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

In this work, we developed personalized test schedules for invasive diagnostic tests in the surveillance of chronic non-communicable diseases. We demonstrate our methodology for the prostate cancer surveillance scenario, using a model fitted to the world's largest prostate cancer active surveillance dataset abbreviated as PRIAS, dated April 2019. More than 100 medical centers from 17 countries contribute to PRIAS using a common study protocol, and web-based tool, both available at www.prias-project.org. The PRIAS dataset consists of patient age at inclusion, and repeated measurements data on biopsy Gleason grade groups (invasive diagnostic test), prostate-specific antigen or PSA (continuous: biomarker ng/mL), digital rectal examination or DRE (binary: tumor palpable or not), and magnetic resonance imaging or MRI (PI-RADS score).

## 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:

The PRIAS dataset is not openly accessible, but access to the dataset can be requested based on a study proposal approved by the PRIAS steering committee. The website of the PRIAS program is www.prias-project.org. We do provide a synthetic dataset instead of the original dataset. Readers can use this dataset along with the provided code to understand and recreate the workflow required to make personalized schedules in their datasets.

The \*\*main contribution of our work\*\* is not the model fitted to the PRIAS dataset, but rather the methodology proposed to schedule invasive diagnostic tests. The methodology is generic for use in other surveillance scenarios (e.g., scheduling endoscopies in Barett's esophagus). Hence, the lack of direct public access to the prostate cancer dataset PRIAS, which we only used for demonstration purposes, is not a major limitation. The simulation study we conducted utilizes the parameter estimates of the model fitted to the PRIAS dataset and can be reproduced without access to the PRIAS dataset itself.

## 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):

SPSS .sav file

### 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:

<https://anonymous.4open.science/r/d862487e-9a1a-4472-9564-ff2be4c625fd/>

### Additional information (optional)

# Part 2: Code

## Abstract

We provide R code for the following purposes:

1. To generate synthetic PRIAS like datasets.
2. For cleaning the original PRIAS dataset for readers who obtain a copy of the raw dataset from the PRIAS consortium.
3. To fit a joint model to the synthetic or real PRIAS dataset using the R package [**JMBayes**](https://cran.r-project.org/web/packages/JMbayes/index.html).
4. For reproducing the simulation study results comparing different biopsy schedules in PRIAS.
5. A generic API that can be used to schedule invasive tests in a personalized manner for any surveillance scenario. This API is compatible with R joint model objects fitted using the JMbayes package.
6. Code to reproduce figures.

## 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):

<https://anonymous.4open.science/r/d862487e-9a1a-4472-9564-ff2be4c625fd/>

### **Supporting software requirements**

Version of primary software used

R 3.6.1

Libraries and dependencies used by the code

In total the provided source code depends on four R packages:

* [JMBayes](https://cran.r-project.org/web/packages/JMbayes/index.html)
* [survival](https://cran.r-project.org/web/packages/survival/index.html)
* [MASS](https://cran.r-project.org/web/packages/MASS/index.html)
* splines: part of the R core now.

### Supporting system/hardware requirements (optional)

The code has been tested to work with 64 bit R-3.6.1 installed on a laptop with 4th generation Intel Core-i7 processor, and 8GB RAM.

### Parallelization used

No parallel code used

Multi-core parallelization on a single machine/node

Number of cores used: 2 to 10, users can customize it in our shell script

Multi-machine/multi-node parallelization

Number of nodes and cores used:

### 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:

Code for Figure 1 to 4 is not provided because they are only illustrative. Code for Figure 5 and Figure 6 has been provided. Supplementary tables showing parameter estimates are based on the MCMC samples of the posterior distribution of the parameters of our model. The fitted model object has been provided with the manuscript.

## 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

<https://anonymous.4open.science/r/d862487e-9a1a-4472-9564-ff2be4c625fd/>

Contains the README file and source code.

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:

Data analysis takes 1 to 8 hours, but simulation study requires a server computer. Computational times are provided in the README file.

<https://anonymous.4open.science/r/d862487e-9a1a-4472-9564-ff2be4c625fd/>

### Additional documentation (optional)

# Notes (optional)