

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.

Commented [A1]: This document will be read by readers after an accepted manuscript is published so please phrase your responses for future readers. Please see <https://jasa-acs.github.io/repro-guide/pages/acc.html> for details, links to examples, and information on how to save your completed form as a PDF. In particular make sure to remove these comments when saving as a PDF.

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).

Commented [A2]: If this box is checked, please skip directly to the Code section. Otherwise, continue.

☒ 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).

Commented [A3]: A short (< 100 words) description of the data. More details can be provided in files accompanying the data.

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 data* 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:

Commented [CP4]: If the data are only available by request to the authors or some other data owner, please make sure to explain the process of requesting access to the data.

☐ 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 Barrett'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.

Commented [A5]: If the data for this manuscript are publicly available, skip to the Description section below. Otherwise, continue.

The Journal of the American Statistical Association requires authors to make data accompanying their papers available to the scientific community except in cases where: 1) public sharing of data would be impossible, 2) suitable synthetic data are provided which allow the main analyses to be replicated (recognizing that results may differ from the "real" data analyses), and 3) the scientific value of the results and methods outweigh the lack of reproducibility.

Commented [A6]: For example:

- why data sharing is not possible,
- what synthetic data are provided, and
- why the value of the paper's scientific contribution outweighs the lack of reproducibility.

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

Commented [CP7]: Check all that apply.

Data dictionary

- ☐ Provided by the authors in the following file(s):
☐ Data file(s) is (are) self-describing (e.g., netCDF files)

Commented [CP8]: A data dictionary provides information that allows users to understand the meaning, format, and use of the data.

☒ Available at the following URL:

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

Additional information (optional)

Commented [CP9]: Provide any details that would be helpful in understanding the data. If relevant, provide unique identifier/DOL/version information and/or license/terms of use.

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](#).
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.

Commented [A10]: A short (< 100 words) description of the code. If necessary, more details can be provided in files that accompany the code.

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):

Commented [CP11]: Check all that apply.

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

Supporting software requirements

Version of primary software used

R 3.6.1

Commented [CP12]: For example, R version 3.6.2.

Libraries and dependencies used by the code

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

- [JMBayes](#)
- [survival](#)
- [MASS](#)
- splines: part of the R core now.

Commented [CP13]: Include version numbers (e.g., version numbers for any R or Python packages used)

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.

Commented [A14]: System/hardware requirements including operating system with version number, access to cluster, GPUs, etc.

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)

Commented [CP15]: By default, submitted code will be published on the JASA Github repository (<https://github.com/jasa-acss>) as well as in the supplementary material. Authors are encouraged to also make their code available in a public repository. If relevant, please provide unique identifier/DOI/version information here.

Part 3: Reproducibility workflow

Commented [CP16]: The materials provided should provide a straightforward way for reviewers and readers to reproduce analyses with as few steps as possible.

Scope

The provided workflow reproduces:

- ☐ Any numbers provided 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)

Commented [CP17]: Check all that apply

- ☐ 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

Commented [A18]: Describe how to use the materials provided to reproduce analyses in the manuscript. Additional details can be provided in file(s) accompanying the reproducibility materials.

<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)

Commented [A19]: Additional documentation provided (e.g., R package vignettes, demos or other examples) that show how to use the provided code/software in other settings.

Notes (optional)

Commented [A20]: Any other relevant information not covered on this form. If reproducibility materials are not publicly available at the time of submission, please provide information here on how the reviewers can view the materials (and make sure to remove this information when submitting the final version of this form for an accepted manuscript).