Pseudo value-based Deep Neural Networks for Multi-state Survival Analysis

Md Mahmudur Rahman
mrahman6@umbc.edu
University of Maryland, Baltimore County
Baltimore, Maryland, USA

Sanjay Purushotham
psanjay@umbc.edu
University of Maryland, Baltimore County
Baltimore, Maryland, USA

ABSTRACT

Multi-state survival analysis (MSA) uses multi-state models for the analysis of time-to-event data. In medical applications, MSA can provide insights about the complex disease progression in patients. A key challenge in MSA is the accurate subject-specific prediction of multi-state model quantities such as transition probability and state occupation probability in the presence of censoring. Traditional multi-state methods such as Aalen-Johansen (AJ) estimators and Cox-based methods are respectively limited by Markov and proportional hazards assumptions and are infeasible for making subject-specific predictions. Neural ordinary differential equations for MSA relax these assumptions but are computationally expensive and do not directly model the transition probabilities. To address these limitations, we propose a new class of pseudo-value-based deep learning models for multi-state survival analysis, where we show that pseudo values - designed to handle censoring - can be a natural replacement for estimating the multi-state model quantities when derived from a consistent estimator. In particular, we provide an algorithm to derive pseudo values from consistent estimators to directly predict the multi-state survival quantities from the subject's covariates. Empirical results on synthetic and real-world datasets show that our proposed models achieve state-of-the-art results under various censoring settings.

CCS CONCEPTS

• **Mathematics of computing → Survival analysis**; • **Computing methodologies → Neural networks**.

KEYWORDS

Multi-state survival analysis, Neural networks, Pseudo values

1 INTRODUCTION

Multi-state survival analysis (MSA) is the problem of analyzing time-to-event data using multi-state models (MSM). Multi-state models [11] are models of a continuous-time stochastic process that capture the movement of subjects among a finite number of healthy and/or disease states. Thus, multi-state modeling can provide insights into the disease progression by providing a detailed view of disease or recovery trajectory in patients. This helps to predict the probability of future events after a given history and thus, can improve the clinician's decision-making ability for survival analysis. Figure 1 shows an example of a multi-state model for breast cancer progression. Here, a patient who is disease-free or had surgery can transition to locoregional relapse or distant relapse before reaching the death state. Multi-state survival analysis deals with the estimations of the multi-state quantities such as (a) **state occupation probability (SOP)** (the probability that a subject will be in a state
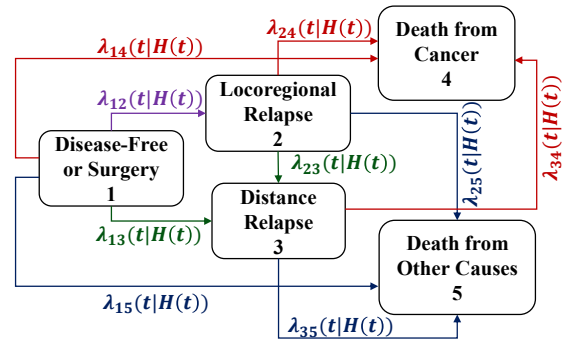

Figure 1: Graphical representation of a breast-cancer patient's multi-state model, where rectangles represent states ($S = \{1, 2, ..., 5\}$) and arrows represent the instantaneous state-to-state transitions. A disease-free patient or a patient who had surgery can move to two clinically relevant intermediate states, i.e., locoregional relapse and distant relapse, until one of the two absorbing states (cancer-specific death and death by other causes) is observed. $\lambda_{jk}(t|H(t))$ denotes the transition intensity functions.

k at time t,), (b) the **transition probability** (the probability of transition to a state k at time t from another state j at time s), and (c) **dynamic SOP** - which is the state occupation probabilities at some future time point t, given the event history (such as clinical information) is available up to a given time point s. A variety of statistical and machine learning approaches have been developed over the years including non-parametric Aalen-Johansen (AJ) estimators [2], Cox-based semi-parametric methods [6], parametric multi-state methods [12], and neural network-based methods -SurvNODE- [9] to estimate these multi-state quantities. Often, these approaches make strong assumptions such as linearity, proportional hazard, and Markov assumptions for each state or transition, which rarely hold in practice [14]. Moreover, many of these methods do not provide subject-specific predictions and do not handle the censoring well. Furthermore, these existing MSA methods cannot obtain subject-specific predictions of transition probability for non-Markov data because finding consistent estimators for non-Markov data has been understudied in the literature. Thus, new approaches that overcome these issues for multi-state survival analysis are in great demand.

In this paper, we introduce pseudo value-based deep learning models for multi-state survival analysis, denoted as **msPseudo**, for estimating the multi-state quantities by treating the complex multi-state survival modeling as a regression analysis problem. msPseudo consists of a deep neural network that takes covariates as inputs to estimate a multi-state quantity (e.g., SOP) via a pseudo

value regression task. msPseudo uses pseudo values as response variables because pseudo values have been shown to efficiently handle censoring for subject-specific survival predictions in survival analysis [20] and competing risks analysis [16]. Inspired by these works, we propose to use pseudo values as a replacement for multi-state quantities to handle censored observations. However, we cannot simply employ the estimators (Kaplan-Meier and Nelson–Aalen estimators) used in the earlier works [16, 20] to derive pseudo values for multi-state quantities since these estimators are inconsistent especially for real-world non-Markov data and can result in large estimation errors. Therefore, we introduce a simple algorithm to derive the pseudo values from consistent estimators such as AJ and Landmark AJ [18] estimators by testing the Markovianity of the data using statistical significance tests - Commenges-Andersen (CA) test [4] and log-rank statistic based tests [19]. Our algorithm provably obtains pseudo values from consistent estimators for both Markov and non-Markov data. Along with consistent pseudo-values, another advantage of our proposed model is that it does not make any underlying linear or proportional hazards assumptions and thus, can model the non-linear covariate effects during the prediction of subject-specific multi-state quantities. Therefore, our proposed msPseudo is simple yet flexible and overcomes the limitations of the existing multi-state survival models. We conducted extensive experiments on both simulated and real-world datasets to show that our proposed models achieve state-of-the-art performance in predicting multi-state survival quantities under various censoring settings.

2 OUR PROPOSED PSEUDO VALUE-BASED DEEP NEURAL NETWORKS

We first describe the derivation of pseudo values for multi-state quantities before discussing our proposed pseudo value-based deep neural networks.

Multi-state Survival Quantities: A multi-state process is a continuous time stochastic process $\{X(t), t \in \mathcal{T}\}$, taking values in the (discrete-state) finite state space $S = \{1, 2, ..., K\}$, where $\mathcal{T} = [0, \tau], \tau < \infty$. MSA deals with the estimation of the multi-state quantities [18]: **Transition probability** is the probability of transition to a state k at time t from another state j at time s, defined as $\mathcal{P}_{jk}(s, t) = P(X(t) = k|X(s) = j, H(s)), \forall j, k \in S$ **State Occupation Probability (SOP)** is the probability that a subject will be in state k at time t, defined as, $\pi_k(t) = P(X(t) = k); k \in S$; and **Dynamic SOP** is the SOP at some future time point t, given the event history is available up to a given time point s, defined as $\pi_k(t|s) = P(X(t) = k|H(s))$.

Pseudo values for multi-state quantities: Multi-state survival datasets are subject to censoring, i.e., incomplete information about the stochastic process (for example, event or transition information missing due to loss to follow-up). Therefore, direct modeling of the event time or status with respect to covariates is challenging for censored observations. Inspired by the recent works [16, 20], we propose to use pseudo values as a substitute for the estimation of subject-specific multi-state model quantities in the presence of censoring. Thus, we estimate the pseudo values for a multi-state quantity of interest as $\hat{y}_i(t^*) = n\hat{y}(t^*) - (n-1)\hat{y}^{-i}(t^*)$, where $\hat{y}(t^*)$ is an estimate of a consistent estimator, based on a n samples,

---

Algorithm 1: Pseudo value derivation algorithm

---
1 **Inputs:** Multi-state data, Selection of CA or log-rank test, ϵ
2 **Output:** Pseudo values
3 **For SOP**:
4 Choose either AJ or LMAJ estimator to derive pseudo values since both are consistent and same quantity [5].
5 **For Dynamic SOP**:
6 Perform CA test on the entire training data
7 **if** *P-value of the test is statistically significant for the violation of overall Markov assumption in the data* **then**
8 | Pseudo values ← LMAJ estimator;
9 **else**
10 | Pseudo values ← AJ estimator;
11 **For TP**:
12 Perform log rank test for checking transition-specific Markovianity.
13 **if** *P-value of the test is statistically significant for the violation of Markov assumption of a transition in the data*
14 **then**
15   **if** *landmark population $< \epsilon$* **then**
16 | Pseudo values ← AJ estimator;
17   **else**
18 | Pseudo values ← LMAJ estimator;
19 **else**
20 | Pseudo values ← AJ estimator;

$\hat{y}^{-i}(t^*)$ is an estimate from the same estimator based on leave-one-out $(n-1)$ samples, obtained by omitting the ith subject.

Consistent pseudo value derivation via Markov assumption testing: Pseudo values for MSA can be derived from an unbiased and consistent estimator such as the AJ estimator. Theoretical analysis of the consistency of the AJ and LMAJ estimators can be found in [15]. However, the AJ estimator is inconsistent for non-Markov data and can result in large estimation errors [19]. Thus, recently, researchers have proposed Landmark AJ (LMAJ) [15] as a consistent and robust estimator for estimating the pseudo values for non-Markov data. However, the AJ estimator is known to be more efficient than LMAJ [19] (when the Markov assumptions are valid), and in any practical scenario, the appropriateness of the Markov assumptions for a specific dataset remains unknown in advance, thus, making it infeasible to use just one estimator for pseudo value estimation. To address this important challenge, we introduce and describe a pseudo value derivation algorithm shown in **Algorithm 1** to efficiently derive pseudo values by selecting consistent estimators by testing the underlying Markovian assumptions. Our algorithm takes as input the multi-state survival data and selects a consistent estimator to obtain pseudo values by testing the Markovianity assumptions in the dataset by using statistical significance tests such as Commenges-Andersen (CA) test [4] or log-rank statistic-based tests [19]. CA and log-tank tests use a test statistic (usually χ^2 statistic) and its corresponding p-values to identify the violation of Markov assumptions in the data. Note that ϵ is chosen based on the minimum size of the population in a landmark state. We fix $\epsilon = 1$ in our experiments.

Proposed Model: We propose **msPseudo** - a first-of-its-kind, pseudo value-based deep learning model for multi-state survival analysis. msPseudo is a simple feedforward deep neural network which performs regression analysis to predict the multi-state quantities, such as state occupation probability (SOP), dynamic SOP, and transition probability (TP), using pseudo values as the response variables, given the covariates. msPseudo captures the complex nonlinear hidden relationship between the patient's characteristics, i.e., the baseline covariates and the multi-state model quantities. For an input $n \times p$ matrix of p baseline covariates with n individuals input, msPseudo returns the predictions of a multi-state quantity (SOP, dynamic SOP, or TP). For a multi-state dataset with K states, predicted output of SOP and dynamic SOP for a subject at a prespecified vector of M time points $\mathbf{t} = \{\tau_1, \tau_2, ..., \tau_M\}$ is a $K \times M$ matrix. For the TP prediction task, the output is a $Q \times M$ matrix, where Q is the number of transitions. We used the mean squared error (the mean squared difference between pseudo values (ground truth) and the predicted multi-state quantity) as the loss function (L) to train our msPseudo model.

3 EXPERIMENTS

We conducted experiments on both simulated and real-world datasets to answer the following questions: (a) How do our proposed models compare against the existing MSA approaches for predicting multi-state quantities? (b) How well do our proposed models perform under a variety of censoring settings compared to other models?

Simulation datasets: We generated the following four simulation datasets (two Markov and two non-Markov datasets): *(1) Time homogeneous Linear Markov Data; (2) Time homogeneous Nonlinear Markov Data; (3) Linear Reversible Non-Markov Data; (4) Nonlinear Reversible Non-Markov Data* - to obtain different datasets with varying Markov and linearity assumptions. For each dataset, we simulated 5000 examples with multiple transitions. The Markov datasets have three states and allow only forward transitions. The Non-Markov multi-state datasets consist of four states: states 1–3 are intermediate and interconnected states, and state 4 is an absorbing state, and they allow reverse transitions [10].

Real-world datasets: We used the following publicly available datasets for our experiments: (1) *METABRIC [17]* dataset contains 1975 breast cancer patient data with multiple transitions and 20 covariates collected over a 360-month study. This multi-state dataset has four states: Surgery, Locoregional Relapse, Distance Relapse, and Death. (2) *EBMT* [7] dataset contains 2279 transplantation patient data collected between 1985 and 1998. In this dataset, an alive patient in remission without recovery or adverse event can move to three possible distinct intermediate states, i.e., recovery, adverse event, and co-occurrence of recovery and adverse event, until one of the two absorbing states (death and relapse) is observed.

Censoring settings: We investigate the impact of the higher rate of censoring (75%) on MSA model performances under two censoring settings: incremental censoring (adding censored observations to a fixed number of uncensored observations) and induced censoring (inducing censored observations by flipping the label of transition status of the uncensored observations) settings [16].

Prediction tasks: Given the covariates, we perform regression for estimating the multi-state quantities such as SOP, dynamic SOP, and TP. We compare the performances of the following multi-state

models for these prediction tasks: **Non-parametric models:** AJ estimator (AJ) [1], LMAJ estimator (LMAJ) [15]; **Parametric models:** Weibull parametric model (msWeibull) [12], linear Pseudo value model (LinearPseudo) [3]; **Semi-parametric model:** Multi-state Cox proportional hazard model (msCox) [6]; **Deep learning multi-state model:** SurvNODE [9]; **Our proposed model:** msPseudo

Evaluations: We evaluate the models in terms of **integrated Brier score (iBS)** [18] and **integrated AUC (iAUC)** [8]. We perform 5 runs of 5-fold cross-validation and report the average of these evaluation metrics. We train our models using Adam optimizer [13] to 10000 epochs with an early stopping criterion. Hyperparameter tuning (over batch size, learning rate, drop-out, number of layers, etc.) is performed to choose the best performing deep learning models. A sigmoid activation function is used in the output layer to obtain the multi-state quantities from the predicted pseudo values.

4 RESULTS AND DISCUSSION

Simulated data: Table 1 shows that our msPseudo performs significantly better than msCox and SurvNODE for SOP prediction in Nonlinear Non-Markov dataset in terms of iAUC and iBS metrics. This shows that our models work well with non-Markov assumptions and can capture non-linearity in the data. Our msPseudo outperforms other models in terms of iAUC and is comparable in terms of iBS for the Non-linear Markov dataset. Table 4 shows that our model msPseudo performs significantly better than msCox and SurvNODE in SOP and Dynamic SOP prediction task on reversible Linear Non-Markov data. In the TP prediction task, msPseudo achieves similar or better results on 7 out of 9 transitions compared to the other models. We also graphically show the time-dependent Brier Score comparison in Figure 2 for dynamic SOP prediction on Linear Non-Markov data. This figure demonstrates that msPseudo achieves ~10% improvement over other multi-state models.

Real-world data: The predictive performances for the real-world clinical data: METABRIC and EBMT are shown in Table 2 & 3. Our model, msPseudo, outperforms all other models on METABRIC data for both SOP and Dynamic SOP prediction tasks. msPseudo also obtains lowest iBS for SOP prediction task on EBMT dataset while msCox performs similar or marginally better for Dynamic SOP predictions. In table 3, we observe that msPseudo gives overall better iBS performance, i.e., better prediction on both METABRIC and EBMT datasets. In some cases, our models give comparable performance to the msCox model due to the absence of covariates interaction effect, negligible violation of the proportional hazards and Markov assumptions. However, when averaged over all states and transitions (shown as the column Avg in all tables), we find that our proposed models outperform msCox and other MSA methods.

Various Censoring Settings: Table 5 shows the iAUC results of different multi-state models on incremental and induced censoring settings for SOP prediction in the time-homogeneous nonlinear Markov dataset. From this table, we see that our msPseudo performs significantly better (more efficient in handling censoring using pseudo values) than other models in both incremental and induced censoring settings with a high censoring rate (75% censoring).

5 CONCLUSION

Multi-state survival analysis (MSA) is an important yet under-studied problem in time-to-event literature. Finding consistent

Table 1: Comparison of SOP predictions on the Nonlinear Markov and Non-Markov datasets.

Algorithm	Nonlinear Markov								Nonlinear Non-Markov									
	iAUC (↑ better)				iBS (↓ better)				iAUC (↑ better)					iBS (↓ better)				
	S1	S2	S3	Avg	S1	S2	S3	Avg	S1	S2	S3	S4	Avg	S1	S2	S3	S4	Avg
msCoxPH	0.94	0.60	0.91	0.82	0.37	0.18	**0.08**	0.21	0.65	0.70	0.58	0.76	0.67	0.20	0.12	0.13	0.01	0.11
LinearPseudo	0.92	0.59	0.92	0.81	0.10	0.16	0.19	0.15	0.85	0.72	0.88	0.71	0.79	0.13	0.11	0.08	0.01	0.08
SurvNODE	0.95	0.67	0.91	0.84	0.16	**0.15**	0.09	0.13	0.83	0.70	0.76	0.54	0.71	0.29	0.14	0.16	0.01	0.15
msPseudo	**0.97**	**0.67**	**0.98**	**0.87**	**0.06**	0.16	0.16	**0.13**	**0.86**	**0.80**	**0.90**	**0.77**	**0.83**	**0.12**	**0.09**	**0.07**	**0.01**	**0.07**

Table 2: Comparison of the iBS (↓ better) scores for METABRIC & EBMT datasets.

| Prediction Task	Model	METABRIC					EBMT						
		State 1	State 2	State 3	State 4	Avg	State 1	State 2	State 3	State 4	State 5	State 6	Avg
State	msCox	0.33	0.03	0.04	0.24	0.16	0.13	0.14	0.11	0.15	**0.003**	0.15	0.12
Occupation	SurvNODE	0.30	0.03	0.04	0.21	0.14	0.34	0.14	0.11	0.17	0.003	0.17	0.16
Probability	msPseudo	**0.21**	**0.03**	**0.04**	0.17	**0.11**	**0.13**	**0.14**	**0.10**	**0.15**	0.01	**0.15**	**0.11**
Dynamic	msCox	0.32	0.03	0.03	0.23	0.15	**0.10**	**0.11**	**0.08**	**0.12**	0.01	**0.14**	**0.09**
SOP	SurvNODE	0.29	0.03	0.03	0.20	0.14	0.13	0.12	0.11	0.14	**0.001**	0.16	0.11
Prediction	msPseudo	**0.20**	**0.03**	**0.03**	**0.17**	**0.11**	0.11	0.12	0.09	0.13	0.02	**0.14**	0.10
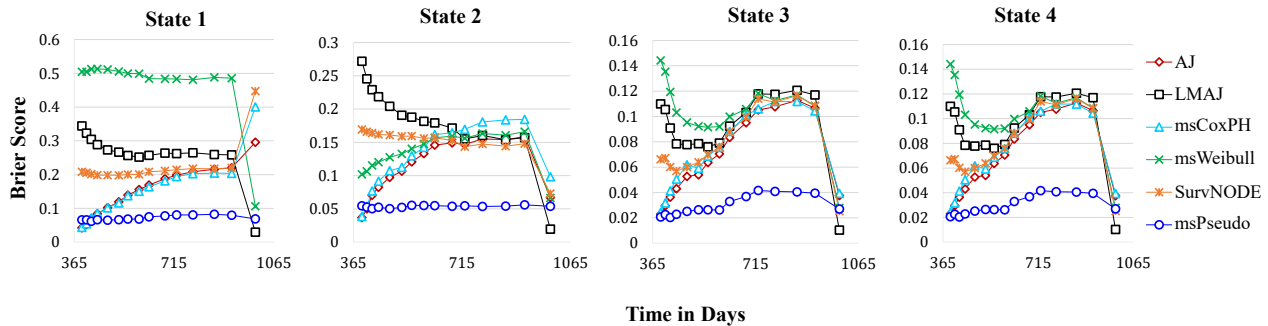
Figure 2: Comparison of time-dependent Brier score for the dynamic SOP prediction at s=1 year for reversible Linear Non-Markov dataset. Blue-circle represents our proposed msPseudo model.

Table 4: Comparison of the iBS (↓ better) scores for the Linear Non-Markov dataset.

Model	SOP					Dynamic SOP (s=1 year)				
	S1	S2	S3	S4	Avg	S1	S2	S3	S4	Avg
AJ	0.21	0.16	**0.11**	0.01	0.12	0.11	0.08	0.05	0.01	0.06
LMAJ	0.21	0.16	0.11	0.01	0.12	0.15	0.10	0.06	0.01	0.08
msCox	0.20	0.15	0.12	0.01	0.12	0.11	0.09	0.05	0.01	0.06
SurvNODE	0.33	0.20	0.14	0.01	0.17	0.14	0.09	0.05	0.01	0.07
msPseudo	**0.19**	**0.15**	0.12	**0.01**	0.12	**0.04**	**0.03**	**0.02**	**0.002**	**0.02**

Model	TP (s=1 year)									
	1→2	1→3	1→4	2→1	2→3	2→4	3→1	3→2	3→4	Avg
AJ	0.05	0.04	0.02	0.11	0.03	0.01	0.05	0.06	0.01	0.04
LMAJ	0.05	0.04	**0.001**	0.10	0.04	0.01	**0.02**	0.10	0.01	0.04
msCox	0.05	0.04	0.02	0.12	0.03	0.01	0.03	0.04	**0.003**	0.04
msWeibull	0.06	0.04	0.14	**0.05**	0.03	0.08	0.03	0.05	0.18	0.07
msPseudo	**0.05**	**0.04**	0.02	0.13	**0.03**	**0.01**	0.05	**0.03**	0.01	**0.04**

Table 5: Comparison of the iAUC on high censoring settings (75%) for SOP prediction on the Nonlinear Markov dataset.

Algorithm	Incremental Censoring				Incremental Censoring			
	S1	S2	S3	Avg	S1	S2	S3	Avg
msCox	0.54	0.54	0.52	0.53	0.52	0.52	0.53	0.52
msWeibull	0.51	0.51	0.52	0.51	0.52	0.53	0.52	0.52
LinearPseudo	0.52	0.53	0.52	0.52	0.51	0.52	0.53	0.52
SurvNODE	0.66	0.61	0.65	0.64	0.69	0.62	0.64	0.65
msPseudo	**0.90**	**0.65**	**0.74**	**0.77**	**0.88**	**0.63**	**0.66**	**0.72**

estimators for non-Markov data is still an open problem in this field. In this paper, we proposed a first-of-its-kind novel pseudo value-based deep learning model, **msPseudo**, for estimating multi-state survival quantities in the presence of censoring without making any assumption on the underlying multi-state processes. We show that pseudo values can be a replacement for the estimation

of multi-state quantities when derived from a consistent estimator. Through empirical experiments on simulated and real datasets, we demonstrated that our proposed models outperform other multi-state survival models under various censoring settings and for both Markov and non-Markov datasets. We believe this work lays the foundation for future investigations on the use of deep models for MSA, including explaining survival predictions and state-specific transition probabilities in real-world datasets.

ACKNOWLEDGEMENT

REFERENCES

[1] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. 2008. *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

[2] Odd O Aalen and Søren Johansen. 1978. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* (1978), 141–150.

[3] Per K Andersen and John P Klein. 2007. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scandinavian Journal of Statistics* 34, 1 (2007), 3–16.

[4] Daniel Commenges and Per Kragh Andersen. 1995. Score test of homogeneity for survival data. *Lifetime data analysis* 1, 2 (1995), 145–156.

[5] Somnath Datta and Glen A Satten. 2001. Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. *Statistics & probability letters* 55, 4 (2001), 403–411.

[6] Liesbeth C De Wreede, Marta Fiocco, and Hein Putter. 2010. The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Computer methods and programs in biomedicine* 99, 3 (2010), 261–274.

[7] Liesbeth C de Wreede, Marta Fiocco, Hein Putter, et al. 2011. mstate: an R package for the analysis of competing risks and multi-state models. *Journal of statistical software* 38, 7 (2011), 1–30.

[8] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 8 (2006), 861–874.

[9] Stefan Groha, Sebastian M Schmon, and Alexander Gusev. 2020. Neural ODEs for Multi-State Survival Analysis. *arXiv preprint arXiv:2006.04893* (2020).

[10] Rune Hoff, Hein Putter, Ingrid Sivesind Mehlum, and Jon Michael Gran. 2019. Landmark estimation of transition probabilities in non-Markov multi-state models with covariates. *Lifetime data analysis* 25, 4 (2019), 660–680.

[11] Philip Hougaard. 1999. Multi-state models: a review. *Lifetime data analysis* 5, 3 (1999), 239–264.

[12] Christopher H Jackson. 2016. flexsurv: a platform for parametric survival modeling in R. *Journal of statistical software* 70 (2016).

[13] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[14] Luís Meira-Machado, Jacobo de Uña-Álvarez, Carmen Cadarso-Suárez, and Per K Andersen. 2009. Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research* 18, 2 (2009), 195–222.

[15] Hein Putter and Cristian Spitoni. 2018. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen–Johansen estimator. *Statistical methods in medical research* 27, 7 (2018), 2081–2092.

[16] Md Mahmudur Rahman, Koji Matsuo, Shinya Matsuzaki, and Sanjay Purushotham. 2021. DeepPseudo: Pseudo Value Based Deep Learning Models for Competing Risk Analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 479–487.

[17] Oscar M Rueda, Stephen-John Sammut, Jose A Seoane, Suet-Feung Chin, Jennifer L Caswell-Jin, Maurizio Callari, Rajbir Batra, Bernard Pereira, Alejandra Bruna, H Raza Ali, et al. 2019. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* 567, 7748 (2019), 399–404.

[18] Cristian Spitoni, Violette Lammens, and Hein Putter. 2018. Prediction errors for state occupation and transition probabilities in multi-state models. *Biometrical Journal* 60, 1 (2018), 34–48.

[19] Andrew C Titman and Hein Putter. 2020. General tests of the Markov property in multi-state models. *Biostatistics* (2020).

[20] Lili Zhao and Dai Feng. 2020. Deep neural networks for survival analysis using pseudo values. *IEEE journal of biomedical and health informatics* 24, 11 (2020), 3308–3314.