



# MSM.app: An Interactive Web Tool for Survival Analysis and Advanced Multi-State Modeling

Gustavo Soutinho

Research on Economics, Management and Information  
Technologies (REMIT) - Portucalense University  
Porto, Portugal  
gustavo.soutinho@upt.pt

Luís Meira-Machado

Centre of Mathematics, University of Minho  
Guimarães, Portugal  
lmachado@math.uminho.pt

## ABSTRACT

The development of applications aimed at providing interpretable results in a concise and user-friendly manner within the framework of multi-state models represents a promising research avenue, particularly when leveraging open-source tools adaptable to biomedical contexts. This paper introduces MSM.app, an interactive web application constructed using the Shiny package for the R language. MSM.app is structured into two main components, each addressing distinct facets of survival analysis and its extension to intricate multi-state models. The first component allows users to perform classical survival analysis techniques using standard functions from R packages such as survival. The second component supports users in achieving key objectives in multi-state analysis, including the inference of regression models and the estimation of transition probabilities, by integrating with the survdm and mstate R packages. The web application employs dynamic web forms, tables, and graphics, harnessing the capabilities of the Shiny package. This architecture ensures that users, regardless of their level of expertise in informatics, can seamlessly engage in dynamic analyses encompassing critical aspects of multi-state models.

## CCS CONCEPTS

• **Mathematics of computing** → **Statistical software.**

## KEYWORDS

R language, shiny package, survival analysis, multi-state models

### ACM Reference Format:

Gustavo Soutinho and Luís Meira-Machado. 2024. MSM.app: An Interactive Web Tool for Survival Analysis and Advanced Multi-State Modeling. In *2024 7th International Conference on Mathematics and Statistics (ICoMS 2024)*, June 23–25, 2024, Amarante, Portugal. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3686592.3686594>

## 1 INTRODUCTION

Over recent years, data visualization has surged in importance and popularity across a diverse range of fields. This trend has been accompanied by the development of a new array of web tools, facilitating researchers in aggregating and disseminating information to

pique the interest of the research community [1]. Such web applications encompass databases, algorithms, services, and software tools, offering several advantages: accessibility, eliminating the need for software package installations; efficient utilization of computational resources; seamless updates and access to the latest versions; and the capability for real-time validation of previous analyses [2].

Since the mid-2000s, R [3] has emerged as one of the most pivotal programming languages in academia and business realms, owing to its robust capabilities in data processing, statistical analysis, and visualization. According to IEEE, R stands as the primary competitor to Python in the realm of data science, consistently maintaining its position at the forefront of the list of preferred languages for employment opportunities [4].

With the increasing availability of data and the growing need for data-driven decision-making, the demand for interactive and engaging data visualization tools is on the rise. In this context, the relevance of Shiny web applications based on the R language is noteworthy. In fact, Shiny has become recognized as an effective tool for creating interactive outputs and graphs via web applications. Once developed, these web tools can be shared through the cloud and run on any browser, enabling users of all computing skill levels to dynamically analyze datasets.

In recent years, particularly within biomedical applications, several examples highlight this trend, such as MSDshiny, MSM-shiny, and MSMplus. These applications streamline various aspects of multi-state model analysis, including planning and powering clinical trials, conducting simulation studies, inferring transition probabilities, and identifying factors influencing transition intensities [5-7]. However, after analyzing existing web applications, we concluded that their use by non-statisticians has been limited. A significant reason for this is the lack of user-friendly software that integrates the main objectives of survival analysis and multi-state models into a single platform.

To address this gap, we developed a Shiny application called MSM.app. This application enables users to explore various types of multi-state models, perform regression inference, and obtain several predictive measures of interest, such as occupation probabilities, transition probabilities, and cumulative incidence functions.

### 1.1 The shiny R package

The introduction of the Shiny R package has revolutionized the presentation of outputs from the R language. These applications offer an interactive display of results that can be accessed either locally or via the internet [8]. With Shiny, developers gain access to a plethora of graphical user interfaces, empowering them to finely control the appearance and behavior of their applications,



This work is licensed under a Creative Commons Attribution International 4.0 License.

*ICoMS 2024, June 23–25, 2024, Amarante, Portugal*  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0722-3/24/06  
<https://doi.org/10.1145/3686592.3686594>

much like the functionality of HTTP wrappers. What’s remarkable is that Shiny operates solely within the R language framework, eliminating the need for expertise in JavaScript or HTTP. As a result, the process of creating and deploying web applications on both Windows and Linux servers has been greatly simplified for programmers [9].

## 1.2 Web applications for survival analysis

Several web tools are available for analyzing multi-state survival data, including MSDshiny, MSM-shiny, and MSMplus. MSDshiny stands out for its ability to efficiently plan and power clinical trials with multi-state outcomes. This application offers a clear view of the multi-state structure, treatment effects, and enables simulations [5]. On the other hand, MSM-shiny utilizes a CSV file containing multi-state data to facilitate modeling and comparison of transition hazard models, as well as prediction of occupation probabilities [6]. A noteworthy addition to this field is MSMplus, which offers flexible visualization of transition probabilities, transition intensities, and the probability of visiting specific states [7].

## 1.3 The MSM.app application

This paper presents the MSM.app application, a comprehensive tool that integrates the Shiny package with essential R packages such as survival [10], mstate [11], and survidm [12] for multi-state analysis. This versatile web tool enables users to perform a range of traditional survival analyses, including: (i) estimating survival functions utilizing the classical Kaplan-Meier estimator; (ii) comparing survival functions across different groups; and (iii) employing semi-parametric and parametric regression models to investigate the association between explanatory variables and survival time.

The MSM.app is not only designed for traditional survival analysis but also serves as a robust platform for multi-state survival data analysis. Multi-state models offer a broader perspective, treating survival as the ultimate outcome while considering intermediate events that individuals may encounter during the study period. Key objectives in multi-state modeling include inferring regression models and estimating transition probabilities, which can be influenced by Markovian processes. MSM.app addresses all these aspects comprehensively.

We believe there’s a gap in user-friendly web applications tailored for analyzing survival data with one or more events of interest, and MSM.app aims to fill this void. Our goal is to offer a comprehensive web tool enabling users to conduct dynamic data analysis covering essential topics in both survival analysis and multi-state models.

This paper is structured as follows: Section 2 provides a concise overview of the mathematical foundations underlying multi-state models and the development of the Shiny application. Section 3 details the various pages within the web tool, accessible through the Shiny Apps repository at <https://gsoutinho.shinyapps.io/appmsm/>. Finally, Section 4 presents the main conclusions drawn from our work.

## 2 METHODS

Survival analysis encompasses a suite of statistical techniques tailored for analyzing data where the primary outcome of interest is

the time until an event occurs. In biomedical contexts, these events could range from time to death to the time until the first recurrence of a tumor following treatment initiation. However, due to deliberate study design or random censoring, the precise time of event occurrence may not always be observable. Right censoring, for instance, occurs when the event of interest has not manifested by the time the data is evaluated.

In many survival studies, individuals may experience multiple events of interest within a given follow-up period. For instance, a patient treated for a tumor may later encounter local recurrence, metastases, or eventual mortality. This can be represented by a multi-state model which is a model for a time continuous stochastic process  $\{X_i(t)\}$  with a state space  $\{0, 1, \dots, K\}$ . These models [13–16] offer a robust framework for analyzing such complex scenarios effectively. The selection of an appropriate multi-state regression model and estimator for transition probabilities hinges on the Markov assumption.

In case of the transition probabilities, for two states  $h, j$  and two time points  $s < t$ , these quantities are given by the conditional probability  $p_{hj}(s, t) = P(X(t) = j \mid X(s) = h, \mathcal{H}_{s-})$ , where  $\mathcal{H}_{s-}$  denotes the history of the process up to time  $s$ . This history can represent information about the process, such as the states previously visited, transition times, etc. Under the Markov assumption,  $P(X(t) = j \mid X(s) = h, X(u) = y) = P(X(t) = j \mid X(s) = h)$  for any  $0 \leq u < s$  and  $y \in S$ , and thus, the future of the process after time  $s$  depends only on the present state at time  $s$  and not on the past history [17].

Typically, the Markov assumption is assessed by incorporating covariates that depend on the individual’s history, often utilizing Cox regression models. Among various methods, MSM.app employs a novel approach based on assessing the discrepancy of the Aalen-Johansen estimator, which yields consistent estimators within Markov processes. Additionally, recent methodologies that circumvent strict reliance on the Markov assumption have been employed [17]. The test statistic of these approaches is derived from the disparity between the areas under two respective curves, offering enhanced robustness and flexibility in analysis.

The structure of Shiny applications revolves around two key components: the user-interface scripts (*ui.R*) responsible for defining the layout of the application and displaying outputs, and the server scripts (*server.R*) that provide instructions for the application’s functionality. These applications can be executed locally on a localhost environment where both R and the Shiny package are installed, offering a straightforward method of sharing when users are familiar with R. Simply executing the instructions suffices to run the app [4].

## 3 THE MSM.APP IN PRACTICE

MSM.app offers an interactive platform featuring web forms, reports, tables, and graphics, all seamlessly integrated to facilitate analysis. To illustrate its capabilities in both survival and multi-state modeling, we present three real-world data examples: The first one involves data from survival in patients with Acute Myelogenous Leukemia [18]. The second corresponds to data from a clinical trial on colon cancer, which can be modeled using the progressive illness-death model [19]. Finally, extensions to progressive

processes beyond the three-state illness-death model are discussed using data from the European Group for Blood and Marrow Transplantation (EBMT) [14].

All these practical examples, drawn from the field of health-care, reflect a variety of similar situations encountered in hospital settings, such as evaluating survival rates or predicting disease diagnoses. In this regard, the authors of this application have initiated discussions about its potential use in a hospital unit in an experimental context. User feedback and usability results from case studies may be included in an extended version of this paper.

For a comprehensive understanding of the upcoming content, we recommend simultaneous use of the MSM.app application, which is accessible at <https://gsoutinho.shinyapps.io/appmsm/>. This hands-on approach will enhance comprehension and engagement with the material.

### 3.1 Introduction page

This page introduces an overview of the theoretical concepts addressed in MSM.app which are presented through check boxes. These HTML elements allow one to show or hide the contents described previously in Section 2: ‘survival analysis’, ‘Multi-state models’ and ‘Shiny applications architecture’ as well as the main goals of the web tool in MSM.app (Figure 1).

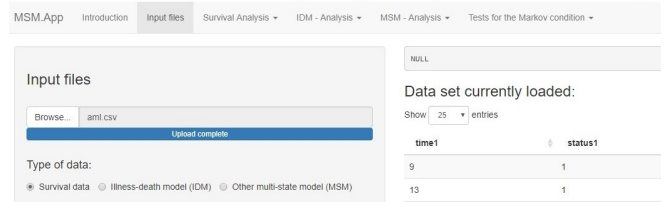


**Figure 1: Introduction web page:** Through four check boxes, the user can see a brief description of the mathematical background underlying MSM.app. The main goals of the web application are also included.

### 3.2 Input file page

This allows the user to select the necessary datasets for implementing the analysis. To this regard, there is a selection box to choose the dataset to be used. Upon file upload, users specify the type of analysis corresponding to the dataset by selecting from three options: "Survival Data", "Illness-Death Model", or "Multi-state Model". The latter should be chosen if the dataset comprises more than three states or reversible transitions, facilitated by radio button elements (Figure 2). In terms of format, the MSM.app only requires a CSV file as input. These files are delimited, utilizing commas to separate values, with each line representing a data record. Additionally, files employing semicolons for value separation are also compatible.

Once the data type is selected, a new web form dynamically appears below the radio buttons. For survival analysis, users are prompted to identify the "event time" and "status indicator" variables. In the case of an illness-death model, users first input the three state names of the model in the "names of the states" text box. Subsequently, they select the "time to the intermediate state", "status indicator of entering the intermediate state", "time to the ultimate



**Figure 2: The input file page with the data table with some results for the *aml* data and three radio buttons for each type of model.**

state", and "status indicator of entering the ultimate state" variables. For more intricate models such as MSM, users are presented with additional options including the "number of states", "transitions schema", "name of states", "variables for event times", and "variables for event status", facilitated by text box HTML elements.<sup>1</sup>

MSM.app further enhances user comprehension by offering three dataset examples, each tailored to represent a distinct type of analysis.<sup>2</sup>

### 3.3 Survival analysis pages

By selecting the "Survival Analysis" button, users gain access to classical methods for survival analysis, as depicted in Figure 3. Utilizing the Kaplan-Meier estimator, users can estimate overall survival or survival within different groups. Furthermore, the log-rank (or Gehan-Wilcoxon) test facilitates comparisons between curves from two or more groups. For assessing multiple predictors simultaneously, users can employ either the Cox model or the accelerated failure time model (AFT). These analyses leverage various functions within the survival R package [10], providing users with a comprehensive toolkit for survival analysis.

**3.3.1 Kaplan-Meier estimator.** The traditional non-parametric estimation of the survival function is commonly conducted using the Kaplan-Meier estimator [20]. Within this page, users can execute this estimation, with the outputs showcasing summaries of estimates and plots displayed on the right side. These estimates of survival curves can be stratified for different groups of categorical variables via the dropdown menu labeled "select variable." Initially, the checkbox for obtaining plots is disabled, but users can enable it to display the survival graph on the right side.

By default, confidence intervals for the plots are not selected. However, users can opt to include confidence intervals by selecting

<sup>1</sup>In many datasets, categorical variables are represented by numeric indices. However, these variables require special handling as they cannot be directly incorporated into regression models. Instead, they need to be recoded into binary variables, a task conveniently facilitated on this page. Additionally, users have the option to remove unnecessary variables from the analysis. This can be achieved via three text boxes, where users input the index order of each variable in the dataset.

Upon clicking the "upload" button, a dynamic data table appears on the right-hand side of the page. Users can interact with this table using filters or by conducting keyword searches. At any point, users can also review the dataset's variable names and types, as well as the most recently uploaded dataset, by accessing the "view structure of data" feature. This comprehensive functionality ensures users have complete control over dataset preparation and exploration within MSM.app.

<sup>2</sup>These CSV files are conveniently accessible via links located on the left side of the page, positioned at the bottom of the "View structure of data" section. This feature streamlines the learning process, enabling users to explore and experiment with different datasets to better grasp the functionality and utility of the web tool.

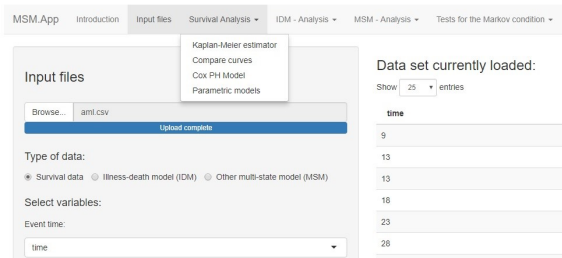


Figure 3: Output with the *tabPanels* shiny elements for the pages to obtain the survival analysis.

"yes" in the corresponding radio button. Additionally, users have the option to export the plots of the survival curves to PDF format by simply clicking the "Download PDF" button (Figure 4). This comprehensive functionality empowers users to conduct and visualize Kaplan-Meier estimations efficiently within MSM.app.

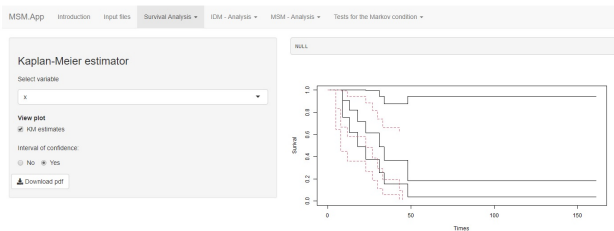


Figure 4: The output of the survival estimation for the categorical covariate “x” of the *aml* dataset using the Kaplan-Meier estimator. Survival curves for each group with confidence intervals are also shown at the bottom.

**3.3.2 Comparison of survival curves.** The log-rank test stands as the most widely utilized statistical test in survival analysis, assigning equal weights to all time points. However, when survival curves intersect at early time points, it becomes essential to employ methods that assign more weight to “deaths” occurring early in the study. The Gehan-Wilcoxon test addresses this need, offering an alternative approach. Both tests are readily accessible within MSM.app.

On the left side of the page, users encounter two radio buttons allowing selection between the two test methods, along with a dropdown list featuring all categorical variables, which are automatically loaded. The outputs displaying results from the tests are dynamically updated whenever users choose a specific test or covariate of interest (Figure 5). This seamless integration facilitates efficient comparison and interpretation of survival curves within the MSM.app environment.

**3.3.3 Cox PH models.** In the subsequent stage of survival analysis, we aim to assess the collective influence of multiple factors on survival. To achieve this, we employ the semi-parametric Cox proportional hazards model [21] on this page. Consistent with previous pages, the outputs detailing the fitted models are displayed on the right-hand side of the page. All potential covariates eligible

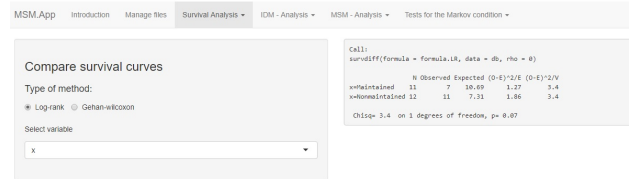


Figure 5: The output of the Log-rank test for the categorical covariate “x” of the *aml* data set for testing the null hypothesis of no difference in survival between the two groups.

for inclusion in the models are conveniently presented via checkbox HTML elements, streamlining the selection process for users. This approach ensures a comprehensive evaluation of the impact of various factors on survival within the MSM.app framework.

**3.3.4 Parametric models.** Parametric survival models can be fitted in a manner similar to the Cox proportional hazards model. However, in this scenario, users are required to not only select covariates from the checkboxes but also specify the distribution type for the regression models. Six distribution options are available on the left side of the page via radio buttons. Leveraging the interactivity of the Shiny package, the outputs of the fitted models are dynamically updated as users choose the appropriate variables and distribution type for the survival times. Additionally, the outputs include the Akaike Information Criterion (AIC) values for each model, facilitating straightforward comparison of model fit across different distributions (Figure 6).

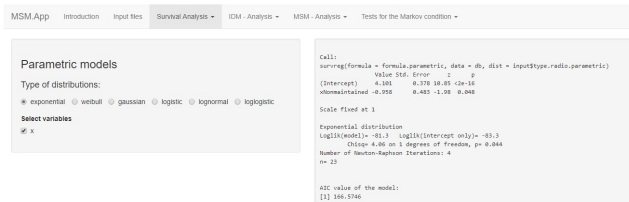


Figure 6: The outputs of the parametric models page with the results of the fitted model for the categorical variable “x” of the *aml* dataset using the exponential distribution.

### 3.4 The Illness-death models analysis pages

In this section, we outline the essential steps for achieving key objectives in analyzing data from a progressive illness-death model. This model typically involves two events, three states, and three transitions (Figure 7).<sup>3</sup>

**3.4.1 Number of events.** This page displays the count of individuals or items undergoing each transition among the three states of the illness-death model. Users can adjust the output format from “count” to “proportion” of transitions using the options provided on the

<sup>3</sup>Beyond calculating occupation probabilities and transition probabilities, it is crucial to compute cumulative incidence functions, which offer valuable predictive probabilities. Moreover, regression models for each of the three transition intensities enable the evaluation of prognostic factors for each transition. To facilitate these analyses, we leverage several functions from the *survdm* R package [12].

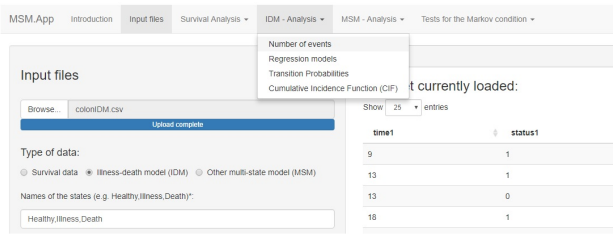


Figure 7: TabPanels shiny elements for the pages to obtain the illness-death model analysis.

left side of the page. This feature enhances flexibility in visualizing and interpreting transition dynamics within the model, catering to varying analytical needs and preferences.

3.4.2 *Regression models.* Covariates play a crucial role in multi-state models by influencing transition intensities, thus explaining differences among individuals in the progression of illness. This is achieved by disentangling the multi-state process into separate survival models, fitting distinct intensities to all permissible transitions using semi-parametric Cox proportional hazard regression models.<sup>4,5</sup>

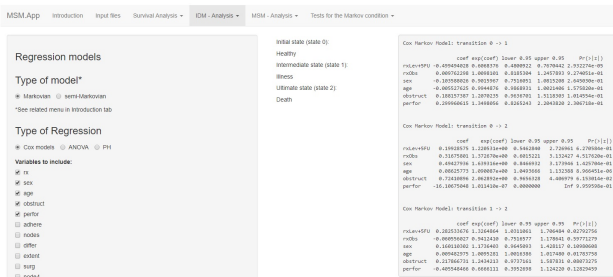


Figure 8: Results of the application of the Cox PH model with the following covariates: *rx*, *sex*, *age*, *obstruct* and *perfor*. Results for each of the three transition intensities of the *colonIDM*. A Markovian process is assumed.

3.4.3 *Transition probabilities.* Transition probabilities play a pivotal role in facilitating long-term predictions of the multi-state process. These quantities are typically estimated non-parametrically using the Aalen-Johansen estimator [22], assuming the system is Markovian. However, in scenarios where the multi-state model deviates from Markovian assumptions, the Aalen-Johansen estimator

<sup>4</sup>The effectiveness of this approach hinges on the assumption of dependence between transition intensities and time, known as the Markov assumption. Specifically, if past and future states are independent given the present state, the Markov assumption holds true. In scenarios where this assumption cannot be assumed, an alternative approach is to employ a semi-Markov model. In this framework, the future of the process is not contingent on the current time but rather on the duration spent in the current state.

<sup>5</sup>This page also provides outputs of ANOVA tests, along with corresponding p-values, to assess nonlinearity. Users can select these options via radio buttons on the web form. Moreover, users can conveniently select covariates for inclusion in the model using checkbox options. Similar to previous pages, as users make selections in the web form, the output of fitted models is automatically updated on the right side of the page (Figures 8). This seamless integration empowers users to explore and interpret regression models within the MSM.app environment efficiently.

may introduce bias. In such cases, it is preferable to utilize methods that do not rely on this assumption. Among these methods are estimators based on subsampling, also known as landmark methods. To be more specific, the landmark methods (LM) proposed by [23], as well as its presmoothed version (PLM) proposed by [24, 25], and the landmark Aalen-Johansen proposed by [26]. Specifically, MSM.app incorporates various non-parametric methods accessible through this page.<sup>6</sup>

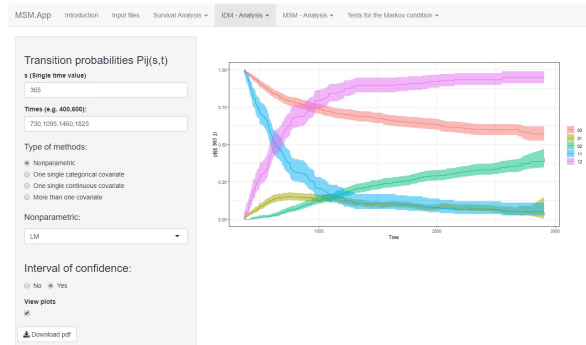


Figure 9: Transition probability estimates with confidence intervals for each transition for  $s = 365$  using the landmark estimators.

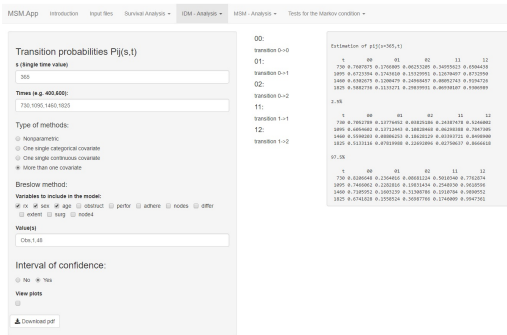
Incorporating categorical covariates using all four methods entails splitting the sample for each level of the covariate and repeating the analysis for each subsample. To accommodate the effect of a continuous covariate, MSM.app implements the nonparametric method proposed by [27] through the "IPCW" option in the radio button and the dropdown list "one single continuous covariate." This method relies on local smoothing, employing regression weights and appropriate reweighting of observations using Inverse Probability of Censoring Weighting (IPCW) to handle right censoring.<sup>7</sup>

### 3.5 MSM-Analysis pages

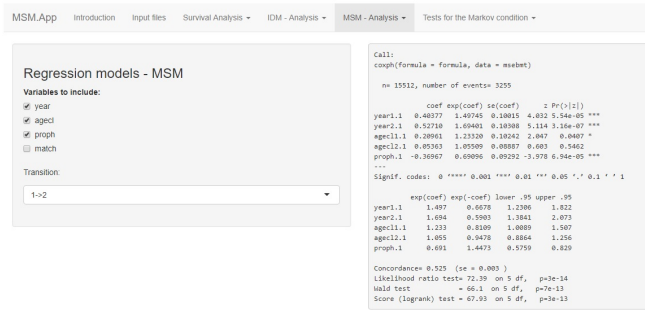
Clicking on the "MSM-analysis" button extends the methods discussed in the previous section to more complex multi-state models

<sup>6</sup>For instance, users can estimate transition probabilities from an initial timepoint (e.g.,  $s = 365$  days) to subsequent intervals over four years (730, 1095, 1460, and 1825 days) using the landmark approach. Users can input the desired timepoints into the first two text boxes on the left side of the page. This procedure remains consistent across all estimation methods, delineated by the subsequent four radio buttons. Upon selecting the landmark method (LM) under "Nonparametric," the results are automatically displayed on the right side of the page. Users can opt to visualize confidence intervals by clicking "yes" on the radio button. Additionally, users can view plots for each transition probability and export them using the last two HTML elements of the form (Figures 9). This comprehensive functionality enables users to explore and analyze transition probabilities using diverse estimation methods within the MSM.app environment.

<sup>7</sup>Another standard approach, particularly suitable for scenarios with multiple covariates, involves considering estimators based on a Cox's model fitted marginally to each type of transition, with the baseline hazard function estimated using Breslow's method [28]. Within the MSM.app web application, users can estimate transition probabilities conditional on several covariates by selecting "more than one covariate" in the radio button "type of methods." Subsequently, all possible covariates to include in the model appear as checkboxes on the left side of the form, along with a text box where users can input corresponding values. For instance, in the example depicted in Figure 10, three covariates ("rx", "sex", and "age") were selected for the model, with the respective values entered as Obs, 1, 48 (without spaces or quotation marks).



**Figure 10:** Results of the estimates of the transition probabilities and the corresponding confidence intervals for  $s = 365$  and times 730, 1090, 1460, and 1825 days for the *colonIDM* data set using the Breslow estimator.



**Figure 12:** The output of the Cox regression model for the transition  $1 \rightarrow 2$  that include the ‘year”, ‘age” and ‘proph” covariates. *ebmt4* dataset.



**Figure 11:** The number of transitions among the states of the *ebmt4* dataset.

involving more than three states, while also considering the possibility of reversible transitions. Three pages are available within this section, each serving to obtain the number of movements of the individual among states, the outputs of the regression models for each transition, and the results for the transition probabilities using the classical Aalen-Johansen estimator that assumes the process to be Markovian and the recent landmark Aalen-Johansen estimator that is free of this condition.

**3.5.1 Number of events.** This page provides visualizations of the movement of individuals among the states within the multi-state model. The outputs on the right side of the page automatically update based on the selection of the radio buttons labeled “count” or “proportion” of transitions (Figure 11).

**3.5.2 Regression models.** In a manner analogous to IDM models, our focus now shifts to estimating covariate effects for each transition within the multi-state model using Cox regression models. As users select which covariates to include in the models for each transition, summary outputs of the models are automatically displayed on the right side of the page. This selection process is facilitated through checkboxes presenting all possible covariates and dropdown lists offering all combinations of transitions. For example, in Figure 12, we observe the results of the Cox regression model for the transition  $1 \rightarrow 2$ , which incorporates the covariates “year,” “age,” and “proph.”

**3.5.3 Transition probabilities.** This page extends the estimation of transition probabilities to multi-state systems, which may involve more than three states with possible reversible transitions. The steps to obtain the results closely mirror those used for the

illness-death model. Users should begin by entering the initial time ( $s$ ) and other relevant times for obtaining the estimates in the text box elements. Two inference methods are available: the markovian Aalen-Johansen estimator (AJ) and the non-markovian LMAJ estimator, which can be selected via the dropdown list. Users can choose to display confidence intervals by selecting “yes”. Additionally, users have the option to export the corresponding plot to a PDF format file.

## 4 CONCLUSION

The shiny package has gained considerable traction for presenting interactive outputs and graphics through web applications. Once developed, these web tools can be easily shared via the cloud and accessed from any browser, enabling users of varying computational skills to dynamically analyze datasets. In the biomedical field, there has been a surge in the development of such web tools, with recent contributions focusing on facilitating multi-state analysis. Examples include MSDshiny, MSM-shiny, and MSMplus applications, which cater to specific aspects of multi-state model analysis such as clinical trial planning, simulation studies, and inference of transition probabilities.

However, despite the recognized importance of these models, existing web applications have seen limited adoption among non-statisticians. One significant reason for this is the absence of user-friendly software that encompasses the main objectives of survival analysis and multi-state models within a single platform. To address this gap, we have developed MSM.app, a shiny application designed to empower users to explore various types of multi-state models, conduct regression inference, and obtain essential predictive measures like occupation probabilities and transition probabilities.

Throughout this paper, we have showcased the main functionalities of MSM.app using three datasets, providing clear steps for conducting the analysis. As part of our ongoing research efforts, we aim to continually enhance MSM.app, addressing its limitations and incorporating recent advancements in the field.

MSM.app is readily accessible through the Shiny Apps repository at <https://gsoutinho.shinyapps.io/appmsm/>. We encourage users to explore the application and provide feedback for further improvements.

## ACKNOWLEDGMENTS

This work was supported by the UIDB/05105/2020 Program Contract, funded by funds through the FCT I.P.

## REFERENCES

- [1] Govan, P. (2016) eAnalytics: Dynamic Web-based Analytics for the Energy Industry. *Journal of Open Research Software*, 4: e45. DOI: <http://dx.doi.org/10.5334/jors.144>.
- [2] Łączky A. and Gyorffy B. (2021) Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. *J Med Internet Res*, 23(7). doi: 10.2196/27633.
- [3] R Core Team. R (2019) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org/>
- [4] Varma, J.R. and Virmani, V. (2017) Shiny alternative for Finance in the Classroom. Indian Institute of Management. W.P.
- [5] Peterson, R. (2019) MSDshiny: The Multistate Simulation Designer shiny Application, <https://github.com/petersonR/MSDshiny>.
- [6] Lacy, S. (2021) msm-shiny, <https://stulacy.shinyapps.io/msm-shiny/>.
- [7] Skourlis, N., Crowther, M.J., Andersson, T., Lambert, P.C. (2021) MSMplus. <https://nskbiostatistics.shinyapps.io/MSMplus/>
- [8] Wojciechowski, J., Hopkins, A. and Upton, R. (2015) Interactive pharmacometric applications using R and the Shiny package. *CPT: pharmacometrics & systems pharmacology*, 4(3), 146–159.
- [9] Powers, R., Kopp, B. and Martinez, W. (2016) Developing Tools for Analysis of Text Data, JSM 2016 - Section on Statistical Computing.
- [10] Therneau, T. M. (2021) A Package for Survival Analysis in R. <https://CRAN.R-project.org/package=survival>.
- [11] Putter, H., de Wreede, L.C., Fiocco, M., Geskus, R.B., Bonneville, E.F., Manevski, D. mstate (2020) Data Preparation, Estimation and Prediction in Multi-State Models. *The R Journal*.
- [12] Soutinho, G., Sestelo, M. and Meira-Machado, L. (2021) survldm: An R package for Inference and Prediction in an Illness-Death Model. *The R Journal*, *Accept for publication*.
- [13] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993) *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- [14] Putter, H., Fiocco, M. and Geskus, R.B. (2007) Tutorial in biostatistics: Competing risks and multi-state models, *Statistics in Medicine*, 26(11), 2389–2430.
- [15] Meira-Machado, L., de Uña-Álvarez, J. and Cadarso-Suárez, C. and Andersen, P.K. (2009) Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, 18, 195–222.
- [16] Meira-Machado, L. and Sestelo, M. (2019) Estimation in the progressive illness-death model: A nonexhaustive review. *Biometrical Journal*, 61 (2), 245–263.
- [17] Soutinho, G. and Meira-Machado, L. (2021) Methods for checking the Markov condition in multi-state survival data. *Comput Stat*. <https://doi.org/10.1007/s00180-021-01139-7>
- [18] Miller, R.G. (1997) *Survival Analysis*. John Wiley & Sons.
- [19] Moertel, G., Fleming, T.R., Macdonald, J.S., Haller, D.G., Laurie, J.A., Goodman, P.J., Ungerleider, J.S., Emerson, W.A., Tormey, D.C., Glick, J.H., Veeder, M.H. and Mailliard, J.A. (1990) Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *New England Journal of Medicine*, 322(6):352–358.
- [20] Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53: 457–481.
- [21] Cox, D.R. (1972) Regression models and life tables. *Journal of the Royal Statistical Society Series B*, 34, 187–220.
- [22] Aalen, O., Johansen, S. (1978) An empirical transition matrix for non homogeneous Markov and chains based on censored observations. *Scandinavian Journal of Statistics*, 5: 141–150.
- [23] de Uña-Álvarez, J. and Meira-Machado, L. (2015) Nonparametric estimation of transition probabilities in the non-markov illness-death model: A comparative study. *Biometrics*, 71(2), 364–375.
- [24] Meira-Machado, L. (2016) Smoothed landmark estimators of the transition probabilities. *SORT-Statistics and Operations Research Transactions*, 40: 375–398.
- [25] Soutinho, G., Meira-Machado, L., & Oliveira, P. (2022). A comparison of presmoothing methods in the estimation of transition probabilities. *Communications in Statistics - Simulation and Computation*, 51(9), 5202–5221. <https://doi.org/10.1080/03610918.2020.1762895>
- [26] Putter, H. and Spitoni, C. Non-parametric estimation of transition probabilities in non-markov multistate models: The landmark aalen-johansen estimator. *Statistical Methods in Medical Research*, 27, 2081–2092; 2018.
- [27] Meira-Machado, L., de Uña-Álvarez, J. and Datta, S. (2015) Nonparametric estimation of conditional transition probabilities in a non-Markov illness-death model. *Computational Statistics*, 30, 377–397.
- [28] Breslow, N. (1972) Discussion of paper by dr cox. *Journal of Royal Statistical Society, Series B*, 216–217. doi:10.1093/biostatistics/kxaa030.