Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

# Part 1: Data

This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Abstract

The data set used in this paper was aggregated data extracted from clinical studies on AML published between January 1, 2000, and Octorber 1, 2018. The data includes reconstructed survival time, hazard ratio estimates and survival rates. Demographic variables and some covariates of interests are also collected, such as study sample size, median patient age, median follow-up time, MRD detecton method, MRD assessment time pointss, AML subtype, specimen source.

## Availability

Data **are** publicly available

Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

### Publicly available data

Data are available online at:

Data are available as part of the paper’s supplementary material.

Data are publicly available by request, following the process described here:

Data are or will be made available through some other mechanism, described here:

### Non-publicly available data

Discussion of lack of publicly available data:

## Description

### File format(s)

CSV or other plain text:

Software-specific binary format (.Rda, Python pickle, etc.):

Standardized binary format (e.g., netCDF, HDF5, etc.):

Other (described here):

### Data dictionary

Provided by the authors in the following file(s): Data dictionary.txt

Data file(s) is (are) self-describiing (e.g., netCDF files)

Available at the following URL:

### Additional information (optional)

# Part 2: Code

## Abstract

The MARS method, including in simulation studies and an application of AML dataset, was conducted by a set of R codes. Figures, tables and real data example results can be reproduced by provided R scripts. The computationally intensive simulations were carried out on high performance computing cluster via batch scripts for job arrays.

## Description

### Code format(s)

Script files

R  Python  Matlab

Other:

Package

R  Python  MATLAB toolbox

Other:

Reproducible report

R Markdown  Jupyter notebook

Other:

Shell script

Other (described here):

### **Supporting software requirements**

Version of primary software used

R version 3.6.1, R version 4.0.2, JAGS version 4.3.0

Libraries and dependencies used by the code

Simsurv, survival, purrr, foreach, doParallel, parallel, R2jags, dclone, rjags, readr, ggpubr, writexl

### Supporting system/hardware requirements (optional)

### Parallelization used

No parallel code used

Multi-core parallelization on a single machine/node

Number of cores used:

Multi-machine/multi-node parallelization

Number of nodes and cores used: 20

### License

MIT License (default)

BSD

GPL v3.0

Creative Commons

Other (described here):

### Additional information (optional)

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

Any numbers proviided in text in the paper

All tables and figures in the paper

Selected tables and figures in the paper, as explained and justified here:

## Workflow details

### Format(s)

Single master code file

Wrapper (shell) script(s)

Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

Text file (e.g., a readme-style file) that documents workflow

Makefile

Other (more detail in 'Instructions' below)

### Instructions

The code provided can be used to reproduce all the results in section 5 and section 6.

Regarding simulation case 1, submit an array of jobs for simulation example 1.R. Pull the results into one R workspace by using Merge output.R. Figure 1 can be generated by Figure 1.R.

The workflow to reproduce the simulation case 2 is similar by replacing the R script by simulation example 2.R

To reproduce the application example, put the data sets provided in the working directory and execute Application example.R and Application example – type 1 data only.R files. Use Figure 3.R to generate the figure.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

<1 minute

1-10 minutes

10-60 minutes

1-8 hours

>8 hours

Not feasible to run on a desktop machine, as described here:

The simulation studies were implemented on a high performance computing system by submitting job arrays. Each job runs 42 replications, and the execution time is about . The analysis of real data example was performed on Intel Core i7. The execution time is about 9 hours.

### Additional documentation (optional)

# Notes (optional)