Notes to user: This repo contains an empty “Results” folder if one would like to generate one’s own data. If one wants to use the archived data (pre-simulated), please contact the authors (since the files are very large).

**Code Repository: step-by-step instructions**

1. Open A.SA\_Main\_Codes
   1. Codes without a “A<#>” prefix are helper codes and not run by themselves.
   2. Run codes starting with A1 (**A1.build\_CFmodel.R** and **A1.build\_DCMmodel.R** - these build the models for both CF and DCM and can be run in any order).
   3. Run **A2.localSA.R** to generate local SA indices
   4. Run **A3.SobolMorris.R** to generate Sobol Indices / Morris Indices
      1. Must change global.type = “Morris” or “Sobol”
      2. Results are saved in Results/SA\_Results folder
   5. Run A4.ReverseDosimetry\_IndpSamples.R for the HED percentiles.
      1. Results are saved in Results/R\_data\_tiefight folder as HEDLIST\_seedx.Rdata for x = seed (10 different seeds); each list in HEDLIST has 4 elements (inhale.DCM, oral.DCM, inhale.CF, oral.CF) and each element has 3 doses; for each dose is a matrix of HED values (n+1 x 10000), where n is the number of influential parameters that have been fixed and the +1 corresponds to the full MC analysis with all parms varying.
2. Open B.PostHocAnalysisCodes
   1. Open and run **B1.Plot\_SAFigs\_Manuscript.R** which generates figures for manuscript and supplementary document. It calls to **data\_arrange.R,** which organizes data.

* SharedCodes contains shared codes. Things like dosing regimes, the solve function, etc. are in **Specs\_funs.R**. Things related to SA are in **SensAnalysis\_funs.R**

**Complete Descriptions**

The code repository includes the following folders (relevant .R code are in **bold**):

1. MCSim: Contains files necessary for installing RMCSim.

* First, MCSim must be able to made to run through R. Relevant files are found in SourceCode/MCSim/. Installation instructions are in SourceCode/MCSim/RMCSim\_Package\_Base/ RMCSim\_install\_instructions.doc and testing RMCSim is found in SourceCode/MCSim/RMCSim\_Package\_Base/ RMCSim\_test\_instructions.doc.

1. A.SA\_Main\_Codes folder contains the following files (note that files that start with “A#” should be run in the order of the number and files without the “A#” prefix are not run on their own):
   1. **pbpk\_dcm.model** and **pbpk\_cf.model**: These .model files contain the PBPK models for DCM and chloroform.
   2. **A1.build\_DCMmodel.R** and **A1.build\_CFmodel.R:** compiles **pbpk\_dcm.model** and **pbpk\_cf.model** respectively to create necessary initial parameter files, .c, .o, and .dll (or .so, depending on the OS), which are all needed when you load the models in **RATPOD\_DCM.R** and **RATPOD\_CF.R.**
   3. **RATPOD\_DCM.R** and **RATPOD\_CF.R:** Calculates the human internal dose for DCM and chloroform inhalation and oral exposure by calculating the rat internal dose and applying a scaling factor.
   4. **A2.LocalSA.R**: Performs local sensitivity analysis and saves results (“localSA\_<subpopulationAge>.Rdata”) in the Results/SA\_Results folder. Has following variables that are set:

numsim = 10000 (generates 10k samples to find average parameter values)

sexm = "female" (“male” or “both”)

agem = 30 (0 for general population)

is.global = FALSE (if set to TRUE, then you are performing GSA and parameters are generated differently).

* 1. **A3.SobolMorris.R:** Performs global sensitivity analysis (both Sobol or Morris methods). Saves results in Results/SA\_Results as “Effects<method>\_GlobalSA\_30female.Rdata” for the output each method generates (matrices of HEDs by dose) and “Indices<method>\_GlobalSA\_30female.Rdata” for the calculated GSA indices. The subfolder global\_MetaData contains metadata for each chemical/route/dose that is saved incrementally.
     1. Loads **Specs\_funs.R** and **SensAnalysis\_funs.R** from SharedCodes folder.
     2. The user can set

If.par = TRUE or FALSE (for parallel processing)

parallelType = “WINDOWS” or “UNIX” (creates .so files and sets up parallel processes depending on the OS)

agem

sexm

global.type = “Morris” or “Sobol” – depending on the type, it will set up the design of experiments based on each method by generating “parmat”, a matrix of size n x M, where n = number of simulated individuals and M = number of parameters, where each value is in [0,1] and pertains to a percentile of the corresponding parameter distribution.

* 1. **A4.ReverseDosimetry\_IndpSamples.R**: For each chemical (chloroform and DCM), route of exposure (oral and inhalation) and dose (1, 50, 400 ppm or 0.5, 45, 400 mg/kg), HEDs are calculated when all parameters vary and then when most influential parameters vary one at a time. Data is saved to Results/R\_data\_tiefight.
     1. Sources SharedCodes/<**Specs\_funs.R>** <**SensAnalysis\_funs.R**> A.SA\_Main\_Codes /<**RATPOD\_DCM.R**><**RATPOD\_CF.R>**
     2. Loads Results/IndicesSobol\_GlobalSA\_.Rdata to load the parameter’s Sobol indices. Source B.PostHocAnalysisCodes/ **data\_arrange.R** to generate parameter ranking.
     3. User can specify ifpar, parallelType, agem, sexm in addition to numsim = 10000 (number of simulations) and seed (any integer. Setting K = 10 different seeds will generate K = 10 independent results).
     4. Parameter values are generated from **DCMhuman\_par.R** and **CFhuman\_par.R.**
  2. **DCMhuman\_par.R** and **CFhuman\_par.R**: Sources SharedCodes/ <**sampling\_funcs.R>** <**human\_par\_funs.R**> and contains parameter means/ standard deviations to generate parameter for each simulation.

1. The SharedCodes folder contains codes shared by multiple files:
   1. **Specs\_funs.R** sets specifications for each chemical/route—doses, dosing structure, time frames, the solve model function, the cost function, and the function if the user wants to parallelize the code (makeParallel = TRUE).
   2. **SensAnalysis\_funs.R** contains functions for sensitivity analysis—setting up Morris and Sobol design of experiments, and organizing parameters.
   3. **human\_par\_funs.R:** Contains functions to generate parameters drawn from physiological parameter distributions shared by both CF and DCM.
   4. **sampling\_funcs.R:** Contains functions to generate distributions for parameter distributions based on mean, standard deviation, and the user’s specification as to the “type” of distribution family (type = “orig” means that the parameters come from distributions originally found in literature and are a mix between truncated normal and truncated lognormal. Anything other than type = “orig” is an artifact from Schacht et al. (2024) and should not be used).
      1. The functions sampledis() and sampledis\_quantile() generate a sample from the distribution, given mean and standard deviation. sampledis() uses normbnd() and lnormbnd() and generates N samples (for the number of individual simulations) and is used when is.global=FALSE; sampledis\_quantile() uses normbnd\_quantile() and lnormbnd\_quantile() and returns N samples based on N percentiles.
2. B.PostHocAnalysisCodes contains codes to generate figures or arrange data after it has been generated:
   1. **data\_arrange.R**: This contains functions to clean up the data so it can be in the correct form for plotting.
   2. **B1.Plot\_SAFigs\_Manuscript.R**: This file contains plotting figures for the Figures/ManuscriptFigs and Figures/SupplementFigs (R will ask if it can create these files or you can create them yourself).
      1. Manuscript Figures: 2-panel-- most influential Sobol Indices (top) and bar graph relative error (bottom); HED Percentile plots (alternative estimate percentiles); grouped parameter GSA methods.
      2. Supplementary: 3-panel --Sobol Indices (all parameters, top), Morris 45 degree plots (middle), local methods (bottom)
3. Results folder contains the SA results in a folder called SA\_Results and the alternative estimate HED results in a folder called R\_data\_tiefight. Within R\_data\_tiefight, the complete HED files that are needed for post hoc analysis are labelled “HEDLIST\_seed<#>.Rdata” (for each seed, 10 total). R\_data\_tiefight contains a subfolder called meta which saves chemical/route data/dose as it becomes available (in case of a crash). R will create subfolders to hold data. Note that the file HEDLIST\_seed1028 is for seed 10 for DCM and seed 28 for CF.