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

Framingham

The Framingham cholesterol study generated a benchmark dataset (Zhang and Davidian, 2001) for longitudinal analysis to examine the role of serum cholesterol as a risk factor for the evolution of cardiovascular disease for 200 randomly selected subjects.

Leukemia Survival

A dataset on the survival of acute myeloid leukemia in 1,043 patients, first analyzed by Henderson et al. (2002). It is of interest to investigate possible spatial variation in survival after accounting for known subject-specific prognostic factors, which include age, sex, white blood cell count (wbc) at diagnosis, and the Townsend score (tpi) for which higher values indicates less affluent areas.

Veteran Lung Cancer Survival

Survival in patients with advanced lung cancer from the North Central Cancer Treatment Group. Performance scores reate how well the patient can perform usual daily activities.

## 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: <https://github.com/manucarl/BCTM>

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:

R packages

Framingham Heart Study: data("Cholesterol", package="qrLMM")

Leukemia survival: data("LeukSurv", package="spBayesSurv")

Veteran lung cancer survival: data(cancer, package="survival")

### 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):

R packages (see above)

### Data dictionary

Provided by the authors in the following file(s):

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

Available at the following URL:

Available on respective GitHub repository and through package documentation.

### Additional information (optional)

# Part 2: Code

## Abstract

All code is available through BCTM.rar and respective GitHub repository

Folder

framingham contains all code to reproduce the results from the Framingham illustration

leukemia contains all code to reproduce the results from the leukemia survival illustration

veteran contains all code to reproduce the results from the veteran lung cancer illustration

sims contains all code to reproduce the results from the simulations

All scripts should be self-contained and can be started by simply clicking on run.

## Description

### Code format(s)

Script files

R  Python  Matlab

Other: rcpp

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-4.1.0

Libraries and dependencies used by the code

splines (4.21)

dplyr (1.0.10)

tidyverse (1.3.2)

mlt (1.4-2)

tram (0.7-2)

tramME (1.0.3)

scam (1.2-13)

mgcv (1.8-41)

bamlss (1.1-8)

Rcpp (1.0.9) RcppArmadillo (0.11.2.3.1) RcppEigen (0.3.3.92)

qrLMM (2.1)

viridis (0.6.2)

MASS (7.3-58.1)

mvtnorm (1.1-3)

cowplot (1.1.1)

gamboostLSS (2.0-6)

lattice (0.20-45)

microbenchmark (1.4.9)

progress (1.2.2)

devtools (2.4.4)

doParallel (1.0.17)

survival (3.4-0)

spBayesSurv (1.1.6)

Matrix (1.5-1)

MCMCpack (1.6-3)

profvis (0.3.7)

sf (1.0-8)

rgeos (0.5-9)

RhpcBLASctl (0.21-247.1)

Loo (2.5.1)

BayesX (0.3-1.1)

caret (6.0-93)

sdPrior (1.0-0)

scoringutils (1.0.1)

### 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)

Code is available in a public GitHub repository.

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

Any numbers provided in text in the paper

The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

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

Every script in code/framingham/, code/leukemia/, code/veteran and code/sims can be executed by clicking on “run” and reproduce plots and numbers mentioned in the paper.

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:

### Additional documentation (optional)

# Notes (optional)