for reading in the script. Supplementary Table S1 summarizes the seven input parameters. Note that the order of the parameters is fixed as listed in Supplementary Table S1. Below the parameter section, the information for each exposure event is described on a line-by-line basis. One line consists of two-column data; the first and second columns indicate the dose in Gy and the age at exposure, respectively. If the users want to define a chronic exposure, they add lines year-by-year with exposure doses from the minimum to maximum age in the exposure period. Supplementary Fig. S2 gives an example of the input data for the Python script. In this example, the baseline risk with class values of 5 years is interpolated linearly at each year, and the latent period from the year at exposure is set with the value of 7.5 years that represents the center of the sigmoid curve. Then the CER of all-solid-cancer incidence to the attained age of 90 years old for Japanese females who had a 0.1-Gy acute exposure at 10 years of age is calculated. After the cal-



Supplementary Fig. S1. Directory-and-file structure of the Python-version SUMRAY code.

Supplementary Table S1. List of the Input Parameters

Order of parameters	Explanation
1 st	Parameter to set the target risk. The values of "1" and "2" are acceptable, and they set the computation to assess the all-solid-cancer mortality and incidence rate risks, respectively.
2 nd	Parameter to set the latent period, which is represented by a Sigmoid curve. This parameter determines the center of the Sigmoid curve drawn from the age at exposure.
3 rd	Parameter to set the interpolation method for the baseline risk. If the users set the parameter to "1", "2", and "3," the script uses the default baseline risk with original class values of 5 years, interpolated values at each year with linear interpolation, and interpolated values at each year with Piecewise Cubic Hermite Interpolating Polynomial (PCHIP). If the parameter is set to "0," the script uses the user-provided baseline risk at each year.
4 th	Parameter to set the gender of an exposed subject. If the value of "1" is chosen, the subject is male. If "2" is chosen, the subject is female.
5 th	Parameter to set the maximum attained age for CER computation. The age range is from 0 to 100 years.
6 th	Parameter to set the risk-transfer weight. The details of the risk-transfer weight are explained in the main body of this paper.
7 th	Parameter to set the output mode. The value should be positive. When the parameter is set to "0" and the exposure event is one which indicates an acute exposure, the script generates the CSV files of calculated CERs for male and female by changing the age at exposure from 1 to 95 years.

CER, cumulative excess risk; CSV, comma-separated value.

```

Inputs.data
1 # -----↓
2 # INPUT data file for computing Cumulative Excess Risks (CERs).↓
3 # -----↓
4 # Japan Health Physics Society (JHPS), 2022.↓
5 # -----↓
6 ↓
7 # Select target risk: -----↓
8 # tgt0 (integer); 1 = solid-cancer mortality, 2 = solid-cancer incidence↓
9 2↓
10 ↓
11 # Set latency period: -----↓
12 # Latency represented by a Sigmoid curve with a parameter (lpy0) in the center of the Sigmoid function↓
13 # lpy0 (float); lpy0 (years) >= 0↓
14 7.5↓
15 ↓
16 # Select step of age for baseline data: -----↓
17 # stp0 (integer); ↓
18 # 0 = for user-provided baseline data at each one year, ↓
19 # 1 = original class values of 5 years, ↓
20 # 2 = values at each one year with linear interpolation, ↓
21 # 3 = values at each one year with Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)↓
22 2↓
23 ↓
24 # Select gender: -----↓
25 # sex0 (integer); 1 = Male, 2 = Female↓
26 2↓
27 ↓
28 # Set maximum attained age for calculation: -----↓
29 # maxage0 (integer); 0 < maxage0 (years) =< 100↓
30 90↓
31 ↓
32 # Set risk-transfer weight: -----↓
33 # 1 if choose EAR model↓
34 # wgt0 (float); 0.0 <= wgt0 <= 1.0; ERR (wgt0), EAR(1-wgt0)↓
35 0.5↓
36 ↓
37 # Select output option: -----↓
38 # mode0 (integer); 0 = outputting the CER data to csv file, 1 >= No output↓
39 1↓
40 ↓
41 # Exposure event: -----↓
42 # dose0 in Gy (float), agex0 in years (float)↓
43 0.100, 10↓
44 # -----↓
45 [EOF]

```

Supplementary Fig. S2. Example of "Inputs.data" for the Python-version SUMRAY code.

ulation, the script outputs the results of CER with its mean, median, 2.5th, and 97.5th values on a standard output. The results are also written in the file named “results_CER.csv” with the input echo. In addition, the calculation results of the attributable probability rates at each attained age based on the EAR and ERR models are output to the files named “results_APR_EAR.csv” and “results_APR_ERR.csv,” respectively. As the 6th parameter of the example is set to “1,” summary files of the male and female CERs whose names are “summary_CER_male.csv” and “summary_CER_female.csv,” respectively, are not generated in this calculation.

3. Acute Exposure

The following box shows how to compute the cumulated excess risk for males exposed to 0.1 Gy at age 10 followed up until age 90 with ERR transfer using CER of the R library.

```
> res <- CER( agex=10, doseGy=0.1, sex=1, maxage=90 )
      mean      median      CI.2.5%      CI.97.5%
      0.012187454  0.011915813  0.007580173  0.018061663
> plotCER(res)
```

4. Chronic Exposure

The following box shows how to compute the cumulated excess risk for females exposed to 0.01 Gy/yr at ages 10 to 19 followed up until age 100 with 7:3 weights for ERR and EAR transfers.

```
> CER( agex=10:19, doseGy=rep(0.01,10), sex=2, maxage=100, wgt=c(.7,.3) )
      mean      median      CI.2.5%      CI.97.5%
      0.01738177  0.01719287  0.01336100  0.02247304
```