Modeling of long-term survival data with unobserved dispersion via neural network

Led Red Teh[1], Vicente Garibay Cancho[1*], Josemar Rodrigues[1]

[1*]Department of Statistics, Institute of Mathematics and Computer Sciences, University of São Paulo, Av. Trab. São Carlense, 400 - Centro, São Carlos, 13566-590, São Paulo, Brazil.

*Corresponding author(s). E-mail(s): garibay@icmc.usp.br;
Contributing authors: ledred21@gmail.com; josemar@icmc.usp.br;

Abstract

Traditional models in survival analysis assume that every subject will eventually experience the event of interest in the study, such as death or disease recurrence, so the survival function is said to be proper. Cure rate model, which was first proposed seven decades ago, has since been used to account for the presence of cure fraction, this means that a certain fraction of the individuals will never experience the occurrence of an event of interest for which they can be treated as immune or cured subjects in the context of cancer treatment. In the literature, various cure rate models have been widely studied and commonly applied to structured data with small quantities of covariates. The use of convolutional neural network, a powerful deep learning technique for image processing problem, has become increasingly more common in the medical field in recent years. Medical images such as histological slides and magnetic resonance images (MRIs) are directly related to a patient's prognostic factors, therefore, it is reasonable to introduce these images as predictors in cure model. In this work, we extend upon the article of Xie and Yu (2021b) in which a neural network was used to model the unstructured predictor's effect in the promotion time cure model's setting to the cases of overdispersed data. We will call our extension as integrated negative binomial cure rate model, and its parameters will be estimated through the Expectation-Maximization algorithm.

Keywords: Survival analysis, Cure rate model, Convolutional Neural Network, OASIS-3 Alzheimer's disease dataset

1 Introduction

Survival analysis models time-to-event data, such as the time from a patient's diagnosis to his death. Traditional survival analysis models assume that all individuals are equally susceptible to the event, which is not always true. With medical advancements, some individuals may achieve complete recovery or have their illnesses controlled, making them immune or less susceptible to the event. This proportion of individuals, referred to as the cure fraction or cure rate, makes traditional models impractical.

Models that incorporate the cure fraction in survival analysis are commonly known as cure models. These models can be categorized into mixture models, promotion time models, and unified models. While they have been effectively applied, their use has been constrained to structured, low-dimensional datasets with few covariates, limiting their applicability to more complex and high-dimensional data scenarios.

In recent years, various machine learning classifiers have been increasingly applied to model the incidence component in cure models, thanks to their ability to capture nonlinear boundaries that separate cured and uncured individuals based on covariate effects. For example, decision trees and random forests have been applied to mixture models by Aselisewine and Pal (2023) and Jiang et al. (2019), respectively. In addition, support vector machines have been integrated in both mixture models (Li et al. (2020); Pal et al. (2023); Pal et al. (2024)) and promotion time models (Pal and Aselisewine (2023)). These approaches have demonstrated improved accuracy in incidence estimates compared to traditional methods such as the logistic link function.

Convolutional Neural Networks (CNN), a key deep learning technique, are widely used in computer vision. Meyer et al. (2018) proposed integrating deep learning into radiotherapy, while Lundervold and Lundervold (2019) studied CNN architectures tailored for MRIs tasks. Since features like cancer cells in histological images and organ conditions shown in MRI may influence oncology outcomes, these medical images can be introduced via CNN as predictors into cure models, as proposed by Xie and Yu (2021a) and Xie and Yu (2021b). Expanding upon the latter approach, where the number of cancer cells is assumed to follow a Poisson distribution, we aim to extend it to accommodate the scenario of overdispersed data by considering the negative binomial distribution, which is well known in the literature for its effectiveness in modeling count data with greater variability. The Poisson distribution is actually a limiting case when the dispersion parameter in negative binomial is approaching zero.

This article is organized as follows: Section 2 introduces the proposed cure model, namely the integrated two-stage cure rate model. The term "integrated" represents the integration of CNN in cure model to use images as predictors, while "two-stage" describes how the model is built with latency and incidence components. This section also explains how the model is constructed and estimated using the Expectation-Maximization (EM) algorithm for Poisson and negative binomial distributions. Section 3 presents the simulation results that compare the performance of the model in Poisson and negative binomial data, with the negative binomial model expected to perform better. Finally, Section 4 applies the negative binomial cure model to the OASIS-3 dataset, and compares the results with those from Xie and Yu (2021b).

2 An integrated two-stage cure rate model

2.1 Model's construction

As Tsodikov et al. (2003) had pointed out, a cure model could be formulated as a two-stage model based on tumor recurrence mechanistic reasoning. At the first stage, one assumes there is an unobservable discrete random variable M which represents a number of risk factors such as surviving carcinogenic cells after an initial treatment, where M is characterized by distribution $P(M|\boldsymbol{\beta}, \boldsymbol{x})$ that depends on a vector of covariates $\boldsymbol{x}$ and a vector of regression coefficients $\boldsymbol{\beta}$. The first stage can be interpreted as a latent tumor generating process where each remaining clonogen is associated with a latent progression time during which it will develop into a detectable tumor. At the second stage, the observed failure time T is generated when one of the risk factors is activated and the observed survival function can be obtained by $S_p(t|\boldsymbol{\beta}, \boldsymbol{x}) = \mathbb{E}\left[S(t)^M|\boldsymbol{x}\right]$.

As mentioned earlier, our goal is to integrate CNN with cure model in order to introduce medical images as predictors of the cure probability. The two-stage, long-term survival model mentioned above is actually a unified long-term survival model as described in Rodrigues et al. (2009). In this section, we propose an integrated two-stage cure rate model combines with CNN to relate clinical images to the latent number of risk factors at the first stage.

2.1.1 First stage

Let M be a discrete random variable denoting the number of risk factors or damaged cells with probability mass function

$$p_m = P_\Theta(M = m),$$

where $\theta = E(M)$. Since the parameter θ is positive, we consider the following parametrization

$$\theta(\boldsymbol{x}) = e^{\eta(\boldsymbol{x})}, \tag{1}$$

in which $\boldsymbol{x} \in \mathbb{R}^p$ is a vector of covariates with p dimensions. From the probabilistic view (Rodrigues et al., 2009, Section 2), the first stage can be characterized by the probability generating function of M given by

$$A_M(s) = \mathbb{E}\left[s^M\right] = \sum_{m=0}^{\infty} s^m p_m. \tag{2}$$

The convolutional neural network link function. The exponent $\eta(\boldsymbol{x})$ in Equation 1 is modeled after CNN. We use $\boldsymbol{w}$ to denote all the weights in the model. The output of the integrated neural network gives estimates of the mean number of risk factors in logarithmic scale.

2.1.2 Second stage

Let $Z_k, k = 1, \ldots, M$ be the unobservable progression time for kth latent risk factor where all Z_k are iid with a distribution function $F(\cdot)$ and survival function $S(\cdot) = 1 - F(\cdot)$. Also, Z_k are independent of M. Thus, the observable lifetime T is defined by

$$T = \min\{Z_1, \ldots, Z_M\}. \tag{3}$$

The susceptible individuals are the observations with $M \geq 1$, and the cured individuals are those with $M = 0, P(Z_0 = \infty) = 1$. As derived in Rodrigues et al. (2009), we can obtain the long-term survival model by

$$S_p(t) = A_M(S(t)). \tag{4}$$

In Equation 4, note that the latency part $S(t)$ is a proper survival function while the populational survival function $S_p(t)$ is improper. The cure probability is defined as the limit as t approaches infinity, $\lim_{t\to\infty} S_p(t) = P(M = 0) = p_0$.

2.2 Estimation procedures

The estimation method for the proposed integrated two-stage cure rate models is based on the frequentist approach. The EM algorithm has been widely implemented in the literature of cure model for finding the Maximum Likelihood Estimator (MLE) of the parameters. Gallardo et al. (2016) proposed a simplified estimation procedure for the power series cure rate model by maximizing the likelihood function of the complete data instead of the observed likelihood function, this approach requires the computation of the expected values of latent variable M which will facilitate the maximization step to be discussed next.

For a sample of size n, the complete data are $\boldsymbol{D}_{\text{comp}} = (\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{M})$ where $\boldsymbol{t} = (t_1, \cdots, t_n)^T$, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_n)^T, \boldsymbol{x} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^T$ and $\boldsymbol{M} = (M_1, \cdots, M_n)^T$, for the ith individual, the variable t_i represents the survival time. The variable δ_i is the censoring indicator, where $\delta_1 = 1$ if t_i is a failure time; otherwise, it indicates the t_i right censored. The variable $\boldsymbol{x}_i$ is a $p \times 1$ covariate vector. The term M_i denotes the number of latent risk factors or competing causes, which are unobservable.

The observed data are defined as $\boldsymbol{D}_{\text{obs}} = (\boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x})$. Let $\boldsymbol{\psi} = (\boldsymbol{w}, \boldsymbol{\alpha})$ denotes the set of parameters in which $\boldsymbol{w}$ is related to the weights of the neural network and $\boldsymbol{\alpha}$ corresponds to the parameters associated with the cdf $F(t)$ of the latent progression times of clonogenic cells. With the assumption of $Z \perp M$ and $Z_k \perp Z_l \mid M$ for $k \neq l$, the complete likelihood function is defined as

$$\begin{aligned} L\left(\boldsymbol{\psi}; \boldsymbol{D}_{\text{comp}}\right) &= \prod_{i=1}^{n} P\left(t_i, \delta_i \mid m_i\right) P_{\Theta_i}\left(M_i = m_i\right) \\ &= \underbrace{\left\{\prod_{i=1}^{n} \left[m_i f\left(t_i; \boldsymbol{\alpha}\right)\right]^{\delta_i} \left[S\left(t_i; \boldsymbol{\alpha}\right)\right]^{m_i - \delta_i}\right\}}_{L_2(\boldsymbol{\alpha}; \boldsymbol{D}_{\text{comp}})} \underbrace{\left\{\prod_{i=1}^{n} P_{\Theta_i}\left(M_i = m_i\right)\right\}}_{L_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})} \end{aligned} \tag{5}$$

where $\theta_i \in \Theta$ depends on some covariates in a relationship of $\theta(\boldsymbol{w}, \boldsymbol{x}_i) = e^{\eta(\boldsymbol{w}, \boldsymbol{x}_i)}$. The assumptions of $Z \perp M$ and $Z_k \perp Z_l \mid M$ are crucial to the factorization in 5 which allows a direct integration of the CNN, via $L_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})$, to obtain the MLE of $\boldsymbol{w}$. Hence, the complete log likelihood can be written as a decomposition

$$l(\boldsymbol{\psi}; \boldsymbol{D}_{\text{comp}}) = l_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}}) + l_2(\boldsymbol{\alpha}; \boldsymbol{D}_{\text{comp}}),$$

where $l_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})$ only depends on the parameters related to the mean number of risk factors (1st stage) and $l_2(\boldsymbol{\alpha}; \boldsymbol{D}_{\text{comp}})$ involves only parameters related to the distribution of the risk factor lifetime (2nd stage) when m_i is given. This complete log likelihood can also be maximized independently and separately.

Let $\boldsymbol{\psi}^{(k)} = (\boldsymbol{\alpha}^{(k)}, \boldsymbol{w}^{(k)})$ be the estimate of $\boldsymbol{\psi}$ at the kth iteration of the EM algorithm. At the $(k+1)$th iteration, we perform:

1. E-step: Compute the conditional expected value of M given $\boldsymbol{D}_{\text{obs}}$ and the current parameter estimates, for $i = 1, \cdots, n$,

$$m_i^{(k+1)} = \mathbb{E}[M_i + \delta_i \mid \boldsymbol{D}_{\text{obs}}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{w}^{(k)}].$$

2. M-step:
   (i) Find $\boldsymbol{w}^{(k+1)}$ that maximizes $l_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})$ via CNN where

$$l_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}}) = \sum_{i=1}^{n} \log P_{\theta_i^{(k)}}(M_i = m_i^{(k+1)}),$$
$$\theta_i^{(k)} = e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)}.$$

   (ii) Find $\boldsymbol{\alpha}^{(k+1)}$ that maximizes

$$l_2(\boldsymbol{\alpha}; \boldsymbol{D}_{\text{comp}}) = \sum_{i=1}^{n} \Big\{ \delta_i \log m_i^{(k+1)} + \delta_i \log \frac{f(t_i; \boldsymbol{\alpha}^{(k)})}{S(t_i; \boldsymbol{\alpha}^{(k)})} + m_i^{(k+1)} \log S(t_i; \boldsymbol{\alpha}^{(k)}) \Big\}.$$

We repeat the iterations until $\|\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)}\|_2^2 < \epsilon$ and $\|\boldsymbol{\alpha}^{(k+1)} - \boldsymbol{\alpha}^{(k)}\|_2^2 < \epsilon$, where ϵ is a defined tolerance and $\|\cdot\|$ a L_2 norm.

2.3 Some integrated models

In this section, we consider two distributions for M–Poisson and negative binomial– which will result in an integrated promotion time cure model (IPCM) and integrated negative binomial cure model (INBCM) respectively. Note that the IPCM is the same as the model used in Xie and Yu (2021b).

We consider a piecewise exponential distribution for T with time partitioned into G intervals, $0 < s_1 < \cdots < s_G$, with $s_G > \max_{1 \le i \le n}(t_i)$. We also assume a constant hazard α_g in the gth interval, $(0, s_1], (s_1, s_2], \cdots, (s_{g-1}, s_g]$. Therefore, the cdf for the

lifetime of risk factors is given by

$$F(t;\boldsymbol{\alpha}) = 1 - \exp\Big\{-\alpha_g\,(t - s_{g-1}) - \sum_{j=1}^{g-1}\alpha_j\,(s_j - s_{j-1})\Big\}, \tag{6}$$

in which $t \in (s_{g-1}, s_g]$ and $S(t;\boldsymbol{\alpha}) = 1 - F(t;\boldsymbol{\alpha})$.

The term of $L_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$ in Equation 5 varies according to its assumed distribution for risk factors, it is related to the weights of the convolutional neural network. The formulas for $L_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$ and $l_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$ are organized in Table 1, C denotes a constant independent of $\boldsymbol{\psi}$. The $L_2(\boldsymbol{\alpha};\boldsymbol{D}_{\text{comp}})$ term is the same for both models.

Table 1: Formulas for the first component of likelihood

$M \sim$ Poisson	
$L_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$	$\prod_{i=1}^{n} \frac{\left(e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}\right)^{m_i} e^{-e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}}}{m_i!}$
$l_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$	$\sum_{i=1}^{n} m_i\eta(\boldsymbol{w},\boldsymbol{x}_i) - e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)} + C$
$M \sim$ Negative binomial	
$L_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$	$\prod_{i=1}^{n} \frac{\Gamma\left(\phi^{-1}+m_i\right)}{\Gamma\left(\phi^{-1}\right)m_i!}\left(\frac{\phi e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}}{1+\phi e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}}\right)^{m_i}\left(\frac{1}{1+\phi e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}}\right)^{\phi^{-1}}$
$l_1(\boldsymbol{w};\boldsymbol{D}_{\text{comp}})$	$\sum_{i=1}^{n} m_i\eta(\boldsymbol{w},\boldsymbol{x}_i) - m_i\log\left(1+\phi e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}\right) - \phi^{-1}\log\left(1+\phi e^{\eta(\boldsymbol{w},\boldsymbol{x}_i)}\right) + C$

Using Equation 6, we can write the second component of the complete log likelihood as follow:

$$\begin{aligned} l_2(\boldsymbol{\alpha};\boldsymbol{D}_{\text{comp}}) &= \sum_{i=1}^{n}\Big\{\delta_i\log m_i + \delta_i\log\frac{f(t_i;\boldsymbol{\alpha})}{S(t_i;\boldsymbol{\alpha})} + m_i\log S(t_i;\boldsymbol{\alpha})\Big\} \\ &= \sum_{i=1}^{n}\Big\{\delta_i\log m_i + \delta_i\log\alpha_g - m_i\big\{\alpha_g(t-s_{g-1}) - \sum_{j=1}^{g-1}\alpha_j(s_j - s_{j-1})\big\}\Big\}. \end{aligned} \tag{7}$$

For each model, we first present the long-term survival representation based on Equation 4; a proposition followed by a corollary and using this corollary, we replace the values of the unobservable variable with its conditional expected values at the E-step of the estimation procedures; we then finalize each subsection with the integrated EM algorithm steps.

2.3.1 Integrated promotion time cure model (IPCM)

If $M \sim \text{Poisson}(\theta)$ with $P(M = m) = \frac{\theta^m}{m!}e^{-\theta},\ m = 0, 1, 2, \ldots,$ then

$$S_p(t) = A_M\big((S(t)\big) = e^{-\theta\left(1-S(t)\right)} = e^{-\theta F(t)}, \tag{8}$$

and $p_0 = e^{-\theta}$.

Proposition 1. *If $M_i \sim Poisson(\theta_i)$, then $M_i \mid \boldsymbol{D}_{obs}; \boldsymbol{\psi} \sim \text{Poisson}\left(\delta_i + \theta_i S\left(t_i; \boldsymbol{\alpha}\right)\right)$.*

Corollary 1. $\mathbb{E}\left(M_i \mid \boldsymbol{D}_{obs}; \boldsymbol{\psi}\right) = \delta_i + \theta_i S\left(t_i; \boldsymbol{\alpha}\right)$, *for $i = 1, \cdots, n$.*

Algorithm 1 Integrated promotion time cure model

---

while $|\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}^{(k-1)}| < \epsilon$, **do**

E step: calculates for $i = 1, \cdots, n$,

$$\widehat{m}_i^{(k+1)} = e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x})} S(t_i; \boldsymbol{\alpha}^{(k)}) + \delta_i,$$

$$\text{where } S(t_i; \boldsymbol{\alpha}^{(k)}) = \exp\Big\{-\alpha_g^{(k)}\left(t - s_{g-1}\right) - \sum_{j=1}^{g-1} \alpha_j^{(k)}\left(s_j - s_{j-1}\right)\Big\}$$

M step:

$$\boldsymbol{w}^{(k+1)} = \arg\max_{\boldsymbol{w}} \left\{\sum_{i=1}^{n} \widehat{m}_i^{(k+1)} \eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i) - e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)}\right\}$$

$$\alpha_g^{(k+1)} = \frac{\displaystyle\sum_{s_{g-1} < t_i \leq s_g} \delta_i}{\displaystyle\sum_{s_{g-1} < t_i \leq s_g} \widehat{m}_i^{(k+1)}\left(t_i - s_{g-1}\right) + \sum_{t_i > s_g} \widehat{m}_i^{(k+1)}\left(s_g - s_{g-1}\right)}$$

end while

2.3.2 Integrated negative binomial cure model (INBCM)

If $M \sim \text{Negative Binomial}(\frac{1}{\phi}, \frac{1}{1+\phi\theta})$ with

$$P(M = m) = \frac{\Gamma\left(\frac{1}{\phi} + m\right)}{\Gamma\left(\frac{1}{\phi}\right) m!}\left(\frac{\phi\theta}{1+\phi\theta}\right)^m \left(\frac{1}{1+\phi\theta}\right)^{\frac{1}{\phi}}, m = 0, 1, 2, \ldots, \tag{9}$$

for $\theta > 0$ and $\phi > -1/\theta$ such that $\mathbb{E}(M) = \theta$ e $\mathbb{V}(M) = \theta + \phi\theta^2$, then

$$S_p(t) = A_M\big((S(t)\big) = \left(\frac{1}{1 + \phi\theta\left(1 - S(t)\right)}\right)^{\frac{1}{\phi}} \tag{10}$$

and $p_0 = (1+\phi\theta)^{-\frac{1}{\phi}}$. As stated in Rodrigues et al. (2009), the Equation 9 contains a dispersion parameter ϕ, where $\phi > 0$ indicates over dispersion and under dispersion when $\phi < 0$. More interestingly, Equation 10 is a flexible model because it converges to the promotion time cure model when $\phi \to 0$; also, it becomes the mixture cure model proposed by Berkson and Gage (1952) when $\phi = -1$.

Proposition 2. *If* $M_i \sim$ Negative binomial$(\frac{1}{\phi}, \frac{1}{1+\phi\theta_i})$, *then*

$$M_i \mid \boldsymbol{D}_{obs}; \boldsymbol{\psi} \sim \text{Negative binomial}\left(\frac{1}{\phi} + \delta_i, \frac{1+\phi\theta_i F\left(t_i; \boldsymbol{\alpha}\right)}{1+\phi\theta_i}\right)$$

Corollary 2. *The conditional expectation of* M_i *given the observed data and parameter values is*

$$\mathbb{E}\left(M_i \mid \boldsymbol{D}_{obs}; \boldsymbol{\psi}\right) = \frac{(1+\phi\delta_i)\theta_i S\left(t_i; \boldsymbol{\alpha}\right)}{1+\phi\theta_i F\left(t_i; \boldsymbol{\alpha}\right)}, \text{ for } i = 1, \cdots, n.$$

Algorithm 2 Integrated negative binomial cure rate model

while $|\boldsymbol{\psi}^{(k)} - \boldsymbol{\psi}^{(k-1)}| < \epsilon$, **do**

E step: calculates for $i = 1, \cdots, n$,

$$\widehat{m}_i^{(k+1)} = \frac{(1+\phi\delta_i)\, e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)} S(t_i; \boldsymbol{\alpha}^{(k)})}{1+\phi e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)} F(t_i; \boldsymbol{\alpha}^{(k)})},$$

$$\text{where } S(t_i; \boldsymbol{\alpha}^{(k)}) = \exp\left\{-\alpha_g^{(k)}\left(t - s_{g-1}\right) - \sum_{j=1}^{g-1} \alpha_j^{(k)}\left(s_j - s_{j-1}\right)\right\}$$

M step:

$$\boldsymbol{w}^{(k+1)} = \arg\max_{\boldsymbol{w}} \left\{\sum_{i=1}^{n} \widehat{m}_i^{(k+1)} \log\left\{\frac{\phi e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)}}{1+\phi e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)}}\right\} - \frac{1}{\phi}\log\left(1+\phi e^{\eta(\boldsymbol{w}^{(k)}, \boldsymbol{x}_i)}\right)\right\}$$

$$\alpha_g^{(k+1)} = \frac{\sum_{s_{g-1} < t_i \le s_g} \delta_i}{\sum_{s_{g-1} < t_i \le s_g} \widehat{m}_i^{(k+1)}\left(t_i - s_{g-1}\right) + \sum_{t_i > s_g} \widehat{m}_i^{(k+1)}\left(s_g - s_{g-1}\right)}$$

end while

3 Simulation study

In this section, we aim to assess the performance of our proposed method through simulations. We explore two scenarios for the number of competing risks: one where M follows a Poisson distribution, and another where it follows a negative binomial distribution. Our implementation procedure closely aligns with the methodology outlined in Xie and Yu (2021b).

We implemented the methodology outlined in Section 2.2 by using Python within the Google Colab Pro environment. The key advantage of utilizing Google Colab Pro is its access to GPU resources which significantly accelerates the execution of deep neural networks. The CNN model was developed using `TensorFlow v2.9.2`. Our approach involved running 30 iterations of the EM algorithm. The image predictors $\boldsymbol{x}$ are samples taken from the Fashion-MNIST dataset, with each entry being a grayscale image with a dimension of 28×28 and associated with one of the 10 clothing classes labeled from 0 to 9. Images were loaded from the built-in datasets of the Keras deep learning library, using the function `keras.datasets.fashion_mnist.load_data()`. These image predictors, denoted as $\boldsymbol{x}$, will be linked to the mean parameter through a neural network, defined as $\theta(\boldsymbol{x}) = e^{\eta(\boldsymbol{x})}$. In our simulation study, we consider only the first five classes as shown in Figure 1, and each labeled class is then assigned a fixed numeric value for $\theta(\boldsymbol{x})$.

The convolutional network model is the same as the one used in the application of Xie and Yu (2021b). We followed the authors' recommendations regarding the determination of network architecture. Among three different network structures, we selected the one that yielded the lowest likelihood value of $l_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})$, as shown in Table 1. The details of the model's architecture are presented in Table 2.

Fig. 1: Examples of the first five clothing classes corresponding to labels 0 through 4, 0 - T-shirt/top, 1 - Trousers, 2 - Pullover, 3 - Dress and 4 - Coat

Table 2: Architecture of the convolutional network model applied on the Fashion-MNIST sample images

Layer	Dimension	Activation σ	Number of parameters
Convolution 1	(28, 28, 4)	ReLU	104
Pooling 1	(14, 14, 4)	max	-
Convolution 2	(14, 14, 12)	ReLU	1212
Pooling 2	(7, 7, 12)	max	-
Convolution 3	(7, 7, 32)	ReLU	9632
Pooling 3	(3, 3, 32)	max	-
Flatten	(288)	-	-
Fully connected	(128)	tanh	36992
Output	(1)	-	128
Total	-	-	48068

The lifetimes for the latent risk factors $Z_1, \cdots, Z_M$ are generated from an exponential distribution characterized by a constant hazard rate of 1, leading to the survival function of the form $S_1(t) = e^{-t}$. In all of the scenarios, we maintain the average cure rate and overall censoring rate at approximately 40% and 45%, respectively. To explore the impact of varying sample sizes, we utilize three distinct training set sizes (n_{train}): 500, 1000, and 5000. And the corresponding test set sizes (n_{test}) are fixed at 125, 250, and 1250, respectively.

To evaluate the estimation accuracy, we follow the methodology outlined in the referenced paper by performing 100 replications on the estimations and calculate the following mean differences on the test sets:

1. The mean squared error for the populational survival function

$$\Delta\overline{S}_p(t) = \frac{1}{R \times n_{test}} \sum_{r=1}^{R} \sum_{i=1}^{n_{test}} \left(\widehat{S}_p\left(t_i\right) - S_p^0\left(t_i\right)\right)^2;$$

2. The mean squared error for the competing risk survival lifetime function

$$\Delta\overline{S}_1(t) = \frac{1}{R \times n_{test}} \sum_{r=1}^{R} \sum_{i=1}^{n_{test}} \left(\widehat{S}_1\left(t_i\right) - S_1^0\left(t_i\right)\right)^2;$$

3. The mean squared error for the cure rate

$$\Delta\overline{p}_0 = \frac{1}{R \times n_{test}} \sum_{r=1}^{R} \sum_{i=1}^{n_{test}} \left(\widehat{p}_0\left(t_i\right) - p_0^0\left(t_i\right)\right)^2;$$

4. The mean squared error for the mean risk factors in logarithm scale

$$\Delta\overline{\eta}(x) = \frac{1}{R \times n_{test}} \sum_{r=1}^{R} \sum_{i=1}^{n_{test}} \left(\widehat{\eta}\left(\boldsymbol{x}_i\right) - \eta^0\left(\boldsymbol{x}_i\right)\right)^2;$$

5. The mean absolute error for the number of risk factors

$$\Delta\overline{m} = \frac{1}{R \times n_{test}} \sum_{r=1}^{R} \sum_{i=1}^{n_{test}} \mid \widehat{m}_i - m_i^0 \mid,$$

where R represents the number of replicates and the superscripted 0 represents the true measures.

To evaluate the predictive capacity of the studied models in terms of the cure fraction, we adopt the Area Under the ROC Curve (AUC) metric as employed in the referenced paper.

3.1 Case where M follows the Poisson distribution

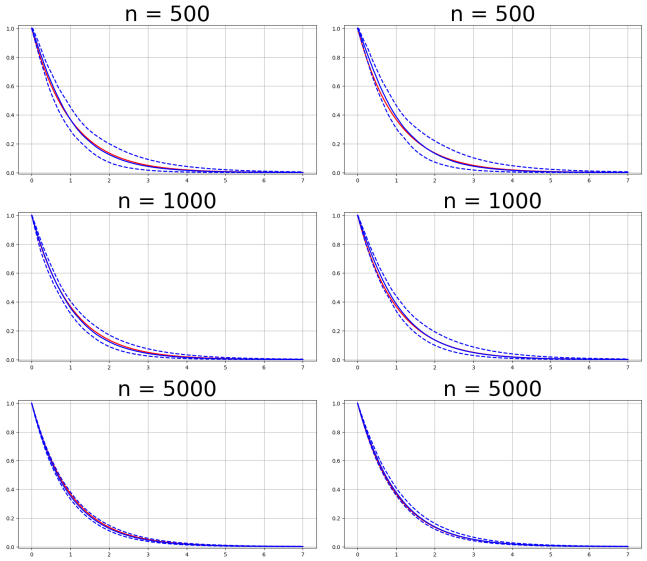
We first study the integrated promotion time cure model case in which the competing risk variable M follows the Poisson distribution. In this case, we set $\theta(\boldsymbol{x}) =$

$0.1, 0.8, 1.5, 2, 2.5$ and then we generate m_i first for each observation $i, \cdots, n$ according to the given value of $\theta(\boldsymbol{x}_i)$. For cases where $m_i = 0$, we assign an infinite lifetime $y_i = \infty$. When $m_i > 0$, we proceed to generate the promotion time of each risk factors $z_1, \cdots, z_{m_i}$, and the corresponding lifetime would be $y_i = \min\{z_1, \cdots, z_{m_i}\}$. Next, we generate the censoring time c_i from a uniform distribution, the observed lifetime is then obtained from $t_i = \min(y_i, c_i)$. If $t_i < c_i$, we would set the censoring indicator variable $\delta_i = 1$, otherwise, $\delta_i = 0$.

Table 3: Estimation accuracy of both models presented as the mean errors calculated from the test set

Model	time	n_{train}	$\Delta\overline{S}_p(t)$	$\Delta\overline{S}_1(t)$	$\Delta\overline{\eta}(x)$	$\Delta\overline{p_0}$	$\Delta\overline{m}$
IPCM	23min 34s	500	0.031885	0.001125	0.811638	0.042243	0.713748
	22min 30s	1000	0.023671	0.000319	0.592476	0.031686	0.629459
	58min 47s	5000	0.017309	0.000123	0.436762	0.023141	0.567622
INBCM with fixed ϕ	28min 7s	500	0.034323	0.001741	0.923800	0.043460	0.797420
	31min 34s	1000	0.026620	0.000755	0.676092	0.033795	0.706891
	63min 9s	5000	0.018399	0.000172	0.461016	0.023513	0.600151

The results of estimation accuracy are displayed in Table 3. When the proposed integrated negative binomial model is applied to this dataset, the dispersion parameter is set at a fixed value of 0.1. It is noteworthy that the outcomes from both models exhibit a high degree of similarity across all sample sizes. The results align consistently with the estimated $S_1(t)$, which can be visualized in Figure 2. The true $S_1(t)$ curves are depicted by solid red lines, while the estimated curves and their corresponding empirical 95% confidence interval are represented by solid and dotted blue lines respectively. In both models, the confidence band narrows as the sample size increases. The estimated AUC values for both models demonstrate exceptional predictive capability as shown in Table 4, with the integrated negative model surpassing 85%.



(a) IPCM (b) INBCM

Fig. 2: Estimated curves of $S_1(t)$, assuming M follows a Poisson distribution in the generated dataset

Table 4: AUC of both models for three different sample sizes

Model	n_{train}	Train set	Test set
IPCM	500	0.862100	0.844204
	1000	0.848263	0.839596
	5000	0.833405	0.830950
INBCM with fixed ϕ	500	0.868450	0.852270
	1000	0.868474	0.859518
	5000	0.855207	0.852680

3.2 Case where M follows the negative binomial distribution

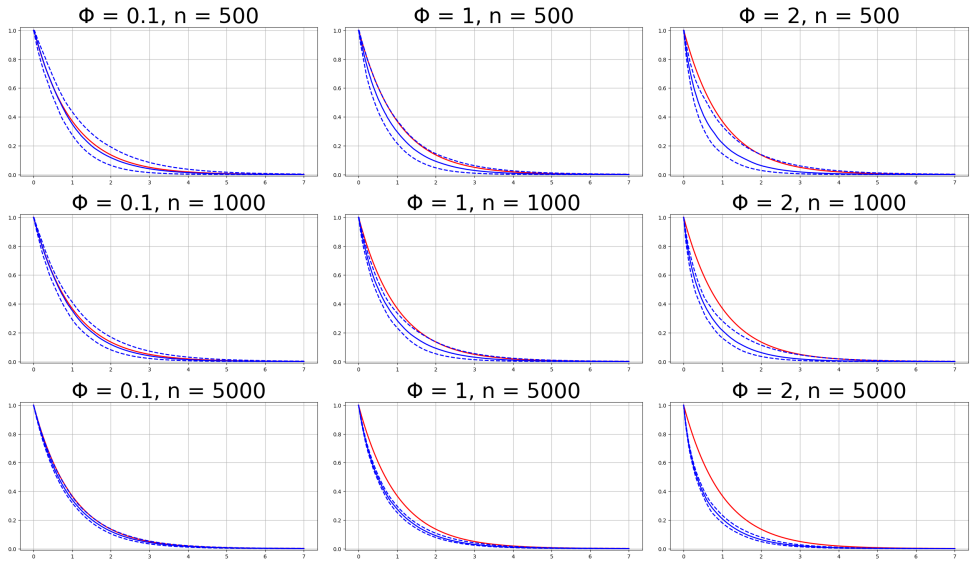
When we assume that the competing risk variable M follows a negative binomial distribution, we explore three distinct values for the dispersion parameter ϕ, 0.1, 1 and 2. The data generation process is analogous to the one described in the Poisson scenario.

From Table 5, it is convincing to show that the integrated promotion time cure model does not yield satisfactory results when applied to the overdispersed dataset.

Table 5: Estimation accuracy of both models presented as the mean errors calculated from the test set

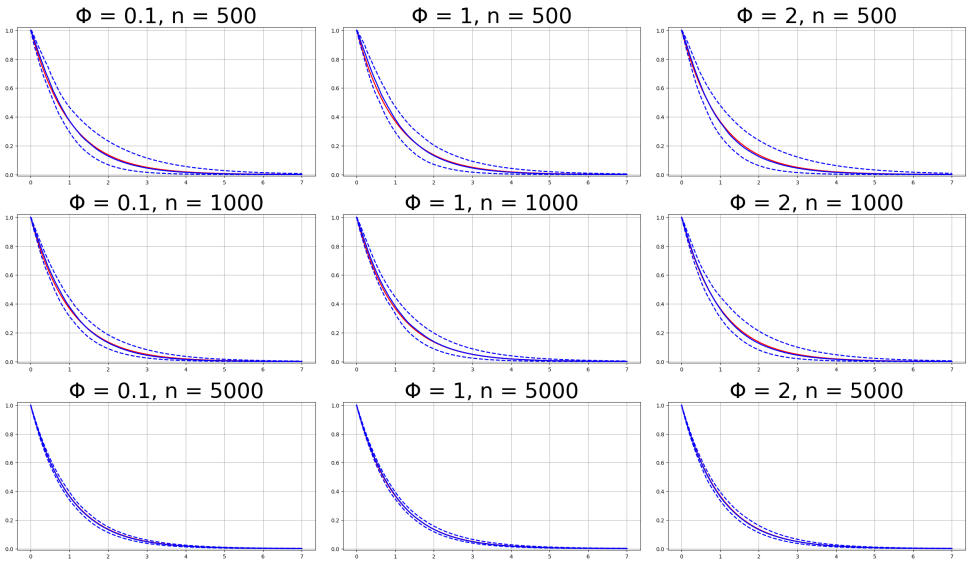
Model	ϕ	time	n_{train}	$\Delta\overline{S}_p(t)$	$\Delta\overline{S}_1(t)$	$\Delta\overline{\eta}(x)$	$\Delta\overline{p_0}$	$\Delta\overline{m}$
IPCM	0.1	25min 29s	500	0.029124	0.000988	0.826139	0.039508	0.711111
		26min 19s	1000	0.022288	0.000480	0.616062	0.031316	0.647315
		59min 57s	5000	0.016053	0.000356	0.437020	0.022225	0.587356
	1	21min 10s	500	0.018581	0.003541	0.625853	0.031741	0.959991
		20min 54s	1000	0.016349	0.004146	0.586346	0.028994	0.977712
		32min 30s	5000	0.008335	0.004666	0.295462	0.016766	0.874762
	2	17min 43s	500	0.020738	0.014814	1.234360	0.041123	1.736038
		18min 39s	1000	0.019442	0.015617	1.175383	0.039116	1.881228
		27min 32s	5000	0.018116	0.016917	1.109764	0.037153	1.839809
INBCM with fixed ϕ	0.1	26min 05s	500	0.032145	0.001381	0.930052	0.041554	0.788707
		36min 10s	1000	0.025263	0.000635	0.707812	0.033504	0.702467
		77min 17s	5000	0.017001	0.000083	0.451592	0.022220	0.603041
	1	21min 11s	500	0.019935	0.001490	0.610912	0.026581	1.076470
		30min 46s	1000	0.016274	0.000740	0.496323	0.020851	1.045644
		29min 32s	5000	0.004367	0.000086	0.102307	0.005509	0.837149
	2	20min 33s	500	0.015165	0.001373	0.668547	0.019151	1.734697
		26min 40s	1000	0.008824	0.000585	0.362249	0.010266	1.750102
		44min 30s	5000	0.002831	0.000092	0.106550	0.003175	1.550582

For datasets generated with a low dispersion value, $\phi = 0.1$, the estimation accuracy obtained by the integrated negative binomial model is slightly lower, but the difference is relatively small. For datasets generated with ϕ equal to 1 and 2, the model, which does not account for the overdispersion parameter, produces larger mean differences on $\Delta\overline{S}_p(t)$, $\Delta\overline{p_0}$, $\Delta\overline{\eta}(x)$ and $\Delta\overline{S}_1(t)$. This is evident in Figure 3, where the estimated $S_1(t)$ curves and their confidence intervals appear shifted to the left from the true curve, indicating that the model overestimates hazard rates. In contrast, the estimated $S_1(t)$ curves and their confidence intervals stay closer to the true curve when using the integrated negative binomial model, as demonstrated in Figure 4. As ϕ increases, the estimated mean absolute errors $\Delta\overline{m}$ also increase. However, both models exhibit similar accuracy in this measurement. Lastly, the AUC values obtained from the integrated negative binomial model consistently surpass those from the integrated promotion time cure model. This suggests that the proposed model is better in predictive capacity particularly as the level of overdispersion intensifies.



(a) $\phi = 0.1$ (b) $\phi = 1$ (c) $\phi = 2$

Fig. 3: Estimated curves of $S_1(t)$ from the IPCM in scenarios where M follows a negative binomial distribution with varying ϕ



(a) $\phi = 0.1$ (b) $\phi = 1$ (c) $\phi = 2$

Fig. 4: Estimated curves of $S_1(t)$ from the INBCM in scenarios where M follows a negative binomial distribution with varying ϕ

Table 6: AUC of both models applied to overdispersed data for three different sample sizes

		IPCM		INBCM with fixed ϕ	
ϕ	n_{train}	Train set	Test set	Train set	Test set
0.1	500	0.839108	0.823816	0.848120	0.833626
	1000	0.826987	0.819386	0.847051	0.838269
	5000	0.806683	0.804761	0.829856	0.827632
1	500	0.684517	0.662885	0.740800	0.714441
	1000	0.638764	0.629546	0.702033	0.689066
	5000	0.578341	0.576895	0.628346	0.625909
2	500	0.594320	0.580597	0.671367	0.650770
	1000	0.578684	0.570287	0.633805	0.623909
	5000	0.536130	0.535386	0.579396	0.578014

4 Application

In this section, we will implement the proposed integrated negative binomial model using the OASIS-3 dataset. OASIS-3 is an open resource available at www.oasis-brains.

org. The resource comprises a collection of MRI and clinical data spanning over 15 years, collected from various studies conducted at the University of Washington Alzheimer's Disease Research Center. The data used in this application were sourced from the supplementary material made available by Xie and Yu (2021b). The sample consists of 352 individuals with the maximum observed time of 12.2 years.

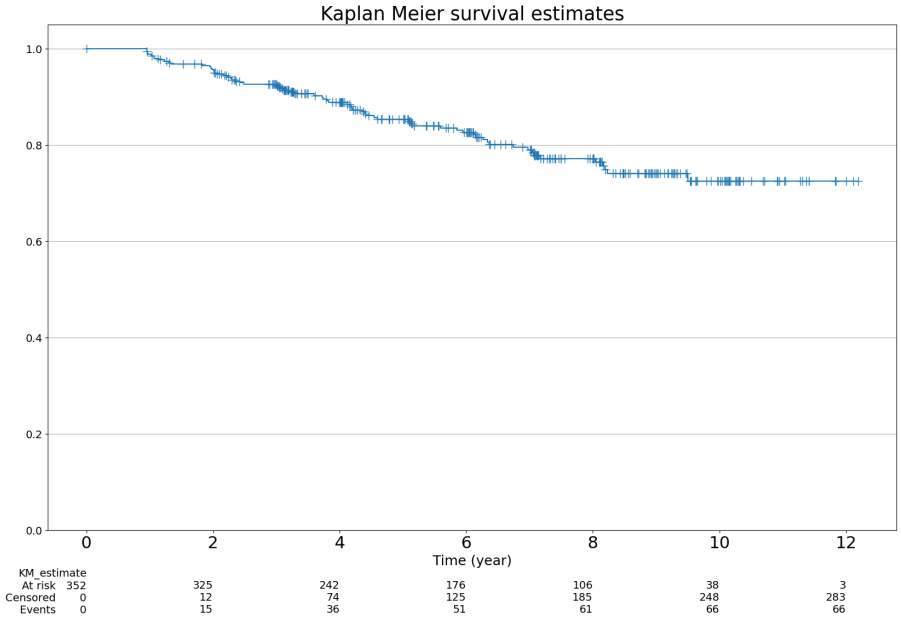


Fig. 5: Estimated Kaplan-Meier survival curve for the OASIS-3 data

The dependent variable is defined as the time until the onset of Alzheimer's disease from the date each individual is enrolled in the study. Event confirmation is determined by the Clinical Dementia Rating (CDR) scale, where an event is defined as CDR > 0. The application is conducted using Python within the Google Colab Pro environment, as described in the simulation study. The event proportion corresponds to approximately one-fifth of the samples. From the fitted Kaplan-Meier survival curve in Figure 5, there is strong evidence of the presence of a non-susceptible group, as the plateau begins at around 9 years. The graph was plotted using the `KaplanMeierFitter` function from the `lifelimes` package. Our interest lies in predicting the cure rate using MRI data, as described in Section 2. The MRIs are in grayscale with dimensions of 160×200, and the pixels are normalized to the range $[0, 1]$ before being fed into the network.

The division of the training and testing sets and the percentage of censoring are shown in Table 7. A moderate choice ($G = 5$) was adopted for the interval division of observed times and the number of events is approximately equal in each sub-interval, as presented in Table 8.

Table 7: Data set partition and censoring percentage

Data	Size	Observed events	Number of censored data	% Censoring
Training	280	52	228	81,43
Test	72	14	58	38,89
Total	352	66	286	81,25

Table 8: Partition of time

g	(0,953, 2,062]	(2,062, 3,360]	(3,360, 4,495]	(4,495, 6,320]	(6,320, 12,2]
Number of events	11	10	10	10	11

Table 9 presents the architecture of the applied CNN and the numbers of parameters are over 2 million units in total. The visualization of the architecture is shown in Figure 6.

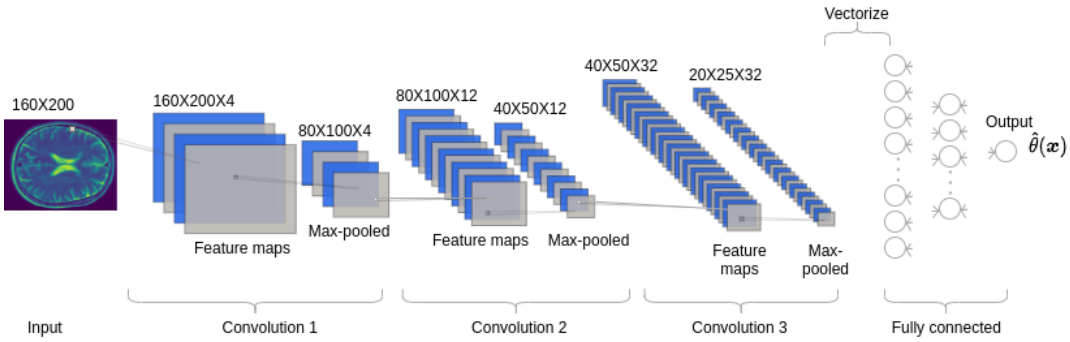


Fig. 6: The convolutional neural network architecture used to estimate θ

The estimation of the overdispersion parameter is based on the profile log-likelihood technique. By choosing a predefined range of fixed values for ϕ, we search for the ϕ' within this set that maximizes the observed log-likelihood function. Initially, we conducted a search within the set of $\{0.1, 0.2, \cdots, 1.4, 1.5\}$, which resulted in $\phi' = 0.1$. Subsequently, we performed a further second search within the set of $\{0.01, 0.02, \cdots, 0.1\}$, leading to the determination of $\phi' = 0.01$.

In Figure 7, the graph on the left illustrates the progression of the loss function presented in Table 1 across epochs during the estimation of $\hat{\theta}$ while the graph on the right displays the estimated AUC values at the conclusion of each iteration in EM algorithm. The blue and orange curves represent the training and test sets respectively. At the end of this iterative process, we attained an AUC of 0.858 for the training set and 0.814 for the test set.

Table 9: Architecture of the convolutional network applied to the MRIs in the OASIS-3 dataset

Layer	Dimension	Activation σ	Number of parameters
Convolution 1	(160, 200, 4)	ReLU	104
Pooling 1	(80, 100, 4)	max	-
Convolution 2	(80, 100, 12)	ReLU	1212
Pooling 2	(40, 50, 12)	max	-
Convolution 3	(40, 50, 32)	ReLU	9632
Pooling 3	(20, 25, 32)	max	-
Flatten	(16000)	-	-
Fully connected	(128)	tanh	2048128
Output	(1)	-	128
Total	-	-	2,059,204

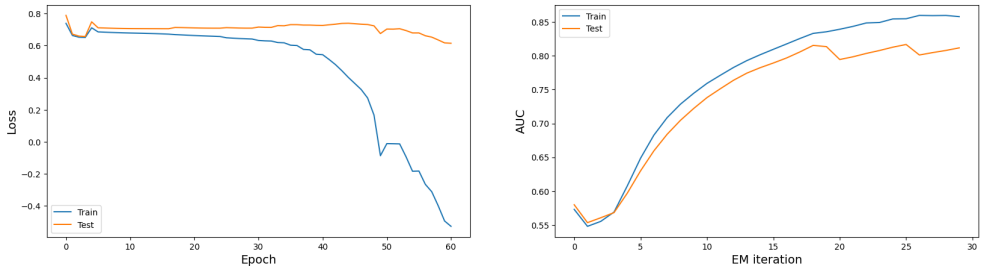


Fig. 7: (left) Loss function monitoring across epochs (right), AUC values increase with the progression of iterations

Alzheimer's disease is known to lead to the degeneration of brain cells and the impairment of mental functions, often affecting older people. Clinically, this neurodegeneration manifests as cerebral atrophy. A recent study conducted by Inglese et al. (2022) suggests that Alzheimer's disease may potentially be detectable and diagnosed through a single magnetic resonance imaging examination. Also, AlSaeed and Omar (2022) stated that CNN have increasingly been used in the diagnosis and prediction of Alzheimer's disease from MRI images. This technique automatically extracts accurate features to measure the brain's size and the number of its cells, thus detecting atrophy areas in the temporal and parietal lobes.

Figure 8 illustrates varying patterns in MRI scans at different levels of $\hat{p}_0$ and $\hat{\theta}(\mathbf{x}) = e^{\hat{\eta}(\mathbf{x})}$. Since these two quantities are inversely proportional, we can observe noticeable enlargement of the cerebral ventricles, which are the chambers within the brain containing cerebrospinal fluid, when $\hat{\theta}(\mathbf{x})$ increases or $\hat{p}_0$ decreases. Simultaneously, the sulci which are the grooves or furrows in the brain tend to deepen and widen.[1] These observations align with clinical evidence.

We anticipate that the proposed integrated negative binomial model will exhibit a good fit similar to the integrated promotion time cure model when ϕ approaches zero.

[1] https://www.brightfocus.org/alzheimers/infographic/brain-alzheimers-disease

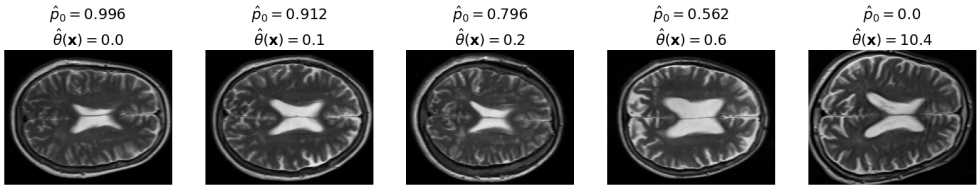


Fig. 8: MRIs (160×200 pixels) at different quantiles of $\hat{p}_0$ and $\hat{\theta}(\mathbf{x})$ from the test set, from left to right: 0%, 25%, 50%, 75% and 100%

To assess the model performances, we will conduct estimations based on 100 bootstrapped samples and compare the results of the two models. This comparison will involve an examination of the estimated survival curves $S_1(t)$ along with their corresponding pointwise 95% confidence intervals and the AUC values. Figure 9 illustrates that the two curves are nearly overlapping. Notably, the confidence intervals of the integrated negative binomial model, represented by the red dotted lines, are slightly narrower than those of the integrated promotion time cure model, depicted by the blue dotted lines. Also, the distributions of AUC of the models are almost identical as demonstrated in Figure 10. The numeric descriptive statistics of the calculated AUCs and the estimated $\boldsymbol{\alpha}$ are presented in Table 10.

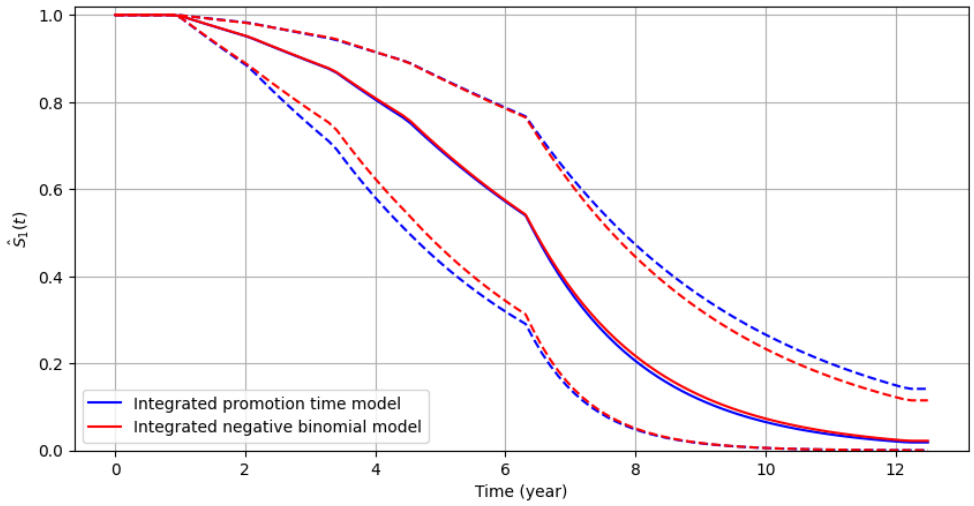


Fig. 9: Comparison of estimated $S_1(t)$ curves from the two models

Based on the aforementioned results, we can conclude that the proposed integrated negative binomial model produces satisfactory outcomes when applied to the OASIS-3 dataset. These results closely align with those presented in the referenced paper, where the model used is an integrated promotion time cure model.

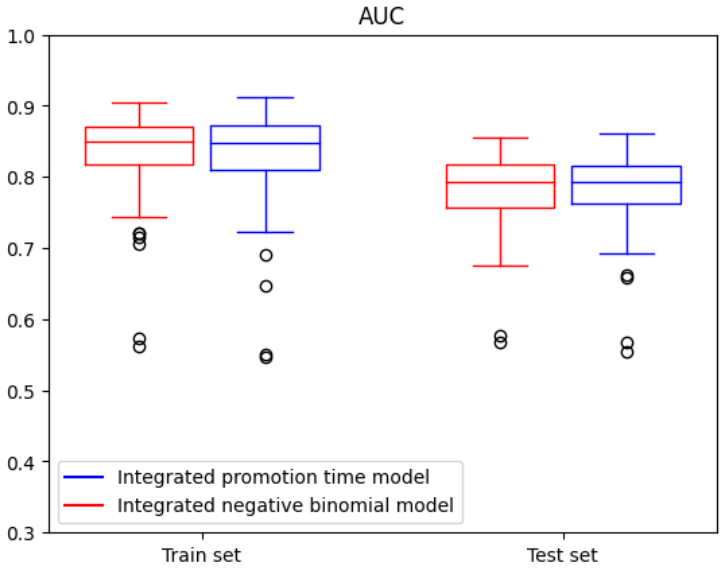


Fig. 10: Comparison of AUC values of the two models

Table 10: Descriptive summary of AUCs and estimated parameters of survival curves $S_1(t)$

	Integrated negative binomial model						
	$\text{AUC}_{\text{train}}$	AUC_{test}	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
mean	0.832066	0.781006	0.048659	0.070641	0.133353	0.181767	0.588516
std deviation	0.059323	0.049832	0.023333	0.029389	0.058779	0.057860	0.201634
1st quartile	0.709700	0.680361	0.017370	0.027592	0.052378	0.084563	0.321970
2nd quartile	0.850377	0.792656	0.046545	0.064515	0.118985	0.188655	0.542635
3rd quartile	0.895008	0.844128	0.111836	0.130714	0.277550	0.304864	1.079108
	Integrated promotion time cure model						
	$\text{AUC}_{\text{train}}$	AUC_{test}	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
mean	0.831481	0.783034	0.050816	0.073376	0.135361	0.184000	0.589287
std deviation	0.063529	0.049128	0.025328	0.037387	0.059180	0.059353	0.202318
1st quartile	0.667735	0.660349	0.016570	0.029967	0.049733	0.083310	0.287260
2nd quartile	0.848589	0.792627	0.047271	0.065014	0.123469	0.187220	0.572247
3rd quartile	0.898700	0.845525	0.115514	0.175137	0.293136	0.301436	1.065154

5 Discussion and conclusion

In this work, we explore the possibility of modeling long-term survival data with unobserved dispersion using unstructured predictors. Taking inspiration from the work of Xie and Yu (2021b), where medical images were utilized to model the effect in the promotion time cure model, we chose to employ the negative binomial distribution to account for the presence of overdispersion in the data. Since all cure models can be

represented as a two-stage model, we proposed the term "integrated two-stage cure rate model" to convey the notion of integrating CNN within the cure rate model. Consequently, our model can be referred to as the "integrated negative binomial cure rate model". The formulation of the proposed integrated two-stage cure rate model was presented in Section 2. To investigate the usefulness of our proposed model, we executed a simulation study, detailed in Section 3. The results affirm that our model is a better choice as it can accommodate both scenarios with and without overdispersion. Concluding our work, we applied the proposed model to the OASIS-3 dataset, and the obtained results closely align with those presented in the referenced paper.

Two principal aspects need to be considered and improved. First, in application chapter, we adopted the time-consuming profiling likelihood method for estimating the dispersion parameter. As the log likelihood function $L_1(\boldsymbol{w}; \boldsymbol{D}_{\text{comp}})$, which deals with the mean number of risk factors and dispersion in the EM algorithm, involves a significant number of parameters, it is desirable to find a more efficient method for estimating the dispersion parameter. Second, it is not feasible to use traditional statistical inference methods to calculate the estimated parameter's standard error.

On top of these, for the purpose of enhancing the clarity and interpretability of the application context, it is advisable to apply the proposed model in future studies on medical images associated with cancer diagnosis.

Statements and Declarations

No funding was received to assist with the preparation of this manuscript. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgments

The authors would like to acknowledge the supplementary resources provided by the authors of the referenced article (Xie and Yu, 2021b). These resources include the Python implementation code and the medical images dataset which have greatly eased the challenges in executing the simulation studies and application within this work.

Data availability

The image dataset employed in the Section 4 is the same used in the referenced article which can be accessed with details described in Appendix C in Xie and Yu (2021b). Additional information is available at www.oasis-brains.org.

Code Availability

All implementation works are available in https://github.com/TehLedRed/SpringerManuscript.

References

AlSaeed, D., Omar, S.F.: Brain mri analysis for alzheimerrsquo;s disease diagnosis using cnn-based feature extraction and machine learning. Sensors **22**(8) (2022) https://doi.org/10.3390/s22082911

Aselisewine, W., Pal, S.: On the integration of decision trees with mixture cure model. Statistics in Medicine **42**(23), 4111–4127 (2023) https://doi.org/10.1002/sim.9850 https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9850

Berkson, J., Gage, R.P.: Survival curve for cancer patients following treatment. Journal of the American Statistical Association **47**(259), 501–515 (1952) https://doi.org/10.1080/01621459.1952.10501187

Gallardo, D.I., *et al.*: A simplified estimation procedure based on the em algorithm for the power series cure rate model. Communications in Statistics - Simulation and Computation (2016) https://doi.org/10.1080/03610918.2016.1202276

Inglese, M., et al.: A predictive model using the mesoscopic architecture of the living brain to detect alzheimer's disease. Communications Medicine **2**(1) (2022)

Jiang, C., Wang, Z., Zhao, H.: A prediction-driven mixture cure model and its application in credit scoring. European Journal of Operational Research **277**(1), 20–31 (2019) https://doi.org/10.1016/j.ejor.2019.01.07

Lundervold, A.S., Lundervold, A.: An overview of deep learning in medical imaging focusing on mri. Zeitschrift für Medizinische Physik **29**(2), 102–127 (2019) https://doi.org/10.1016/j.zemedi.2018.11.002 . Special Issue: Deep Learning in Medical Physics

Li, P., Peng, Y., Jiang, P., Dong, Q.: A support vector machine based semiparametric mixture cure model. Computational Statistics **35**(3), 931–945 (2020) https://doi.org/10.1007/s00180-019-00931-

Meyer, P., Noblet, V., Mazzara, C., Lallement, A.: Survey on deep learning for radiotherapy. Computers in Biology and Medicine **98**, 126–146 (2018) https://doi.org/10.1016/j.compbiomed.2018.05.018

Pal, S., Aselisewine, W.: A semiparametric promotion time cure model with support vector machine. The Annals of Applied Statistics **17**(3), 2680–2699 (2023) https://doi.org/10.1214/23-AOAS1741

Pal, S., Peng, Y., Aselisewine, W.: A new approach to modeling the cure rate in the presence of interval censored data. Computational Statistics **39**(5), 2743–2769 (2024) https://doi.org/10.1007/s00180-023-01389-7

Pal, S., Peng, Y., Aselisewine, W., Barui, S.: A support vector machine-based cure rate

model for interval censored data. Statistical Methods in Medical Research **32**(12), 2405–2422 (2023) https://doi.org/10.1177/09622802231210917 . Epub 2023 Nov 8

Rodrigues, J., Cancho, V.G., de Castro, M., Louzada-Neto, F.: On the unification of long-term survival models. Statistics Probability Letters **79**(6), 753–759 (2009) https://doi.org/10.1016/j.spl.2008.10.029

Tsodikov, A.D., Ibrahim, J.G., Yakovlev, A.Y.: Estimating cure rates from survival data. Journal of the American Statistical Association **98**(464), 1063–1078 (2003) https://doi.org/10.1198/016214503000000001007 . PMID: 21151838

Xie, Y., Yu, Z.: Mixture cure rate models with neural network estimated nonparametric components. Comput. Stat. **36**(4), 2467–2489 (2021) https://doi.org/10.1007/s00180-021-01086-3

Xie, Y., Yu, Z.: Promotion time cure rate model with a neural network estimated nonparametric component. Statistics in Medicine **40**(15), 3516–3532 (2021) https://doi.org/10.1002/sim.8980