

Stochastic EM Algorithm-based Likelihood
Inference for Spatial Cure Rate Models Based on
Some Flexible Distributions

STOCHASTIC EM ALGORITHM-BASED LIKELIHOOD
INFERENCE FOR SPATIAL CURE RATE MODELS BASED ON
SOME FLEXIBLE DISTRIBUTIONS

BY
XINYI WANG

A THESIS
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS AND STATISTICS
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

© Copyright by Xinyi Wang, March 2025

All Rights Reserved

Doctor of Philosophy (2025)
(Mathematics and Statistics)

McMaster University
Hamilton, Ontario, Canada

TITLE: Stochastic EM Algorithm-based Likelihood Inference for
Spatial Cure Rate Models Based on Some Flexible Dis-
tributions

AUTHOR: Xinyi Wang
B.Sc.(McMaster University)
M.Sc.(McMaster University)

SUPERVISOR: Prof. Narayanaswamy Balakrishnan

CO-SUPERVISOR: Prof. Wei Xu

NUMBER OF PAGES: xv, 172

To Yi

To my parents, Hong & Hao

Abstract

With geographic information systems (GIS) software having become readily accessible, a natural interest arises in modeling spatial-referenced data. Spatial survival analysis is one of the common approaches to account for spatial dependence in the data, and it refers to the modelling and analysis of location-referenced time-to-event data. One of the popular methods for conducting survival analysis for lifetime data is to use cure rate models, which have been studied extensively assuming a competing risks scenario. In this work, we adopt the competing risks scenario for modelling lifetimes by assuming the number of competing causes related to the occurrence of an event to follow a discrete power series (PS) distribution as proposed in Noack (1950). The PS cure rate model is extremely flexible and can be transformed into many well-known cure models through choices of its power parameter and series function. In this thesis, we restrict our attention to the three scenarios, called the first activation scheme, the random activation scheme, and the last activation scheme. Under the first activation scheme, we focus on competing causes following Poisson, geometric, and logarithmic distributions, and the corresponding cure rate models are the promotion time cure rate model, geometric cure rate model, and logarithmic cure rate model. For the last activation scheme, the three distributions for the competing causes are also the same. But, the corresponding cure models are the complementary

promotion time cure rate model, the complementary geometric cure rate model, and complementary logarithmic cure rate model. The random activation is defined when competing risk follows a Bernoulli distribution and forms a Bernoulli cure rate model.

The spatial effect model is constructed through a Gaussian process, also known as a Gaussian random field (Li and Ryan (2002); Wilson and Wakefield (2020)). The spatial effect is then added to cure rate models as spatial frailties to reveal the effect of geographical location on survival time for susceptible individuals within the region.

In addition, we propose the baseline to follow the generalized extreme value (GEV) distribution (Coles (2001) and Kotz *et al.* (2001)). By doing so, we empower the model with a more flexible baseline hazard function that combines a family of continuous distributions into one. By adjusting the scale and shape parameters, GEV includes to the Gumbel (Type I), Fréchet (Type II), and Weibull (Type III) distributions as special cases. Widely used hazard distribution like Weibull assumes the distribution to be monotone, while in fact the hazard function can have a bell shape, U-shape, or combine both of them. The GEV distribution accommodates tail behaviour and is more inclusive for modelling survival data (Li *et al.* (2016)). Additionally, the GEV simplifies the implementation by allowing the data to determine the most suitable type of tail behaviour through inference on this shape parameter. It is not required to have prior knowledge of which extreme value family is appropriate to adopt. While the GEV distribution has gained popularity in various disciplines, its application under the survival model set up and in survival analysis is relatively unexplored.

Despite the fact that Bayesian inference and the EM algorithm have been carried out for conducting survival analysis in various papers, the inference of the spatial survival model based on Stochastic EM has not been studied much. In this work,

the required methodology pipelines are developed for applying the Stochastic EM algorithm (Celeux and Diebolt (1985)) to find the optimal estimates of parameters for the proposed models. In many cases, the stochastic step (S-step) of the stochastic EM is preferred to the expectation step (E-step) of the EM algorithm due to the fact that the S-step is based on a single draw from the conditional distribution, and it avoids the necessity of computing the conditional expectations involved in the EM algorithm. Unlike the Monte Carlo EM algorithm and Newton-Raphson method, SEM is of a stochastic nature and is free of the saddle point problem. Besides, it is insensitive to the starting values and performs well for small and moderate sample sizes, which are quite common in clinical data. In the following Chapters, the improvements in the robustness of the SEM algorithm are illustrated. Moreover, extensive simulation studies demonstrate that the Stochastic EM algorithm converges well under various settings, and the performance is also assessed through model discrimination using information-based criteria. The selection rates for choosing the true model among all candidate models are examined. The spatial effect of a particular geographical location on the survival times of susceptible individuals is shown. The proposed spatial cure rate models and associated methodologies are then applied to a smoking cessation dataset. The spatial effect on hazard is visualized using maps. The cure rate and survival probabilities are compared and contrasted with and without including spatial effects. The necessity of adding spatial effects to the survivals is illustrated through the likelihood ratio test (LRT).

KEY WORDS: Stochastic EM; Spatial survival analysis; Competing cause scenario; Power series cure rate model; Generalized extreme value distribution; Likelihood ratio test; Model discrimination; Hypothesis test; Goodness-of-fit test.

Acknowledgements

I would like to express my heartfelt gratitude to my dear doctoral supervisor, Prof. Narayanaswamy Balakrishnan, who not only introduced me to the field of survival analysis, but also inspired me in so many areas of life. Without his invaluable guidance, constant support, patience, and care, I would not be the person I am today. I will always cherish the memories of attending conferences with him, having discussions in his office, and being inspired by the researchers he introduced me to. Working under his supervision these past few years has been an incredible experience, and I'm certain the bond we've formed will last a lifetime.

I would like to thank my co-supervisor, Prof. Wei Xu, who is always available to answer my questions and has inspired me with research suggestions that are applicable to real-life situations. His kindness, deep knowledge, and passion for statistics have greatly influenced me and will continue to shape my career.

My thanks also go to my doctoral committee member, Prof. Shui Feng, for years of support, valuable suggestions, constructive criticisms, and for being there for me since I joined the graduate program in 2018.

I am grateful to my parents for their unwavering support throughout my years of studies. The infinite love and patience they have given me have shaped me into the easygoing and optimistic person I am today.

I would also like to extend my thanks to Yi Liu, who brought laughter into my life. Your kindness, guidance, love, and support have made me stronger. I truly appreciate the time you spent helping me with the theoretical and computational challenges I faced in my thesis, as well as everything you've done for me and my family.

Lastly, I wish to express my appreciation to Kai Liu, Xiaojun Zhu, Chenxi Yu, Angela Wang, and all my research friends for their kind assistance, inspiration, and moral support. I am also deeply grateful to McMaster University for awarding me the scholarship and to the Department of Mathematics and Statistics for providing me with financial support during my graduate study.

Contents

Abstract	iv
Acknowledgements	vii
1 Introduction	5
1.1 Common survival models	9
1.1.1 Power series (PS) model	10
1.1.2 PS model with spatial frailties (Scenario 1)	12
1.1.3 PS model with spatial frailties (Scenario 2)	14
1.2 A brief literature review	15
1.3 Distributions for competing causes	15
1.4 Gaussian spatial effect	16
1.5 Baseline Distribution: GEV distribution	20
1.6 Data structure and the likelihood	20
1.6.1 Non-informative censoring and likelihood	21
1.6.2 Non-informative censoring and likelihood for spatially correlated data	22
1.7 Likelihood inference	24

1.7.1	Type I: SEM	25
1.7.2	Type II: SEM algorithm for cure rate models with spatial effect	27
1.7.3	Standard errors and asymptotic confidence intervals based on MLEs	30
1.8	Simulation study	31
1.8.1	General setup	31
1.8.2	Cure rate models with spatial survival analysis	31
1.9	Model discrimination	33
1.9.1	Information-Based criterion	33
1.9.2	Likelihood ratio test	35
1.9.3	Parameter evolution of SEM	36
1.10	Smoking cessation dataset	37
1.11	Scope of the thesis	38

2 Stochastic EM-based Likelihood Inference for a Class of Cure Rate

	Models Based on GEV Distribution	41
2.1	Introduction	41
2.2	Cure rate models	42
2.3	The expression of log-likelihood function for cure rate models	46
2.3.1	Bernoulli mixture cure rate model (Model 1)	46
2.3.2	Poisson cure rate model (Model 2)	48
2.3.3	Geometric cure rate model (Model 3)	48
2.3.4	Logarithmic cure rate model (Model 4)	49
2.4	Stochastic Expectation-Maximization (SEM) algorithm for cure rate models	50

2.5	Simulation study	55
2.6	Analysis of smoking cessation data	64
3	Stochastic EM-based Likelihood Inference for First Activation Scheme of PS Cure Rate Model with Gaussian Spatial Frailties	72
3.1	Introduction	72
3.2	Cure models	74
3.3	The log-likelihood function	78
3.3.1	Bernoulli mixture cure rate with spatial frailties (Model 5) . .	78
3.3.2	Promotion time cure rate Poisson model with spatial frailties (Model 6)	81
3.3.3	Geometric cure rate model with spatial frailties (Model 7) . .	83
3.3.4	Logarithmic cure rate model with spatial frailties (Model 8) .	83
3.4	Stochastic EM	85
3.5	Simulation Study for Spatially Correlated Data	94
3.5.1	Model discrimination	96
3.6	Analysis of smoking cessation data	104
4	Stochastic EM-based Likelihood Inference for Last Activation Scheme of PS Cure Rate Model with Gaussian Spatial Frailties	111
4.1	Introduction	111
4.2	Cure models	113
4.3	The likelihood function	116
4.3.1	Model 9	116
4.3.2	Model 10	117

4.3.3	Model 11	118
4.4	Stochastic EM	119
4.5	Simulation study	124
4.6	Analysis of smoking cessation data	134
5	Concluding Remarks	141
5.1	Summary of research	141
5.2	Future work	143
A	Appendix A for Chapter 2	145
B	Appendix B for Chapter 3	153
C	Appendix C for Chapter 4	161

List of Figures

1.1	Map of 5 zip codes in Minnesota	32
1.2	Map of 51 zip codes in Minnesota	38
2.3	Maps of cure rate and survival probabilities stratified by zip codes. . .	66
2.4	Survival Plots (C.P.): Female vs Male w/ SI	69
2.5	Treatment: Female, Consump:6, Duration: 30	70
2.6	Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm when M_{ij} follow Model 2 for the smoking cessation dataset. (2500 iterations after burn-in period of 500)	71
3.7	Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota	106
3.8	Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota	106
3.9	Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota	106
3.10	QQ plot	109
3.11	Surviving function: stratified by location (Cannon Falls, Minnesota; Stewartville, Minnesota) and gender with Duration (30 years-mean) and cigarette consumption: 6 and 31.	109

3.12	Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period) . .	110
4.13	Demonstration of different spatial frailties for subjects from 51 zip codes in Minnesota	136
4.14	Demonstration of different spatial frailties for subjects from 51 zip codes in Minnesota	136
4.15	Cure rate and survival probabilities when considering spatial frailties for subjects from 51 zip codes in Minnesota	136
4.16	Survival plot of Model 9 stratified by treatments and gender with Duration: 30 years, cigarette consumption: 6/day and 31/day (Mean), and spatial frailty: Rochester, MN (Zip: 55066).	139
4.17	Survival plot of Model 9 for a male smoker with smoking habit of 33 yrs and consumption= 20 cigarettes/day	140
A.18	Parameter evolution plots of the SEM algorithm when M_{ij} follow Model 2 for the smoking cessation dataset. (2500 iterations after burn-in period of 500)	150
B.19	Parameter evolution plots for b_4 , β_1 and β_2 of the SEM algorithm, when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)	158
B.20	Parameter evolution plots for β_3 , β_4 and μ of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period) . .	159
B.21	Parameter evolution plots for σ , γ of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)	160

C.22	Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period)	165
C.23	Parameter evolution plots for b_4 , β_1 and β_2 of the SEM algorithm, under the scenario when all competing causes are present and follow the complementary promotion time cure rate model (Model 9). (2500 iterations after the 500 burn-in period)	166
C.24	Parameter evolution plots for β_3 , β_4 and μ of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period)	167
C.25	Parameter evolution plots for σ , γ of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period) .	168

List of Tables

1.1	A list of distributions for the competing causes are assumed to follow in the analysis	16
1.2	Latitude, longitude, zip code, and city names of the 5 regions in Minnesota, US, used in the simulation study.	33
1.3	The Euclidean distance matrix of the 5 cities and their zip codes in Minnesota, US.	33
1.4	Variables in the smoking cessation dataset in southeastern corner of Minnesota	39
2.5	The settings of the shape parameter, γ , considered for the baseline function	56
2.6	Some of the cure rates and levels of censoring for the simulated datasets	57
2.7	The choice of true values for the model parameters	57
2.8	Simulation results of estimated means, Bias, and RMSE, for different choices of baseline shape parameter, γ , and true model and fitted model are the same, based on 1200 iterations.	59

2.9	Simulation results of mean estimates, bias, and root mean square error (RMSE) based on 1200 iterations. ($n = 1000$, censored proportion = (Case 1: 0.68, Case 2: 0.73), moderate and high cure rate = (0.611, 0.711))	63
2.10	Cure Probabilities: $n = 1000$	64
2.11	Estimated mean, standard error, and 95% CI when competing causes follow Bernoulli and Poisson (Model 1 and Model 2) for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3). . . .	67
2.12	Estimated mean, standard error, and 95% CI when M_{ij} follow geometric and logarithmic models (Model 3 and Model 4) for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3). . . .	68
2.13	Negative log-likelihood values, AIC, BIC, AICc for Model 1, Model 2, Model 3, and Model 4 for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).	69
3.14	Some of the true values selected for the model	95
3.15	The setup of the spatial correlation, ϕ_s , considered for the simulated dataset	95
3.16	The settings of the shape parameter, γ , considered for the baseline function	96
3.17	Simulation results on mean estimates, bias, and root mean square error (RMSE) under different choices of baseline, $\gamma = (-1.3, 0, 1.6, 2.5)$, and different spatial dependencies $\phi_s = (0.5, 0.8, 1, 1.1, 1.2)$, based on 1200 iterations	98

3.18	Comparison of simulation results on mean estimates, bias, and root mean square error (RMSE) for different fitted models based on 1200 iterations.	102
3.19	Selection rates based on AIC ($n = 500$)	103
3.20	Mean estimates, standard error, and 95% CI for model parameters for the smoking cessation data assuming M_{ij} follow Models 5 and 6 (Iterations: 3000, Burn-in: 500, and Spacing: 3).	107
3.21	Estimates of mean, standard error, and 95% CI for model parameters for the smoking cessation data assuming M_{ij} follow Models 7 and 8 (Iterations: 3000, Burn-in: 500, and Spacing: 3).	108
3.22	Negative log-likelihood, AIC, BIC, and AICc values for the smoking cessation data assuming M_{ij} follow Models 5 - 8 (Iterations: 3000, Burn-in: 500, and Spacing: 3).	109
4.23	Examples of the cure rates and the levels of censoring for the simulated datasets	125
4.24	The settings of the shape parameter, γ , considered for the baseline function	125
4.25	The settings of different spatial correlation, ϕ_s , considered for the simulated dataset	126
4.26	The different settings of the true values of the model parameters . . .	126
4.27	The estimated means, bias, and RMSE for selected models, with different choices of shape parameter, $\gamma = (2, 2.1, 2.7)$, and spatial dependency, $\phi_s = (0.5, 1.5, 2)$, when the true model of M_{ij} and the fitted model are the same.	129

4.28	The estimated means, bias, and RMSE for selected models, with different choices of shape parameter, $\gamma = (-1.25, 0, 2.5)$, and spatial dependency, $\phi_s = (0.8, 1, 1.2)$, assuming the true models of M_{ij} and the fitted models are the same.	130
4.29	The selection rates for Models 9 - 11	133
4.30	Estimated mean, standard error, and 95% CI for the case when competing cause follow Bernoulli mixture model (Model 5) and complementary promotion time cure Poisson model (Model 9) with spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).	137
4.31	Estimated mean, standard error, and 95% CI for the cases when M_{ij} follow the complementary geometric cure rate model (Model 10) and complementary logarithmic cure rate model (Model 11) with spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).	138
4.32	Negative log-likelihood, AIC, BIC, and AICc, values for the distribution of the competing causes, M_{ij} , following the selected models under the consideration of all possible underlying competing causes and spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).	139

Chapter 1

Introduction

With geographic information systems (GIS) software becoming readily accessible, a natural interest arises in modeling spatial-referenced data. Disease mapping and spatial survival analysis are two common approaches to account for spatial dependence in the data. Disease maps are often used to spotlight areas with high or low incidence, mortality rates of disease and changes in the rates among regions. Spatial survival analysis refers to the modelling and analysis of location-referenced time to event data. The event of interest has either occurred (death or disease) or been censored, whichever occurs first. Spatial survival analysis is often used to model clustered survival data, and the clusters arise according to the geographic regions.

Spatial survival analysis can be further broken down into two components, which are survival analysis and the spatial effect on the survival. One of the popular methods for conducting the survival analysis for lifetime data is by the use of cure rate models. Cure rate models are also called survival models with a cure fraction. The subjects who are non-susceptible to the event of interest or are long-term disease-free survivors are considered cured, and the percentage of those cured patients in the lifetime data

is called the cured fraction. Yakovlev *et al.* (1996) discussed a common situation where multiple causes of failure create problems in the analysis of lifetime data, and the competing event precludes the event of interest from happening. In such a case, a model with competing risks becomes necessary. One of the ways to do it is to construct a competing-risks model based on unobserved latent time to each type of failure. Cure rate models have been studied extensively assuming a competing risks scenario.

The two most widely used cure rate models are the Bernoulli cure rate model and the promotion time cure rate model. In the Bernoulli cure rate model and the promotion time cure rate model, the number of competing risks follows a Bernoulli distribution and a Poisson distribution, respectively. In this work, the number of competing causes is modeled using the power series cure rate (PS) model (Noack (1950)). We restrict the scenario to three broad aspects, which are first activation, random activation, and last activation. In the first activation, we focus on competing causes following Poisson, geometric, and logarithmic distributions, and construct the promotion time cure rate model, the geometric cure rate model, and the logarithmic cure rate model, respectively. In the last activation, we assume competing causes following the same distributions as in the first activation scenario. However, the cure rate models get modified into the complementary promotion time cure rate model, the complementary geometric cure rate model, and the complementary logarithmic cure rate model, respectively. The random activation is defined as the case in which the competing risk follows a Bernoulli distribution, leading to a Bernoulli cure rate model.

In the literature, many papers have modeled the cluster-specific random effect in

the clustered data by using frailties (Vaupel *et al.* (1979)). This is due to the natural characteristics of clustered data, where the observations within the same region or cluster share similar conditions and environment, and these features sometimes are not easy to observe directly. Using frailties, the similarity and heterogeneity in the data can be captured within the same region. To achieve this goal of capturing spatial effect in the data, as shown in Section 1.20, the spatial effects are constructed through a Gaussian process, also known as a Gaussian random field (Li and Ryan (2002); Wilson and Wakefield (2020)). The spatial effect then gets added to the cure rate models as spatial frailties to reveal the effect of particular geographic location on survival times of susceptible individuals.

In existing research on survival analysis, numerous papers have modeled the data using the Weibull distribution, since the Weibull has monotone hazard rates, depending on the shape parameter. When we assume the hazard function to follow a Weibull distribution, we assume the hazard function to be monotonically decreasing, increasing, or staying constant. In reality, the hazard function is not necessarily guaranteed to be monotone; it could be bell-shaped, U-shaped, or a combination of both. Coles (2001) and Kotz *et al.* (2001) have discussed a flexible model called the generalized extreme value (GEV) distribution. By assuming the baseline to be the generalized extreme value (GEV) distribution, we empower our model with a more flexible baseline hazard function that combines a family of continuous distributions into one. The GEV is quite flexible. By adjusting the scale parameter, σ , and shape parameter, γ , GEV can change between Gumbel (Type I), Fréchet (Type II), and Weibull (Type III) distributions. Hence, the GEV is more appropriate, and we adopt it here for constructing our survival models. In addition, the GEV simplifies the implementation

by allowing the data to determine the most suitable type of tail behaviour through inference on its shape parameter, γ . Thus, we are not required to have prior knowledge of which extreme value family is appropriate to use. The GEV distribution has gained popularity in various disciplines, but its application under the survival model setup and in survival analysis is relatively new (Li *et al.* (2016)).

Despite the fact that Bayesian inference and the EM algorithm have been carried out to conduct survival analysis in many articles, the inference of the spatial survival model based on Stochastic EM has not been studied much. In this work, we develop the required steps of inference based on the Stochastic EM algorithm (Celeux and Diebolt (1985)) for our proposed models. In many scenarios, the stochastic step (S-step) in the stochastic EM is preferred over the expectation step (E-step) of the EM algorithm due to the fact that the S-step is based on a single draw from the conditional distribution, and it avoids the necessity for computing complex conditional expectations involved in the EM algorithm. Unlike the Monte Carlo EM algorithm or Newton-Raphson method, which do not guarantee convergence to a global maximum or a local maximum (since they may lead to convergence to a stationary point close to the starting value, and that stationary point might be a saddle point), SEM is of a stochastic nature and is free of the saddle point problem. Moreover, SEM is insensitive to starting values and performs well for small and moderate sample sizes, which are commonly seen in clinical data. The improvements in the robustness of the convergence algorithm have been well illustrated in the literature. (Bordes (2007); Davies (2020)).

Along with model development, the extensive simulation studies (Section 2.5; Section 3.5; Section 4.5) carried out demonstrate that the algorithm converges well under

various settings, and the model performance is also assessed via model discrimination using information-based criteria (i.e., AIC, BIC, AICc). The selection rates for choosing the true model among all candidate models are checked and summarized in Tables 3.19 and 4.29. For illustration, the proposed models and methodologies are illustrated with a smoking cessation dataset on the relapse of smoking. The spatial effect on hazard has been demonstrated and also visualized via heat maps. The cure rate and survival probabilities are compared and contrasted with and without the spatial effect. The necessity of adding spatial effects to the survival models is confirmed through hypothesis testing using the likelihood ratio test (LRT).

1.1 Common survival models

In survival analysis, we are often interested in studying the differences between subjects' lifetime that are present due to different factors. One of the popular survival models is the proportional hazards model proposed by Cox (1972). Let T_i be the event time and $\mathbf{z}_i$ be the covariates associated with the i th subject. Then, the proportional hazards model is defined in terms of hazard function and is expressed as

$$\lambda(t_i|\mathbf{z}_i) = \lambda_0(t_i)\exp(\mathbf{z}_i^T\boldsymbol{\beta}), \quad (1.1)$$

where λ_0 is the baseline hazard function, and $\boldsymbol{\beta}$ is the effect of covariates associated with survival function. The proportional hazards model assumes that the hazard ratio is invariant over time, and it can be developed either parametrically or semi-parametrically. A parametric proportional hazards model is developed in this thesis where the choice of baseline hazard function λ_0 is predetermined (Section 1.5). The

survival function under the proportional hazards model is given by

$$S(t_i|\mathbf{z}_i) = S_0(t_i|\mathbf{z}_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}, \quad (1.2)$$

where S_0 is the baseline survival function corresponding to λ_0 .

1.1.1 Power series (PS) model

Suppose there are n individuals, indexed by $i = 1, \dots, n$, and the i th individual is exposed to competing risk M_i , where M_i denotes the initial number of competing causes related to the occurrence of an event. For M_i , we use the probability mass function (pmf) of the discrete power series distribution (Noack (1950)) given by

$$P(M_i = m) = \frac{a_m \theta_i^m}{\sum_{m=0}^{\infty} a_m \theta_i^m} = \frac{a_m \theta_i^m}{G(\theta_i)}, \quad (1.3)$$

where $m = 0, 1, 2, \dots$, $G(\theta_i) = \sum_{m=0}^{\infty} a_m \theta_i^m$, $a_m > 0$, and $\theta_i \in (0, s)$ is selected such that $G(\theta_i)$ is finite, and its first, second, and third derivatives are all well defined.

Given $M_i = m$, let the random variable W_{k^*i} denote the event time due to the k^* th latent risk for the i th individual, with distribution function $F(\cdot) = 1 - S(\cdot)$ and survival function $S(\cdot)$, where $k^* = 1, 2, \dots, m$. Any survival distribution $S(\cdot)$ may be used to represent the uncertainty in W_{k^*i} , as discussed in later sections. Although the number of competing causes, M_i , and the latent event time associated with a specific cause, W_{k^*i} , are unobservable, they can be conceptually constructed as follows.

In Chapter 2, we focus on the case when the event of interest only takes place after first possible latent cause has presented (first activation scheme). Under this assumption, the time to the event of interest for the i th individual is $T_i = \min\{W_{1i}, \dots, W_{M_i}\}$

for $M_i \geq 1$, and $T_i = \infty$ if $M_i = 0$ with $P[T_i = \infty | M_i = 0] = 1$.

The general form of the population survival function is $S_p^F(t_i) = P(T_i > t_i) = \sum_{m=0}^{\infty} S(t_i)^m \frac{a_m \theta_i^m}{G(\theta_i)}$, where $t_i > 0$, θ_i is the power function. The cure fraction is $p_{0i} = G(0)/G(\theta_i) > 0$, the density function is $f_p^F(t_i) = G'(\theta_i S(t_i)) \theta_i f(t_i)/G(\theta_i)$, and the hazard function is $\lambda_p^F(t_i) = G'(\theta_i S(t_i)) \theta_i f(t_i)/G(\theta_i S(t_i))$. Here, $G'(\theta_i S(t_i)) = dG(\theta_i S(t_i))/dt$ and $f(t_i) = -dS(t_i)/dt$, with $f(t_i)$ being the proper density function to event of interest T_i .

In addition, given M_i , the event of interest would occur when some of the possible causes are activated. Then, the number of initiated causes is a random variable with discrete uniform distribution on $\{1, \dots, M_i\}$ (Noack (1950); Bao *et al.* (2020)). This scenario is defined as random activation scheme. Its population survival function is given by $S_p^R(t_i) = P(T_i > t_i) = G(0)/G(\theta_i) + (G(\theta_i) - G(0))S(t_i)/G(\theta_i)$ for $t_i > 0$. The population density and hazard function for the random activation scheme are given by $f_p^R(t_i) = (1 - G(0)/G(\theta_i))f(t_i)$, and $\lambda_p^R(t_i) = \frac{(G(\theta_i) - G(0))f(t_i)}{G(0) + (G(\theta_i) - G(0))S(t_i)}$, respectively. The relation between the first activation scheme and random activation scheme is that $S_p^F(t_i) \leq S_p^R(t_i)$, for all $t_i > 0$.

In Chapter 2, we restrict our attention to four special cases. By adjusting the power function and series function, the competing causes are set to follow Bernoulli, Poisson, geometric, and logarithmic distributions. The related cure models become Bernoulli mixture cure rate model, Poisson (promotion time cure rate model), geometric (odds cure rate model), and logarithmic cure rate model.

The detailed forms of density functions, survival functions and form of cure rates under first or random activation are all discussed in Section 2.2. Hereafter in Chapter 2, we omit the superscripts F and R and write $f_p(t_i)$, $S_p(t_i)$, and $\lambda_p(t_i)$ for simplicity.

Although under the general form, $f_p(t_i)$ and $S_p(t_i)$, are not proper density function and not a proper survival function, it is not an issue for developing inference for the parameters.

1.1.2 PS model with spatial frailties (Scenario 1)

In Chapter 3, we discuss the spatial effect on the cure rate and survival probabilities.

First, we assume in the i th region ($i = 1, \dots, I$), there are n_i number of subjects indexed by $j = 1, \dots, n_i$. The observed event time for the (i, j) th subject is denoted by T_{ij} .

The initial number of competing causes of the event, M_{ij} , for (i, j) th subject is assumed to follow a discrete PS distribution, with probability mass function (p.m.f)

$$P(M_{ij} = m) = \frac{a_m \theta_{ij}^m}{\sum_{m=0}^{\infty} a_m \theta_{ij}^m} = \frac{a_m \theta_{ij}^m}{G(\theta_{ij})}, \quad (1.4)$$

where $m = 0, 1, 2, \dots$, $G(\theta_{ij}) = \sum_{m=0}^{\infty} a_m \theta_{ij}^m$, $a_m > 0$, and $\theta_{ij} \in (0, 1)$. The power series function $G(\theta_{ij})$ changes under different cases and form well-known distributions, like Bernoulli, Poisson, geometric, and logarithmic distributions.

Given $M_{ij} = m$, let random variable W_{k^*ij} be the failure time of the (i, j) th individual due to the k^* th latent risk, with distribution function $F(\cdot) = 1 - S(\cdot)$, $k^* = 1, 2, \dots, m$. Again, M_{ij} and W_{k^*ij} are unobservable. Hence, we proceed as follows.

Under the first activation scheme, assuming that any competing cause will eventually trigger the event, the failure time of the ij th individual is defined by the random variable $T_{ij} = \min\{W_{1ij}, \dots, W_{m_{ij}ij}\}$ for $M_{ij} > 0$, and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P[T_{ij} = \infty \mid M_{ij} = 0] = 1$.

The general form of its survival function for the population under the first activation scheme is given by $S_p^F(t_{ij}) = P(T_{ij} > t_{ij}) = \sum_{m=0}^{\infty} S(t_{ij})^m \frac{a_m \theta_{ij}^m}{G(\theta_{ij})}$, where $t_{ij} > 0$, and $G(\theta_{ij})$ is the power function. The cure fraction is $p_{0ij} = G(0)/G(\theta_{ij}) > 0$, the density function is $f_p^F(t_{ij}) = G'(\theta_{ij}S(t_{ij}))\theta_{ij}f(t_{ij})/G(\theta_{ij})$, and the hazard function is $\lambda_p^F(t_{ij}) = G'(\theta_{ij}S(t_{ij}))\theta_{ij}f(t_{ij})/G(\theta_{ij}S(t_{ij}))$. Here, $G'(\theta_{ij}S(t_{ij})) = dG(\theta_{ij}S(t_{ij}))/dt$ with $f(t_{ij}) = -dS(t_{ij})/dt$, and $f(t_{ij})$ being the proper density function to event of interest T_{ij} .

Under the random activation scheme, the event occurs when some possible competing causes are presented, given the number of competing causes M_{ij} , the number of activated competing causes is then a random variable with the discrete uniform distribution on $\{1, \dots, M_{ij}\}$. Its population survival function is given by $S_p^R(t_{ij}) = P(T_{ij} > t_{ij}) = G(0)/G(\theta_{ij}) + (G(\theta_{ij}) - G(0))S(t_{ij})/G(\theta_{ij})$ for $t_{ij} > 0$. The population density and the hazard function for the random activation scheme are given by $f_p^R(t_{ij}) = (1 - G(0)/G(\theta_{ij}))f(t_{ij})$, and $\lambda_p^R(t_{ij}) = \frac{(G(\theta_{ij}) - G(0))f(t_{ij})}{G(0) + (G(\theta_{ij}) - G(0))S(t_{ij})}$. The relationship between the first activation scheme and random activation scheme is $S_p^F(t_{ij}) \leq S_p^R(t_{ij})$, for all $t_{ij} > 0$.

In this work, we restrict our attention to four special cases, which are Bernoulli mixture model, Poisson (promotion time cure rate model), geometric (odds cure rate model), and logarithmic cure rate model. The spatial frailties get added to the cure rate models through linear form of covariates, as discussed in Section 1.4. The detailed information on the four cure models with spatial frailties are discussed in Section 3.2, and we omit the superscript F and R and use $f_p(t_{ij})$, $S_p(t_{ij})$, $\lambda_p(t_{ij})$ for simplicity. Although $f_p(t_{ij})$ and $S_p(t_{ij})$ are not proper density function and not proper survival function, it is not an issue for developing inference of the parameters.

Even though the general form of the population survival and density functions under the first activation scheme and random activation scheme are similar to these in Chapters 2 and 3, and when we add the spatial frailties to account for spatial correlation in the data and its effects on survival, the underlying cure models become quite different.

1.1.3 PS model with spatial frailties (Scenario 2)

The assumption in Chapters 2 and 3 is that the event of interest takes place after the first possible competing cause presents. In Chapter 4, to make our analysis applicable to more complicated scenario, we consider an additional case where event, T_{ij} , only take place after all the competing cause have occurred (last activation scheme).

Under the last activation scheme, the time to the event of interest is denoted by the random variable $T_{ij} = \max\{W_{1ij}, \dots, W_{M_{ij}}\}$ for $M_{ij} \geq 1$, and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P[T_{ij} = \infty | M_{ij} = 0] = 1$.

The survival function for the population in this case is $S_p^L(t_{ij}) = P(T_{ij} > t_{ij}) = 1 + G(0)/G(\theta_{ij}) - G(\theta_{ij}F(t_{ij}))/G(\theta_{ij})$ for $t_{ij} > 0$. The cure fraction is given by $p_{0ij} = G(0)/G(\theta_{ij}) > 0$. The corresponding density function is $f_p^L(t_{ij}) = G'(\theta_{ij}F(t_{ij}))\theta_{ij}f(t_{ij})/G(\theta_{ij})$ and the hazard function is $\lambda_p^L(t_{ij}) = \frac{G'(\theta_{ij}F(t_{ij}))f(t_{ij})}{G(0)+G(\theta_{ij})-G(\theta_{ij}F(t_{ij}))}$, where $G'(\theta_{ij}F(t_{ij})) = dG(\theta_{ij}F(t_{ij}))/dt$ and $f(t_{ij}) = -dS(t_{ij})/dt$. The density function for the population, $f_p^L(t_{ij})$, and survival function for the population, $S_p^L(t_{ij})$, are constructed in detail in Section 4.2. As shown in Section 1.4, the spatial frailties are from Gaussian and haven been added to the cure models though linear form of covariates. In Chapter 4, we omit the superscript L and use $f_p(t_{ij})$, $S_p(t_{ij})$, $\lambda_p(t_{ij})$ for simplicity.

1.2 A brief literature review

The spatial survival analysis of medical data has emerged over the last two decades. For example, the spatial survival analysis has been conducted on leukemia survival (Zhou and Hanson (2018)), asthma (Li and Ryan (2002); Li and Lin (2006)), infant mortality (Banerjee *et al.* (2003)), breast cancer (Hanson *et al.* (2012); Zhou *et al.* (2015)), smoking cessation (Pan *et al.* (2014); Bao *et al.* (2020)), HIV/AIDS survival (Martins *et al.* (2016)), the lifespan of tooth (Schnell *et al.* (2015)), lip cancer (Wilson and Wakefield (2020)), and some more.

Numerous methods have been proposed for spatial survival analysis, which can be broadly classified as applications of conditional model and marginal model. The conditional model assumes the regression coefficients are conditioned on the spatial frailties. The conditional model was introduced in the proportional hazards framework (Li and Ryan (2002); Banerjee *et al.* (2003)) and the proportional odds framework (Banerjee and Dey (2005)). Li and Lin (2006) proposed a marginal model that requires the data to be transformed to Gaussian and modelled directly (i.e., without frailties) by a Gaussian spatial model. Recently, Schnell (2016) proposed to model the frailties as linear combinations of positive stable random variables. Bao *et al.* (2020) assumed the spatial frailty to follow MCAR prior and used Bayesian analysis for estimating the model parameters.

1.3 Distributions for competing causes

In this thesis, we focus on several well-known distributions that belong to the family of power series cure rate model. As listed in Table 1.1, the competing causes are assumed

to follow Bernoulli, Poisson, geometric, and logarithmic distributions, giving rise to complementary cure rate model, complementary promotion time cure rate model, complementary geometric cure rate model, and complementary logarithmic cure rate model, respectively, along with spatial frailties.

Table 1.1: A list of distributions for the competing causes are assumed to follow in the analysis

	Power series cure rate model	
Chapter 2	Model 1	Bernoulli
	Model 2	Poisson
	Model 3	Geometric
	Model 4	Logarithmic
Chapter 3	Model 5	Bernoulli with Spatial Frailties
	Model 6	Poisson with Spatial Frailties
	Model 7	Geometric with Spatial Frailties
	Model 8	Logarithmic with Spatial Frailties
Chapter 4	Model 9	Complementary Poisson with Spatial Frailties
	Model 10	Complementary Geometric with Spatial Frailties
	Model 11	Complementary Logarithmic with Spatial Frailties

1.4 Gaussian spatial effect

Clustered data are frequently seen in survival analysis and it has been modelled by many using frailties for describing the cluster-specific random effect. In clustered data, observations within the same region or cluster, share similar condition and environment which are not easy to observe directly. Using frailties, the similarity and heterogeneity in the data can be captured within the same region. To achieve this goal, the spatial effect model has been constructed using Gaussian process, also

known as Gaussian random field. We use the spatial frailty, U_i , to interpret the effect of the i th geographic location on survival time for susceptible individuals within the region. First, let u_i be the realization of a stationary spatial frailty U_i and have U_i coming from a Gaussian process (Li and Ryan (2002); Wilson and Wakefield (2020)).

In the clustered dataset, we assume individuals belong to I disjoint regions (clusters). Then, to take into account the spatial frailties in each region, we construct the Gaussian process to be the stationary I -dimensional multivariate Gaussian, as proposed in Li and Ryan (2002). We define $\mathbf{U} = (U_1, U_2, \dots, U_I)^T$ and $\mathbf{U} \sim N(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}(\sigma_s, \phi_s))$, where $\boldsymbol{\mu}_s$ is a I -dimensional constant mean vector and $\boldsymbol{\Sigma}(\sigma_s, \phi_s)$ is the $I \times I$ dimensional variance-covariance matrix given by

$$\boldsymbol{\Sigma}(\sigma_s, \phi_s) = \begin{bmatrix} \text{Cov}(U_1, U_1) & \text{Cov}(U_1, U_2) & \text{Cov}(U_1, U_3) & \dots & \text{Cov}(U_1, U_I) \\ \text{Cov}(U_1, U_2) & \text{Cov}(U_2, U_2) & \text{Cov}(U_2, U_3) & \dots & \text{Cov}(U_2, U_I) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(U_I, U_1) & \text{Cov}(U_I, U_2) & \text{Cov}(U_I, U_3) & \dots & \text{Cov}(U_I, U_I) \end{bmatrix} \quad (1.5)$$

The pairwise covariance function, $\text{Cov}(U_i, U_{i^*})$, is the function of distance between subjects i and i^* . The isotropic covariance between U_i and U_{i^*} , takes an exponential form, $\Sigma_{ii^*}(\sigma_s, \phi_s) = \text{cov}(U_i, U_{i^*}) = \sigma_s^2 \exp\{-\|\text{Loc}_i - \text{Loc}_{i^*}\|/\phi_s\}$, where Loc_i and Loc_{i^*} are locations of subject i and subject i^* , respectively, $\|\cdot\|$ is the Euclidean distance, and I_s is the indicator function taking value 1 if $i = i^*$, and 0 otherwise.

Eq. (1.5) can be rewritten as

$$\Sigma(\sigma_s, \phi_s) = \begin{bmatrix} \sigma_s^2 & \sigma_s^2 \exp\{-d_{12}/\phi_s\} & \sigma_s^2 \exp\{-d_{13}/\phi_s\} & \dots & \sigma_s^2 \exp\{-d_{1I}/\phi_s\} \\ \sigma_s^2 \exp\{-d_{12}/\phi_s\} & \sigma_s^2 & \sigma_s^2 \exp\{-d_{23}/\phi_s\} & \dots & \sigma_s^2 \exp\{-d_{2I}/\phi_s\} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_s^2 \exp\{-d_{1I}/\phi_s\} & \sigma_s^2 \exp\{-d_{I2}/\phi_s\} & \sigma_s^2 \exp\{-d_{I3}/\phi_s\} & \dots & \sigma_s^2 \end{bmatrix}, \quad (1.6)$$

where $d_{ii^*} = \|\text{Loc}_i - \text{Loc}_{i^*}\|$, $\Sigma(\sigma_s, \phi_s)$ is positive semi-definite, $\mu_s \in \mathbb{R}$, $\sigma_s > 0$, and $\phi_s > 0$. The parameter ϕ_s depends on the spatial dependence between two locations, and a larger value of ϕ_s indicates a higher dependence. The density function of the multivariate normal is given by

$$f_U(u_1, u_2, \dots, u_I) = \frac{1}{\sqrt{(2\pi)^I |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{u} - \boldsymbol{\mu})\right). \quad (1.7)$$

The spatial version of cure models

The spatial effect on survival times of individuals, u_i , gets included in the models through linear form. The components for the effects of covariates on power parameter θ_{ij} is modified as in (1.8) below, and θ_{ij} is related to the cure probability p_{0ij} . The components for the effects of covariates on the survival models are as expressed in (1.9) below:

$$\alpha_{ij}^* = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i, \quad (1.8)$$

$$\phi_{ij}^* = \mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i, \quad (1.9)$$

for region $i = 1, \dots, I$, and individual j in region i , where $j = 1, \dots, n_i$. In addition, $\mathbf{x}$ and $\mathbf{z}$ are covariates, $\mathbf{b}$ is the effect of covariates associated with cure probabilities p_{0i} , and $\boldsymbol{\beta}$ is the effect of covariates associated with survival function.

The spatial version of the survival, hazards, and density functions can be re-written as

$$S(t_{ij}) = S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}, \quad (1.10)$$

$$\lambda(t_{ij}) = \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i), \quad (1.11)$$

$$f(t_{ij}) = \lambda(t_{ij}) S(t_{ij}), \quad (1.12)$$

where $S_0(t_{ij})$ is the baseline survival function and $\lambda_0(t_{ij})$ is the baseline hazard function defined in Section 1.5.

The spatial version of the power parameter θ_{ij} consider in this work is as follows:

$$\theta_{ij} = \begin{cases} \exp(x_{ij}^T \mathbf{b} + u_i), & \text{for Models 5, 6, 9.} \\ \frac{\exp(x_{ij}^T \mathbf{b} + u_i)}{1 + \exp(x_{ij}^T \mathbf{b} + u_i)}, & \text{for Model 7, 8, 10, 11,} \end{cases} \quad (1.13)$$

where Model 5 is the Bernoulli mixture model with multivariate Gaussian spatial effect, Model 6 is the promotion time cure model with Gaussian spatial effect, and Model 9 is the complementary promotion time cure model with Gaussian spatial effect. Models 7, 8, 10, and 11 are geometric, logarithmic, complementary geometric, and complementary logarithmic models with a Gaussian spatial effect.

Details of model development for Models 5-8 are discussed in Chapter 3. In Chapter 4, we focus on the model development of Models 9, 10, and 11.

1.5 Baseline Distribution: GEV distribution

The baseline hazard function is assumed to follow a GEV distribution, and $\log T \sim \text{GEV}$. The GEV distribution is a family of continuous probability distributions combining three distributions. As shape parameter γ changes, GEV distribution becomes the light-tailed Gumbel distribution (Type I variation of GEV) when $\gamma = 0$. The Type II of GEV is the heavy-tailed Fréchet distribution when $\gamma > 0$. Finally, $\gamma < 0$, it becomes the Weibull distribution (Type III of GEV); See Kotz *et al.* (2001). We can write $\log T \sim \text{GEV}(\mu, \sigma, \gamma)$, and the general form of baseline survival function is given by

$$S_0(t \mid \mu, \sigma, \gamma) = \begin{cases} 1 - \exp\left\{-\left(1 + \gamma \frac{\log t - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}}\right\}, & \text{if } \gamma \neq 0, \\ 1 - \exp\left\{-\exp\left(\frac{\log t - \mu}{\sigma}\right)\right\}, & \text{if } \gamma = 0, \end{cases} \quad (1.14)$$

where location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, shape parameter $\gamma \in \mathbb{R}$, and $x_+ = \max(0, x)$ (Kotz *et al.* (2001)).

The general form of the baseline hazard function is given by

$$\lambda_0(t \mid \mu, \sigma, \gamma) = \begin{cases} \frac{1}{\sigma t} \left(1 + \gamma \frac{\log t - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}} \left[\exp\left\{-\left(1 + \gamma \frac{\log t - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}}\right\} - 1\right], & \text{if } \gamma \neq 0, \\ \frac{1}{\sigma t} \exp\left(\frac{\log t - \mu}{\sigma}\right) \left[\exp\left\{\exp\left(\frac{\log t - \mu}{\sigma}\right)\right\} - 1\right]^{-1}, & \text{if } \gamma = 0. \end{cases} \quad (1.15)$$

1.6 Data structure and the likelihood

Let $T_i = \min\{T_i, C_i\}$, where C_i denotes the right censoring time for subject i . Let $\delta_i = I(T_i \leq C_i)$ and δ_i takes a value 1 if T_i is the observed lifetime, and 0 if it is right

censored. We define a cured status variable, J^* , as

$$J^* = \begin{cases} 1, & \text{if a subject is susceptible} \\ 0, & \text{if a subject is cured/immune.} \end{cases}$$

Let $P[J^* = 0] = p_{0i}$ and $P[J^* = 1] = 1 - p_{0i}$. Then, the population survival functions can be expressed as

$$\begin{aligned} S_p(t_i) &= P[T_i > t_i] \\ &= P[T_i > t_i | J_i^* = 0]P[J_i^* = 0] + P[T_i > t_i | J_i^* = 1]P[J_i^* = 1] \\ &= p_{0i} + (1 - p_{0i})S_s(t_i), \end{aligned} \tag{1.16}$$

where $S_s(t_i)$ is survival function of susceptible individuals.

The missing data are the unobserved cured status of the subjects with censored lifetimes. If the lifetime is censored and its cure status is known as $J_i^* = 0$, its contribution to the likelihood is p_{0i} . If the lifetime is censored and its cure status is unknown, $J_i^* = 1$, its contribution to the probability is $(1 - p_{0i})S_s(t_i)$, which equals $S_p(t_i) - p_{0i}$ from (1.16).

1.6.1 Non-informative censoring and likelihood

Under the non-informative censoring assumption, the likelihood function can be written as

$$L(\boldsymbol{\xi}; \mathbf{t}, \boldsymbol{\delta}, \mathbf{x}, \mathbf{z}) \propto \prod_{i=1}^I \{f_p(t_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi})\}^{\delta_i} \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i)\}^{1-\delta_i}, \tag{1.17}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$, $t_i = \min\{t_i, c_i\}$, $\delta_{ij} = I(T_i \leq C_i)$, $\mathbf{x}$ and $\mathbf{z}$ are covariates, $\mathbf{b}$ is the effect of covariates associated with cure probabilities p_{0i} , $\boldsymbol{\beta}$ is the effect

of covariates associated with survival function, and $\mathbf{t}, \mathbf{c}$ are the vectors of observed lifetimes and censored lifetimes.

The log-likelihood function takes on the form

$$\begin{aligned}
 l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) = & \sum_{i: \delta_{ij}=1} \log f_p(t_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) \\
 & + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
 & + \sum_{i: \delta_{ij}=0} J_i^* \log \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) - p_{0i}(\mathbf{x}_i; \mathbf{b})\}.
 \end{aligned} \tag{1.18}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$ is the model parameter.

1.6.2 Non-informative censoring and likelihood for spatially correlated data

For the i th region, $i = (1, \dots, I)$, j indicates the j th individual in the i th region, $j = (1, \dots, n_i)$, the complete likelihood function, with consideration to spatial frailties in the data, is given by

$$L(\boldsymbol{\xi}; \mathbf{t}, \boldsymbol{\delta}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \propto \prod_{i=1}^I \prod_{j=1}^{n_i} \{f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi})\}^{\delta_{ij}} \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i)\}^{1-\delta_{ij}} f_U(u_i; \boldsymbol{\theta}_s) \tag{1.19}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma, \boldsymbol{\theta}_s)'$, $t_{ij} = \min\{t_{ij}, c_{ij}\}$, $\delta_{ij} = I(T_{ij} \leq C_{ij})$, $\mathbf{x}$ and $\mathbf{z}$ are covariates, $\mathbf{b}$ is the vector of covariates associated with cure probabilities p_{0ij} , $\boldsymbol{\beta}$ is the vector of covariates associated with the survival function, and $\mathbf{y}, \mathbf{c}$ are the vectors of observed lifetimes and censored lifetimes. The u_i are random variables that denote the spatial effect from a Gaussian process with parameter $\boldsymbol{\theta}_s = (\mu_s, \sigma_s, \phi_s)'$.

The corresponding log-likelihood function under the consideration of spatial frailties is given by

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) = & \sum_{(i,j): \delta_{ij}=1} \sum \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) \\
& + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|u_i; \mathbf{b})\} \\
& + \sum_{(i,j): \delta_{ij}=0} \sum J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|u_i; \mathbf{b})\} \\
& + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{1.20}$$

where $\boldsymbol{\xi}$ is the vector of model parameters to be estimated, $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma, \boldsymbol{\theta}_s')'$. Also, the f_U is the density function of spatial frailties that follows a Gaussian process with parameter $\boldsymbol{\theta}_s = (\mu_s, \sigma_s, \phi_s)'$, f_p is the population density function, p_{0ij} is the cure probability, and S_p is the population survival function. The f_p , p_{0ij} , and S_p are all provided in Sections 3.2 and 4.2. The log-likelihood function (1.20) becomes different versions under Models 5 - 8 as shown in Section 3.3, and Models 9 - 11 in Section 4.3.

Recall that the survival data can be right censored, left censored, and interval censored. But, we focus here on the right censoring scenario in this thesis, since this is the most common form of censoring encountered in practice.

So far, the non-informative likelihood function of right censored data with spatial frailties has been shown; but, we still do not have complete information for the subjects who are right censored. The survival status remains unknown for the censored subjects, they can be either cured or susceptible. We overcome this difficulty by implementing the stochastic step (S-step) and then finding the optimal MLEs by

using the Stochastic EM algorithm. The general steps of the SEM algorithm for inference for the cure rate models with spatial frailties are provided in Section 1.7.2. The detailed formulas are all provided in Sections 2.4, 3.4, and 4.4.

1.7 Likelihood inference

The SEM algorithm, proposed by Celeux and Diebolt (1985), is an alternative approach to the EM algorithm. The SEM algorithm replaces the expectation step (E-step) in the EM algorithm with a stochastic step (S-step). The stochastic step is easy to implement when the missing data are not fairly imputable.

In this thesis, we have developed two versions of the SEM algorithm. Type I of the SEM algorithm is developed for the first scenario (as outlined in Chapter 2), where individuals in the data are treated as if they were from one cluster. It is equivalent to setting the spatial frailties to 0. In Chapters 3 and 4, the non-zero spatial frailties are included in the model, and studies are then conducted to show that adding the spatial frailties improves the performance of the model. When we shift our focus to include the spatial frailties for analyzing clustered data, we no longer treat the individuals to be from one big region, but we stratify them by the actual region they belong to. The Stochastic EM algorithm in such a case (Type II) is developed so that the parameter θ_s that describes the spatial effect in the model gets determined along with other model parameters.

The general steps of the SEM algorithms are presented in Sections 1.7.1 and 1.7.2, and the detailed steps of the corresponding algorithms are provided in Chapters 2, 3, and 4.

1.7.1 Type I: SEM

Step 1: Initialization

To properly start the SEM algorithm, we find the set of initial values using the grid search method based on 2000 values within the parameter space. The best set of parameter values, that are recorded as the initial values, is the one that maximizes the observed data log-likelihood function. The notation of the initialized parameters is given by $\xi^{(0)} = (\mathbf{b}^{(0)'}, \boldsymbol{\beta}^{(0)'}, \mu^{(0)}, \sigma^{(0)}, \gamma^{(0)})'$.

If the lifetime is observed, we have $\delta_i = 1$ and t_i , which implies that the subject is susceptible, and so $J_i^* = 1$. If the lifetime is unobserved, we have $\delta_i = 0$ and $t_i = c_i$, and the lifetime status of the subject is then unknown, with $J_i^* = 0$ if cured and $J_i^* = 1$ if uncured.

Step 2: Stochastic step (S-step)

Recall that for censored subject i , we have $\delta_i = 0$, and J_i^* can be generated from a Bernoulli distribution with conditional probability of success as

$$p_{0i}^{(0)} = P[J_i^* = 1 | T_i > c_i; \boldsymbol{\xi}^{(0)}] = \frac{P[T_i > c_i | J_i^* = 1]P[J_i^* = 1]}{P[T_i > c_i]} \Bigg|_{\boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}} \quad (1.21)$$

$$= \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}, \quad (1.22)$$

where p_{0i} and S_p are the cure probabilities and the survival function that are presented in the following chapters.

Step 3

If the censored subject is susceptible, $J_i^* = 1$, the complete lifetime t_i^* is from the

truncated distribution with density function and cumulative density function (cdf) as

$$f_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i | u_i; \boldsymbol{\xi}^{(0)}) = \frac{f_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}, \quad (1.23)$$

$$F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - S_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}, \quad (1.24)$$

where $c_i < t_i^* < \infty$. Here, $F_T(t_i^*)$ is not a proper cdf since $\lim_{t_i^* \rightarrow \infty} F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) \neq 1$.

To generate t_i^* from (1.23) under the susceptible scenario, we adopt inverse transformation sampling techniques. It is easy to show that $F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ follows an $\text{Uniform}(a^* = 0, b^* = \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})})$ distribution.

The cured/immunized subject with $J_i^* = 0$ is treated as long term survivor and the subject's lifetime is infinite with respect to the event of interest. Hence, it takes the form $\lim_{t_i^* \rightarrow \infty} S_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) = p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$.

The detailed derivation and definitions of cure probabilities p_{0i} and population survival probability S_p are all presented in Section 2.2.

Step 4: Maximization step (M-step)

We fill the censored data with the generated data from Step 3. Now, the improved estimate of $\boldsymbol{\xi}$ can be found using the pseudo-complete data and

$$\boldsymbol{\xi}^{(1)} = (\mathbf{b}^{(1)'}, \boldsymbol{\beta}^{(1)'}, \mu^{(1)}, \sigma^{(1)}, \gamma^{(1)})' = \arg \max_{\boldsymbol{\xi}^{(1)}} \log L_c(\boldsymbol{\xi}; (\mathbf{t}, \mathbf{J}^*), (\mathbf{t}^*, \mathbf{J}^{**}), \mathbf{x}, \mathbf{z}),$$

where $\mathbf{t}^*$ and $\mathbf{J}^{**}$ are vectors of t_i^* and J_i^{**} . The optimal value of $\boldsymbol{\xi}$ is obtained using the 'L-BFGS-B' package in R software, where the algorithm is set to be converged when the desired tolerance level, i.e., $|\hat{\xi}_{r+1} - \hat{\xi}_r| < 10^{-6}$, is achieved.

Step 5: Iterative step

Using the estimate, $\hat{\xi}^{(1)} = (\hat{\mathbf{b}}^{(1)'}, \hat{\beta}^{(1)'}, \hat{\mu}^{(1)}, \hat{\sigma}^{(1)}, \hat{\gamma}^{(1)})'$, that we obtained in **Step 4**, repeat **Steps 2 - 4** R times, to generate $\hat{\xi}^{(r)} = (\hat{\mathbf{b}}^{(r)'}, \hat{\beta}^{(r)'}, \hat{\mu}^{(r)}, \hat{\sigma}^{(r)}, \hat{\gamma}^{(r)})'$, $r = 1, \dots, R$. The result, a sequence of estimates, is a Markov Chain which, instead of converging to a single value, converges to a stationary distribution under standard conditions as discussed by Diebolt and Ip (1995).

Step 6: Burn-in and MLE step

To obtain the stationary distribution, we discard the first r^* iterations as a burn-in, and then compute the estimates by averaging every third of the remaining iterates to avoid auto-correlation. By adopting the burn-in period, the random perturbations of the Markov chains preclude the influence of local maximum, so that the estimates become more reliable. The choice of spacing is somewhat arbitrary (e.g., every 2nd, 3rd, or 5th iteration) and should be large enough to reduce autocorrelation between retained draws, but not so large that too much information is discarded.

Convergence of Algorithm

In this study, the optimal estimates are obtained using the ‘L-BFGS-B’ package in R software, where the algorithm is set to converge when the desired tolerance level, $|\hat{\xi}_{r+1} - \hat{\xi}_r| < 10^{-6}$, is achieved. The standard errors are obtained from the Hessian matrix with respect to the parameters, and they are obtained numerically.

1.7.2 Type II: SEM algorithm for cure rate models with spatial effect

Next, we shift our focus to include the spatial frailties to the proposed cure rate models for modelling clustered data. Recall that we assume individuals in the data are from

I regions, $i = (1, \dots, I)$. In the i th region, we have j individuals, $J = (1, \dots, n_i)$. In this case, we no longer treat the individuals in the data as being from one group, but we have stratified them by the region they belong to. Hence, when the lifetime of the (i, j) th individual in the i th region is observed, we denote it by $\delta_{ij} = 1$ and t_{ij} , which implies that subject j in region i is susceptible, and so $J_{ij}^* = 1$. If the lifetime is unobserved, we denote it by $\delta_{ij} = 0$ and $t_{ij} = c_{ij}$, and the lifetime status of the subject is then unknown, and so $J_{ij}^* = 0$ if cured or $J_{ij}^* = 1$ if uncured.

Now, a new set of initial values, $\xi^{(0)} = (\mathbf{b}^{(0)'}, \boldsymbol{\beta}^{(0)'}, \mu^{(0)}, \sigma^{(0)}, \gamma^{(0)}, \boldsymbol{\theta}_s^{(0)'})'$, is generated using the grid search method mentioned in the previous section. The expressions in Steps 2 and 3 are then modified to accommodate the spatial frailties captured in the model. The modified steps of SEM algorithms are as follows.

Step 2 (Spatial version)

Recall that for censored subject (i, j) in region i , we have $\delta_{ij} = 0$. The cure status of individual j in the i th region, J_{ij}^* , can be generated from a Bernoulli distribution with conditional probability of success as

$$p_{0ij}^{(0)} = P[J_{ij}^* = 1 | T_{ij} > c_{ij}; \boldsymbol{\xi}^{(0)}] = \frac{P[T_{ij} > c_{ij} | J_{ij}^* = 1] P[J_{ij}^* = 1]}{P[T_{ij} > c_{ij}]} \Bigg|_{\boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}} \quad (1.25)$$

$$= \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}, \quad (1.26)$$

where p_{0ij} and S_p are conditioned on the spatial frailties u_i . Also, p_{0ij} and S_p are modified to be the spatial versions of the cure probabilities and the survival function, and detailed information are then provided in Sections 3.2 and 4.2. Here, T_{ij} is the lifetime for the (i, j) th individual and c_{ij} is the censoring time.

Step 3 (Spatial version)

If the censored subject is susceptible, i.e., $J_{ij}^* = 1$, the complete lifetime T_{ij}^* is from the truncated distribution with density function and cdf given by

$$f_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = \frac{f_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}, \quad (1.27)$$

$$F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - S_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}, \quad (1.28)$$

where $c_{ij} < t_{ij}^* < \infty$. The cdf is not a proper cdf since

$$\lim_{t_{ij}^* \rightarrow \infty} F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})} \neq 1. \quad (1.29)$$

The cured/immunized subject with $J_{ij}^* = 0$ is treated as long term survivor and the subject's lifetime is infinite with respect to the event of interest. Hence, it takes the form of $\lim_{t_{ij}^* \rightarrow \infty} S_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})$. To generate t_{ij}^* from (1.27) under the susceptible scenario, we adopt inverse transformation sampling technique. It is easy to show that $F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})$ follows an $\text{Uniform}(a^*, b^*)$ distribution with parameters

$$a^* = 0 \text{ and } b^* = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}. \quad (1.30)$$

where p_{0ij} and S_p are conditioned on the spatial frailty u_i . Also, p_{0ij} and S_p are the cure probabilities and the survival function for Models 5 - 10. Derivation of all relevant formulas are provided in Sections 3.2 and 4.2.

1.7.3 Standard errors and asymptotic confidence intervals based on MLEs

The standard errors of estimators are calculated using the observed information matrix. To be more specific, the variance-covariance matrix of $\hat{\xi}$ can be obtained by the inverse of the information matrix as

$$Var(\hat{\xi}) = [I(\hat{\xi})]^{-1} \quad (1.31)$$

The observed information matrix is the negative of the expected value of the Hessian matrix, where the Hessian matrix is the second derivatives of the log-likelihood function with respect to the parameters given by

$$[I(\hat{\xi})] = -E[H(\hat{\xi})] = -E \left[\frac{\partial^2 l(\xi)}{\partial \xi \partial \xi'} \right] \bigg|_{\xi=\hat{\xi}}. \quad (1.32)$$

Then, the standard errors of the estimator, $\hat{\xi}$, is given by the square roots of the diagonal terms in the variance-covariance matrix. The asymptotic 100(1- δ)% confidence interval of ξ is given by

$$(\hat{\xi} - z_{1-\delta/2} se(\hat{\xi}), \hat{\xi} + z_{1-\delta/2} se(\hat{\xi})), \quad (1.33)$$

where $se(\hat{\xi})$ is the standard error of the MLE $\hat{\xi}$ and z_α is the upper α percentage point of the standard normal distribution.

1.8 Simulation study

1.8.1 General setup

The simulation studies conducted include different sample sizes: 500 and 1000. Various different sets of true values are chosen to ensure different levels of cure rates. In addition, the simulated dataset have low to high level of censoring. The underlying competing cause, M_i , is fixed to follow Bernoulli, Poisson, geometric, and logarithmic distributions, which is equivalent to producing data with Bernoulli cure rate (Model 1), Promotion time cure rate (Model 2), Geometric cure rate (Model 3), and logarithmic cure rate (Model 4) models. One categorical variable, X_i , is considered, and is set to follow Bernoulli distribution with success probability 0.6. Censoring time, C_i , is set to follow an exponential distribution, $\text{Exp}(cc)$, where $cc > 0$ controls percentage of censored data. The lifetime, Y_i , is generated from the quantile function of the logGEV distribution with parameters, μ , σ , and γ . If $Y_i \leq C_i$, we set censoring indicator $\delta_i = 1$, and $\delta_i = 0$ otherwise.

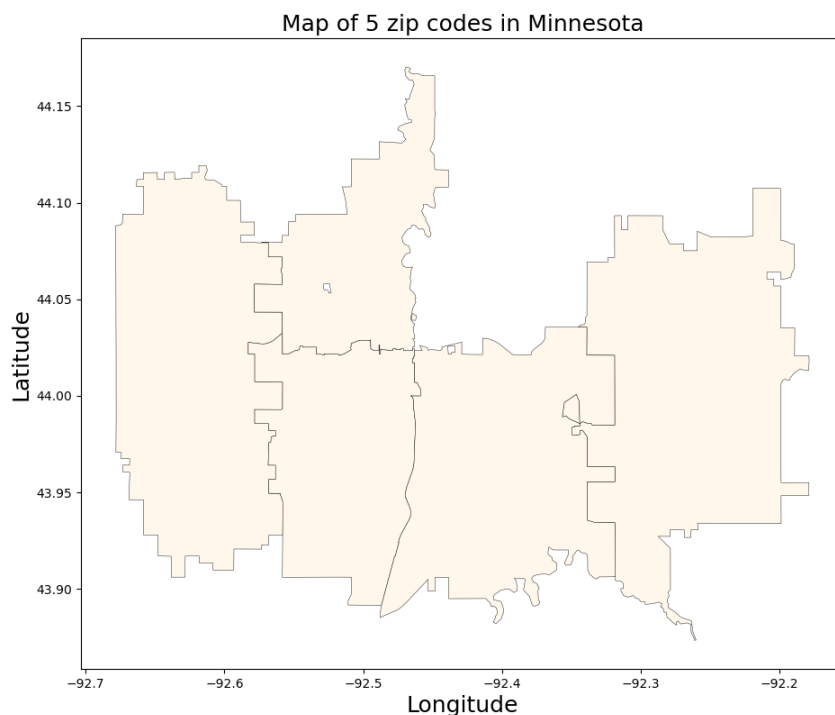
We monitor the censoring proportion for each simulated data and run the algorithm 1200 times. The burn-in period is set to 200 iterations, and the spacing is set to 3.

1.8.2 Cure rate models with spatial survival analysis

In Chapters 3 and 4, to reveal the spatial effects on survival time, the location of patients is taken into account in the simulation studies. We assume that 500 and 1000 subjects were from 5 zip codes in Minnesota, as shown in Figure 1.1. In each region we have 100 and 200 patients, corresponding to 500 and 1000 sample sizes,

respectively. The latitude, longitude, and zip code are all listed in Table 1.2, and the euclidean distance matrix of the 5 zip codes is presented in Table 1.3.

Figure 1.1: Map of 5 zip codes in Minnesota



For the j th individual in the i th region, we assume the corresponding competing causes, M_{ij} , follow Bernoulli, Poisson, geometric, and logarithmic distributions. Then, we obtain their corresponding cure rate models with Gaussian spatial effect, which are Model 5 (Bernoulli cure rate model with spatial effect), Model 6 (promotion time cure rate model with spatial effect), Model 7 (geometric cure rate model with spatial effect), Model 8 (logarithmic cure rate model with spatial effect), Model 9 (complementary promotion time cure rate model with spatial effect), Model 10

Table 1.2: Latitude, longitude, zip code, and city names of the 5 regions in Minnesota, US, used in the simulation study.

Latitude	Longitude	Zip code	City
44.016	-92.624	55920	Salem Minnesota
44.061	-92.504	55901	Rochester Minnesota
43.985	-92.499	55902	Rochester Minnesota
43.973	-92.414	55904	Bear Creek Marion Minnesota
43.999	-92.259	55934	Eyota Township Minnesota

(complementary cure rate model with spatial effect), and Model 11 (complementary logarithmic cure rate model with spatial effect). The algorithm is run for 1200 times on the simulated data. The burn-in period is set to be 200, and the spacing is set to be 3. Let covariate, x_{ij} , be a categorical variable that takes values 0 and 1, and gets generated from a Bernoulli distribution with success probability as 0.6.

Table 1.3: The Euclidean distance matrix of the 5 cities and their zip codes in Minnesota, US.

	(55920)	(55901)	(55902)	(55904)	(55934)
Salem (55920)	0.00000	0.12816	0.12879	0.21436	0.36540
Rochester (55901)	0.12816	0.00000	0.07616	0.12587	0.25272
Rochester (55902)	0.12879	0.07616	0.00000	0.08584	0.24041
Bear Creek Marion(55904)	0.21436	0.12587	0.08584	0.00000	0.15717
Eyota (55934)	0.36540	0.25272	0.24041	0.15717	0.00000

1.9 Model discrimination

1.9.1 Information-Based criterion

As one would expect, an increase in the likelihood value can be achieved by increasing the number of parameters. In order to avoid over fitting, the Akaike information

criterion (AIC; Akaike (1974)) and Bayesian information criterion (BIC; Schwarz (1978)) are used to resolve this problem by including a term to penalize based on the number of free parameters. The AIC is given by

$$\text{AIC} = -2\ln(L(\hat{\xi})) + 2k, \quad (1.34)$$

where $\ln(L(\hat{\xi}))$ represents the value of the maximum log-likelihood of the considered model and k is the number of parameters in the model. The idea of AIC is to evaluate the performance of the models by the goodness of fit while penalizing for an increase in the number of parameters. The model that gives the minimum AIC value is the best model to be selected among the candidate models.

The Bayesian information criterion (BIC) has been widely used as a criterion for model selection. It is an alternative to AIC, and this criterion is given by

$$\text{BIC} = -2\ln(L(\hat{\xi})) + k \ln(n), \quad (1.35)$$

where $\ln(L(\hat{\xi}))$ represents the value of the maximum log-likelihood for the estimated model, k is the number of free parameters to be estimated, and n is the number of observations. The model with the lowest value of BIC is the one that is preferred. The value of BIC increases when the variation in the dependent variable and the number of explanatory variables increase. A lower BIC value indicates that the model has fewer variables, provides a better fit to the data, or both. BIC is more strict with free parameters than AIC, and BIC does not require the test model to be nested.

The Akaike Information Criterion corrected (AICc) is the corrected version of AIC designed to address the issue of over fitting for the models with small sample sizes.

It is particularly useful when the ratio $\frac{n}{k} < 40$, where n is the number of observations and k is number is parameters in the model. AICc is given by

$$\text{AICc} = -2\ln(L(\hat{\xi})) + 2k + \frac{2k(k+1)}{n-k-1} \quad (1.36)$$

where $\ln(L(\hat{\xi}))$ represents the value of the maximum log-likelihood of the model and k is the number of parameters in the model. AICc converges to AIC when n gets large (Sugiura (1978)). All three information-based criteria have been used in this thesis.

1.9.2 Likelihood ratio test

The idea of the likelihood ratio test (LRT) is to compare the unrestricted and restricted models, provided that the simpler model is a special case of the unrestricted model. It is also called the comparison of two nested models. In terms of hypothesis testing, the null hypothesis is defined as the test parameters are all not all zero, which means the restricted model is the correct model to choose. The alternative hypothesis is defined as the test parameters are not all zero, which means the unrestricted model is the correct model to choose. In order to conduct the LRT, the likelihood values of both models need to be determined.

The LRT statistics is defined as

$$LRT = -2\ln\left(\frac{\hat{L}_r}{\hat{L}_u}\right) = -2(\hat{l}_r - \hat{l}_u) \sim \chi^2(g), \quad (1.37)$$

where $\hat{l}_r$ is the maximized log-likelihood value for the reduced (constrained) model and $\hat{l}_u$ is that of the full model (unconstrained) model, and g is the number of testing

parameters.

For the real data analysis in Chapter 3, we apply the LRT to compare Bernoulli (Model 1) and Bernoulli with spatial frailties (Model 5), Poisson (Model 2) and Poisson with spatial frailties (Model 6), geometric (Model 3) and geometric with spatial frailties (Model 7), and logarithmic (Model 4) and logarithmic with spatial frailties (Model 8). Comparing these four pairs of models is equivalent to investigating the necessity of adding parameters to include spatial frailties to the cure rate models. The null hypotheses is that the cure model, Models 1 - 4 without spatial frailties ($u_i = 0$), versus the alternative hypothesis is the cure rate models with a spatial component included in them. Models 1 - 4 are special cases of Models 5 - 8, and so the LRT is appropriate to apply here.

1.9.3 Parameter evolution of SEM

It is worth mentioning that the convergence characteristic of Stochastic EM is meant to better help one to detect convergence of the proposed iterative algorithm. One can visualize the results obtained from the iterations. A successful convergence of SEM is equivalent to its results of SEM iterations oscillating around the horizontal line without indication of either upward or downward trend. As established in the work by Nielsen (2000), a parameter evolution plot with horizontal average value indicates the convergence of SEM iterations to a stationary distribution, and so it is reasonable to use the average of the iterations after the burn-in to obtain the estimate of the parameter.

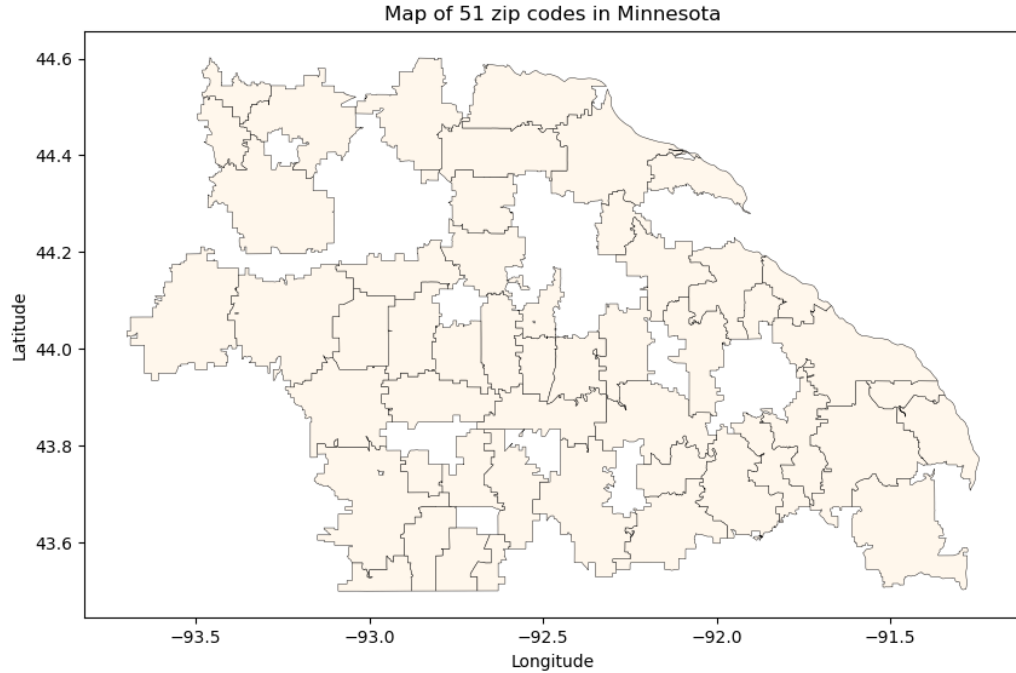
1.10 Smoking cessation dataset

In the real data analysis section, we apply the proposed cure rate models to the smoking cessation data that were collected in the 51 zip codes in the southeastern corner of Minnesota, United States. The 51 zip codes have been visualized via a map in Figure 1.2. The original data contain records of 223 smokers who enrolled in the study between 1986 and 1989. They were randomly split into two groups, with one group receiving smoking intervention (SI) and the other receiving the usual care (UC). As shown in Table 1.4, some other information is also recorded in the study, such as the gender (Male=0, Female=1), duration as smoker in years, the average number of cigarettes smoked per day over the last 10 years, and their corresponding zip codes (51 zip codes).

The smokers were randomly put into two groups: usual care (UI) and smoking intervention (SI). At the end of this study, 65 of them had relapsed, and 158 of them were censored, which give a high level of censoring proportion, 0.708. However, the cure status of the patients who had not relapsed remains unknown. It will be of interest to better understand the effect of the location of patients on his/her relapse of smoking while excluding the possibility of labelling the individual as cured due to early stopping of the experiment.

Due to limited information in this dataset, we do not have access to the detailed clinical definitions of usual care and the smoking intervention used in the experiment. Other potential confounding factors are outside the scope of this thesis and therefore are not explored further.

Figure 1.2: Map of 51 zip codes in Minnesota



1.11 Scope of the thesis

In Chapters 2 - 4, the proposed cure rate models are all discussed in detail along with the corresponding steps of the stochastic EM algorithm. Simulation study and real data analysis for proposed models are conducted, and the obtained results are provided. In Chapter 2, we adopt the stochastic EM algorithm to a family of cure rate models from the power series cure rate model. They are Bernoulli cure rate model

Table 1.4: Variables in the smoking cessation dataset in southeastern corner of Minnesota

Treatment	1 = special intervention [SI] 0 = usual care [UC]	SI = 169 UC = 54
Gender	male = 0 female = 1	136 (male) 87 (female)
Consumption	the average number of cigarettes smoked/day	range: 5-60 mean: 27.1 /day
Duration	duration of smoking habit (years)	range: 12-46 yrs mean: 30.5 yrs

(Model 1), promotion time cure rate model (Model 2), geometric cure rate model (Model 3), and logarithmic cure rate model (Model 4), with the GEV distribution as baseline hazard distribution. The performance of the proposed models is evaluated through a simulated study along with model discrimination using information-based criteria. In addition, the proposed cure rate models are applied to a real dataset, and cure rate and survival probabilities are visualized through heat maps. With variation of colours, the heat map indicates high and low values of cure and survival probabilities among different regions. Next, in Chapter 3, we then extend the work to include spatial frailties and study the spatial effect on cure rate and survival probabilities. The Bernoulli cure rate model with spatial effect (Model 5), promotion time cure rate model with spatial effect (Model 6), geometric cure rate model with spatial effect (Model 7), and logarithmic cure rate model with spatial effect (Model 8) are all constructed with baseline GEV distribution. The stochastic EM algorithm is then used to find the optimal estimates of the model parameters. A simulation study is performed with patients being fixed to be from specific regions. Additionally, the samples were generated from a true model and fitted all the relevant candidate models. The fitted results are then compared using information-based criteria. The selection

rates of all candidate models are also reported to illustrate the model performance. Furthermore, model discrimination using LRT confirms the improvement in the models by adopting spatial frailties using the spatial information of the patients (longitude and latitude). The presence of spatial effect is successfully captured by the models proposed, and the necessity of adding spatial effect to the model are also visualized using heat maps.

In Chapter 4, we extend the research work to the case when the event takes place after all of the competing causes are presented, along with the consideration of the spatial effect on to cure rate and survival probabilities. Complementary promotion time cure rate model with spatial effect (Model 9), complementary geometric cure rate model with spatial effect (Model 10), and complementary logarithmic cure rate model with spatial effect (Model 11) are constructed, and the MLEs of these model parameters are obtained by using the stochastic EM algorithm. The model performance is evaluated by a simulation study and also by model discrimination. The samples are generated from a true model and then fitted with some candidate models. The fitted results are then compared using information-based criteria. The selection rates of all candidate models are also reported to illustrate the model performance. The proposed models are then applied to the real data on smoking cessation. The difference in cure rate and survival probabilities is also demonstrated using maps.

The concluding remarks, along with some future research directions, are finally mentioned in Chapter 5.

Chapter 2

Stochastic EM-based Likelihood Inference for a Class of Cure Rate Models Based on GEV Distribution

2.1 Introduction

Suppose the i th individual is exposed to competing risk M_i , where M_i denotes the initial number of competing causes relating to the occurrence of an event. Given $M_i = m$, let random variable W_{k^*i} be the time-to-event due to k^* th latent risk for the i th individual, with distribution function $F(\cdot) = 1 - S(\cdot)$, survival function $S(\cdot)$, and $k^* = 1, 2, \dots, m$.

In reality, we cannot observe the competing causes M_i and the lifetime related to

a specific cause, W_{k^*i} . Recall that under the first activation scheme, assuming that any competing cause will eventually trigger the event, the time to the event of interest is denoted by the random variable $T_i = \min\{W_{1i}, \dots, W_{m_i}\}$ for $M_i \geq 1$, and $T_i = \infty$ if $M_i = 0$ with $P[Y_i = \infty \mid M_i = 0] = 1$. Under the random activation scheme, the event occurs when some possible competing causes are presented, given the number of competing causes M_i , the number of activated competing causes is then a random variable with the discrete uniform distribution on $\{1, \dots, M_i\}$.

In Section 2.2, details of the proposed cure rate models, namely, Bernoulli cure rate model (Model 1), Poisson cure rate model (Model 2), geometric cure rate model (Model 3), and logarithmic cure rate model (Model 4), are all presented. The likelihood function and estimation methods using SEM are described in Section 2.3 and Section 2.4, respectively. In Section 2.5, various settings with different sample sizes are considered for the simulation study. The flexible baseline of GEV successfully captures the tail behaviour of the data. In addition, the purposed models are applied to analyze a real data in Section 2.6. The cure rate and survival probabilities are then visualized using plots and maps.

2.2 Cure rate models

When competing cause, M_i , follows Bernoulli or Poisson distributions, we obtain Bernoulli mixture cure rate model (Model 1) or Promotion time cure rate model (Model 2) from the PS cure rate model. In this case, the power parameter of the distribution is $\theta_i = \exp(\mathbf{x}_i^T \mathbf{b})$, where $\theta_i > 0$, and $\mathbf{b}$ is the vector indicating the effects of covariates on θ_i , with θ_i being associated with cure probability.

Bernoulli mixture cure rate model (Model 1)

By choosing the series function of PS model, $G(\theta_i)$, to be $G(\theta_i) = (1 + \theta_i)$, and combining it with the choice of the power parameter of distribution $\theta_i = \exp(\mathbf{x}_i^T \mathbf{b})$, the competing cause M_i follows a Bernoulli distribution. In this case, we obtain a Bernoulli mixture cure model (Model 1), with

$$p_{0i} = \frac{1}{1 + \theta_i}, \quad (2.38)$$

$$S_p(t_i) = p_{0i} + (1 - p_{0i})S(t_i), \quad (2.39)$$

$$f_p(t_i) = \frac{\theta_i}{(1 + \theta_i)} f(t_i). \quad (2.40)$$

where p_{0i} are the cure probabilities, $S_p(t_i)$ is the population survival function for Model 1, and $S(t_i)$ is the survival function calculated from the baseline survival function as $S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}$. The density function is $f(t_i) = \lambda(t_i) * S(t_i)$, and the hazard function $\lambda(t_i)$ is calculated from the baseline hazard function as $\lambda = \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})$. Also, $\boldsymbol{\beta}$ is a vector representing the effects of covaraites on the survival model component.

Promotion time cure rate model (Model 2)

By choosing the power parameter of the PS distribution as $\theta_i = \exp(\mathbf{x}_i^T \mathbf{b})$ and set the series function $G(\theta_i)$ to be $G(\theta_i) = \exp(\theta_i)$, M_i follows a Poisson distribution, and the promotion time cure rate model (Model 2) is obtained in this case from the PS cure rate model. The corresponding cure probability, population survival and population

density functions are given by

$$p_{0i} = \exp(-\theta_i) \quad (2.41)$$

$$S_p(t_i) = \exp(-\theta_i F(t_i)), \quad (2.42)$$

$$f_p(t_i) = \theta_i f(t_i) \exp(-\theta_i F(t_i)), \quad (2.43)$$

where $F(t_i) = 1 - S(t_i)$, $S(t_i) = S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}$, and $\boldsymbol{\beta}$ is a vector representing the effects of covaraites on the survival model component. The hazard function λ is $\lambda = \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})$, $\lambda_0(t_i)$ is the baseline hazard function that follows the GEV distribution as discussed in (1.15), and $S_0(t_i)$ is the baseline survival function from GEV distribution in (1.14).

When the competing cause follows the geometric and logarithmic distributions, the geometric cure rate model (Model 3) and logarithmic cure rate model (Model 4) are obtained from the PS cure rate model. The power parameter θ_i in these case is set to be

$$\theta_{ij} = \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}, \quad (2.44)$$

where $0 < \theta_i < 1$, and $\mathbf{b}$ is the vector representing the effects of covariates on θ_i . Also, θ_i is associated with cured probability p_{0ij} .

Geometric cure rate model (Model 3)

Setting series function to be $G(\theta_i) = \frac{1}{1-\theta_i}$, and combining it with the choice of power parameter in (2.44), M_i follows a geometric distribution, and the geometric cure rate model (Model 3) is then obtained from the PS cure rate model. The cure probability,

population survival and population density functions given by

$$p_{0i} = 1 - \theta_i \quad (2.45)$$

$$S_p(t_i) = \frac{1 - \theta_i}{1 - \theta_i S(t_i)}, \quad (2.46)$$

$$f_p(t_i) = \theta_i(1 - \theta_i)f(t_i)[1 - \theta_i S(t_i)]^{-2}, \quad (2.47)$$

where $S(t_i) = S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}$, the density function constructed using baseline hazard function is $f(t_i) = \lambda S(t_i)$, and $\lambda = \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})$. $\lambda_0(t_i)$ and $S_0(t_i)$ are the baseline hazard and survival functions from the GEV distribution as presented in (1.15) and (1.14). Unlike $\mathbf{b}$ focusing on the effect on cure probabilities, $\boldsymbol{\beta}$ is a vector representing the effects of covariates on the survival component.

Logarithmic cure rate model (Model 4)

The PS cure rate model changes to a logarithmic cure rate model (Model 4) if M_i follow a logarithmic distribution with series function $G(\theta_i) = \frac{-\log(1-\theta_i)}{\theta_i}$ and power parameter θ_i as in (2.44). The cure probability, population survival and population density functions in this case are given by

$$p_{0i} = \frac{-\theta_i}{\log(1 - \theta_i)}, \quad (2.48)$$

$$S_p(t_i) = \frac{\log(1 - \theta_i S(t_i))}{S(t_i) \log(1 - \theta_i)}, \quad (2.49)$$

$$f_p(t_i) = -\frac{f(t_i)}{S(t_i) \log(1 - \theta_i)} \left[\frac{\log(1 - \theta_i S(t_i))}{S(t_i)} + \frac{\theta_i}{1 - \theta_i S(t_i)} \right]. \quad (2.50)$$

where the survival function $S(t_i) = S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}$, and $\boldsymbol{\beta}$ is the vector representing the effects of covariates on the survival model component. The density function

$f(t_i) = \lambda S(t_i)$, with the hazard function being $\lambda = \lambda_0(t_i)\exp(\mathbf{z}_i^T \boldsymbol{\beta})$, where $\lambda_0(t_i)$ and $S_0(t_i)$ are the baseline hazard and survival functions from the GEV distribution as given in (1.15) and (1.14).

2.3 The expression of log-likelihood function for cure rate models

2.3.1 Bernoulli mixture cure rate model (Model 1)

In case the competing cause of event M_i follow a Bernoulli distribution, the log-likelihood as the complete likelihood function is given by

$$\begin{aligned}
 l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) = & \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{(1 + \exp(\mathbf{x}_i^T \mathbf{b}))} \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\} \\
 & + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
 & + \sum_{i: \delta_i=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\},
 \end{aligned} \tag{2.51}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$. The detailed steps of obtaining log-likelihood function in this case are presented in Appendix A.

Recall that the baseline survival function and baseline hazard function have two forms depending on parameter γ . When $\gamma = 0$, the log-likelihood function is given

by

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) = & \sum_{i: \delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{(1 + \exp(\mathbf{x}_i^T \mathbf{b}))} \frac{1}{\sigma t_i} \exp\left(-\frac{\log(t_i) - \mu}{\sigma}\right) \right. \\
& \times \left[\exp \left\{ \exp\left(-\frac{\log(t_i) - \mu}{\sigma}\right) \right\} - 1 \right]^{-1} \\
& \times \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \left[1 - \exp \left\{ -\exp\left(\frac{\log(t_i) - \mu}{\sigma}\right) \right\} \right]^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \Big\} \\
& + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
& + \sum_{i: \delta_{ij}=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right) \right. \\
& \quad \times \left. \left[1 - \exp \left\{ -\exp\left(\frac{\log(t_i) - \mu}{\sigma}\right) \right\} \right]^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\}.
\end{aligned} \tag{2.52}$$

When $\gamma \neq 0$, the log-likelihood function has the general form of

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z})p = & \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{(1 + \exp(\mathbf{x}_i^T \mathbf{b}))} \frac{1}{\sigma t_i} \left(1 + \gamma \frac{\log(t_i) - \mu}{\sigma} \right)_+^{\frac{-1}{\gamma} - 1} \right. \\
& \times \left[\exp \left\{ \left(1 + \gamma \frac{\log(t_i) - \mu}{\sigma} \right)_+^{\frac{-1}{\gamma}} \right\} - 1 \right]^{-1} \\
& \times \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \left[1 - \exp \left\{ - \left(1 + \gamma \frac{\log(t_i) - \mu}{\sigma} \right)_+^{\frac{-1}{\gamma}} \right\} \right]^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \Big\} \\
& + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
& + \sum_{i: \delta_i=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right) \right. \\
& \quad \times \left. \left[1 - \exp \left\{ - \left(1 + \gamma \frac{\log(t_i) - \mu}{\sigma} \right)_+^{\frac{-1}{\gamma}} \right\} \right]^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\},
\end{aligned} \tag{2.53}$$

where $X_+ = \max(0, X)$, $\mu \in R$ and $\sigma > 0$. Then, it can be further split into cases where $\gamma > 0$ and $\gamma < 0$ corresponding to Fréchet and Weibull baseline distributions, respectively.

2.3.2 Poisson cure rate model (Model 2)

In case the competing cause of event M_i follow a Poisson distribution and the assumption of the event of interest occur when first competing cause presents, by combining (2.41) - (2.43) the log-likelihood the complete likelihood function is given by

$$\begin{aligned}
 l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) = & \sum_{i: \delta_i=1} \log \left\{ \exp(\mathbf{x}_i^T \mathbf{b}) \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right. \\
 & \times \exp \left[- \exp(\mathbf{x}_i^T \mathbf{b}) (1 - S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}) \right] \left. \right\} \\
 & + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{ \exp(-\exp(\mathbf{x}_i^T \mathbf{b})) \} \\
 & + \sum_{i: \delta_i=0} J_i^* \log \left\{ \exp \left[- \exp(\mathbf{x}_i^T \mathbf{b}) (1 - S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}) \right] - \exp(-\exp(\mathbf{x}_i^T \mathbf{b})) \right\}.
 \end{aligned} \tag{2.54}$$

It changes to two forms depending on the shape parameter, γ , of the baseline. The detailed steps of obtaining this log-likelihood function for Model 2 are presented in Appendix A.

2.3.3 Geometric cure rate model (Model 3)

When the competing cause of event, M_i , follow a geometric distribution, the geometric cure rate model has the complete log-likelihood function as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) &= \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \left(\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right) \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \right. \\
&\quad \times S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \left[1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right]^{-2} \Big\} \\
&+ \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
&+ \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}} - \left[\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right] \right\}.
\end{aligned} \tag{2.55}$$

This changes to two forms depending on the shape parameter, γ , of the baseline. The detailed steps of obtaining the log-likelihood function for Model 3 are presented in Appendix A.

2.3.4 Logarithmic cure rate model (Model 4)

In case the competing cause of event M_i follow a logarithmic distribution, PS cure rate model reduced to the logarithmic cure rate model, with its complete log-likelihood

function as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) = & \sum_{i: \delta_i=1} \log \left\{ \frac{-\lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right. \\
& \times \left[\frac{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}} + \frac{\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}} \right] \Bigg\} \\
& + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right\} \\
& + \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} - \left(\frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right) \right\}.
\end{aligned} \tag{2.56}$$

The log-likelihood has two forms depending on shape parameter, γ , of the baseline. The detailed steps of obtaining the log-likelihood function for Model 4 are presented in Appendix A.

In fact, we do not have complete information for the subjects that are right censored. The survival status remain unknown for the censored subjects since they can be either cured or susceptible. We overcome this difficulty by implementing the stochastic step (S-step) described in the following section.

2.4 Stochastic Expectation-Maximization (SEM) algorithm for cure rate models

To properly start the SEM algorithm, we fix a set of initial values using the grid search method based on 2000 points within the parameter space. The best set of

parameter values, that are chosen as the initial values, is the one that maximizes the observed data log-likelihood function.

Step 1: Initialization

We use the recorded set of parameters that optimize the log-likelihood function using grid search. The notation of the initialized parameter is

$$\xi^{(0)} = (\mathbf{b}^{(0)'} , \boldsymbol{\beta}^{(0)'} , \mu^{(0)} , \sigma^{(0)} , \gamma^{(0)})'.$$

If lifetime is observed, we have $\delta_i = 1$ and t_i , which implies the subject is susceptible, and so $J_i^* = 1$. If the lifetime is unobserved, we have $\delta_i = 0$ and $t_i = c_i$, and the lifetime status of the subject is unknown, and in this case $J_i^* = 0$ if cured and $J_i^* = 1$ if uncured.

Step 2: Stochastic step (S-step)

Recall that for censored subject i , we have $\delta_i = 0$, and J_i^* can be generated from a Bernoulli distribution with conditional probability of success as

$$p_{0i}^{(0)} = P[J_i^* = 1 | T_i > c_i; \boldsymbol{\xi}^{(0)}] \quad (2.57)$$

$$= \frac{P[T_i > c_i | J_i^* = 1] P[J_i^* = 1]}{P[T_i > c_i]} \Big|_{\boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}} \quad (2.58)$$

$$= \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}, \quad (2.59)$$

where p_{0i} and S_p are the cure probabilities and the survival function are as in (2.38),

and (2.39) presented in Section 2.2.

Step 3

If the censored subject is susceptible, i.e., $J_i^* = 1$, the complete lifetime t_i^* is from the truncated distribution with density function as

$$f_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i | u_i; \boldsymbol{\xi}^{(0)}) = \frac{f_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i |; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i |; \boldsymbol{\xi}^{(0)})}, \quad (2.60)$$

where $c_i < t_i^* < \infty$. The corresponding cdf is

$$F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - S_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}, \quad (2.61)$$

where $c_i < t_i^* < \infty$, which is not a proper cdf since

$$\lim_{t_i^* \rightarrow \infty} F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})} \neq 1. \quad (2.62)$$

The cured/immunized subject with $J_i^* = 0$ is treated as long term survivor and its theoretical lifetime is infinity with respect to the event of interest. Hence, it takes form of $\lim_{t_i^* \rightarrow \infty} S_p(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) = p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$.

To generate t_i^* from (2.60) under the susceptible scenario, we adopt inverse transformation sampling technique. It is easy to show that $F_T(t_i^*, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ follows an $\text{Uniform}(a^*, b^*)$ distribution with parameters

$$a^* = 0 \text{ and } b^* = \frac{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)}) - p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})}{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})}. \quad (2.63)$$

Model 1

When assuming latent cause, M_i , follow Bernoulli distribution, we look at the Bernoulli mixture cure rate (Model 1) from the PS cure rate model, $S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ and $p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$ are replaced by (2.38) and (2.39). Now b^* is given by

$$b^* = \frac{\exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{1 + \exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}. \quad (2.64)$$

Model 2

When assuming latent cause, M_i , follow Poisson (promotion time cure rate (Model 2)), $S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ and $p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$ are replaced by (2.41) and (2.42). Also, b^* is given by

$$\begin{aligned} b^* &= \frac{\exp(-\theta_i) \exp(\theta_i S(c_i)) - \exp(-\theta_i)}{\exp(-\theta_i) \exp(\theta_i S(c_i))} = \frac{\exp(\theta_i S(c_i)) - 1}{\exp(\theta_i S(c_i))} \\ &= 1 - \exp\{-\theta_i S(c_i)\} = 1 - \exp\{-\exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}\} \end{aligned} \quad (2.65)$$

Model 3

When assuming latent cause, M_i , follow geometric distribution, we obtain geometric cure rate model (Model 3), with $S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ and $p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$ being given by (2.45) and (2.46). Also, b^* is given by

$$\begin{aligned} b^* &= \frac{\frac{1-\theta_i}{1-\theta_i S(c_i)} - (1-\theta_i)}{\frac{1-\theta_i}{1-\theta_i S(c_i)}} = \theta_i S(c_i) \\ &= \exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} / \{1 + \exp(\mathbf{x}_i^T \mathbf{b})\} \end{aligned} \quad (2.66)$$

Model 4

When assuming latent cause, M_i , follow logarithmic distribution (logarithmic cure rate model (Model 4) from PS cure rate model), $S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}^{(0)})$ and $p_{0i}(\mathbf{x}_i; \boldsymbol{\xi}^{(0)})$ are as in (2.48) and (2.49). Also, b^* is given by

$$\begin{aligned} b^* &= \frac{\frac{\log(1 - \theta_{ij}S(t_i))}{S(c_i)\log(1 - \theta_i)} - \frac{-\theta_i}{\log(1 - \theta_i)}}{\frac{\log(1 - \theta_i S(c_i))}{S(c_i)\log(1 - \theta_i)}} = 1 + \frac{\theta_i S(c_i)}{\log(1 - \theta_i S(c_i))} \\ &= \frac{\exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{\left[1 + \exp(\mathbf{x}_i^T \mathbf{b})\right] \log \left[1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b}) S_0(c_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}\right]} \end{aligned} \quad (2.67)$$

Step 4: Maximization step (M-step)

We fill the censored data with the generated data from **Step 3**. Now, the improved estimate of $\boldsymbol{\xi}$ can be found using the pseudo-complete data as

$$\begin{aligned} \boldsymbol{\xi}^{(1)} &= (\mathbf{b}^{(1)'} , \boldsymbol{\beta}^{(1)'}, \mu^{(1)}, \sigma^{(1)}, \gamma^{(1)})' \\ &= \arg \max_{\boldsymbol{\xi}^{(1)}} \log L_c(\boldsymbol{\xi}; (\mathbf{t}, \mathbf{J}^*), (\mathbf{t}^*, \mathbf{J}^{**}), \mathbf{x}, \mathbf{z}), \end{aligned}$$

where $\mathbf{t}^*$ and $\mathbf{J}^{**}$ are vectors of t_i^* and J_i^{**} . The optimal value of $\boldsymbol{\xi}$ is obtained using the ‘L-BFGS-B’ package in R software where the algorithm is set to be converged when the desired tolerance level is achieved, i.e., $|\hat{\xi}_{r+1} - \hat{\xi}_r| < 10^{-6}$.

Step 5: Iterative step

Using the estimate $\hat{\boldsymbol{\xi}}^{(1)} = (\hat{\mathbf{b}}^{(1)'}, \hat{\boldsymbol{\beta}}^{(1)'}, \hat{\mu}^{(1)}, \hat{\sigma}^{(1)}, \hat{\gamma}^{(1)})'$, that we obtained in **Step 4**, we repeat **Step 2** to **Step 4** R times, to generate $\hat{\boldsymbol{\xi}}^{(r)} = (\hat{\mathbf{b}}^{(r)'}, \hat{\boldsymbol{\beta}}^{(r)'}, \hat{\mu}^{(r)}, \hat{\sigma}^{(r)}, \hat{\gamma}^{(r)})'$,

$r = 1, \dots, R$. The result, a sequence of estimates, is a Markov Chain; instead of converging to a single value, it converges to a stationary distribution under the standard conditions as discussed in Diebolt and Celeux (1993) and Diebolt and Ip (1995).

Step 6: Burn-in and MLE step

To obtain the stationary distribution, as mentioned in **Step 5**, we discard the first r^* (i.e., 200 and 500) iterations as a burn-in step, and compute the estimates by averaging every third of the remaining iterated estimates to avoid auto-correlation. By adopting the burn-in period, the random perturbations of the Markov chains preclude the influence of local maximum, thus making the estimates more reliable.

2.5 Simulation study

We adopt here the simulation setup described in Section 1.13. We consider two sample sizes, $n = (500, 1000)$, to mimic the data with moderate and large sample sizes. We further fix the X_{ij} to be a categorical variable that follow a Bernoulli distribution with success probability of 0.6, $\text{Ber}(p = 0.6)$.

In different settings, the competing cause, M_i , are assumed to follow Model 1, Model 2, Model 3, and Model 4, and baseline is set to follow Type I, Type II, and Type III of GEV distribution. We then fit all four models with the three choices of baseline to the simulated dataset, and assess the model performance.

When M_{ij} follow Bernoulli and Poisson, i.e., when we consider Bernoulli cure rate model (Model 1) and promotion time cure rate model (Model 2), we have

$$\theta_i = \exp(b_0 + b_1 * x_i), \quad (2.68)$$

and $i = 1, \dots, 500(1000)$. When M_{ij} follow geometric and logarithmic distributions, i.e., the PS cure rate model changes to a geometric cure rate model (Model 3) and a logarithmic cure rate model (Model 4), we have

$$\theta_i = \exp(b_0 + b_1 * x_i) / (1 + \exp(b_0 + b_1 * x_i)), \quad (2.69)$$

and $i = 1, \dots, 500(1000)$. Cure rates and survival probabilities for Model 1 to Model 4 are given in Table 2.6. The simulated dataset considers low to moderate level of censoring, high level of censoring and extra high level of censoring. In addition, we conduct the analysis on the true parameter settings that generate types of cure rate to be the low level cure rate, moderate level of cure rate, and high level cure rate.

Censored lifetime for individual i is generated from an exponential distribution with parameter cc , $C_i \sim \text{Exp}(cc)$. We generate Y_i from the quantile function of the GEV distribution with parameters μ , σ , and γ . If $Y_i \leq C_i$, we set $\delta_i = 1$, and $\delta_i = 0$ otherwise. We adjust the parameter of the exponential distribution, leading to different censored lifetimes and different censored proportions for the data. The number of iterations was set to 1200, with the first 200 as burn-in, and the spacing was fixed to 3 to avoid autocorrelation.

Table 2.5: The settings of the shape parameter, γ , considered for the baseline function

Baseline	γ			
Type I	0.000			
Type II	1.600	2.550		
Type III	-0.230	-1.100	-1.200	-1.500

Table 2.6: Some of the cure rates and levels of censoring for the simulated datasets

Censoring		Cure Rate
low to moderate	0.529 moderate	0.560
extra high	0.781 low	0.334
high	0.626 high	0.643
high	0.682 high	0.643
high	0.652 high	0.614
extra high	0.743 low	0.434
moderate	0.619 moderate	0.569
high	0.643 high	0.672
high	0.696 low to moderate	0.497

Table 2.7: The choice of true values for the model parameters

	b_0	b_1	β_1	μ	σ	γ (Type I, II, III)
TV	-0.700	0.520	0.640	0.200	0.170	0.000
TV	-0.600	0.150	0.300	0.100	0.150	1.600
TV	-0.890	0.530	0.875	0.140	0.440	-1.100
TV	0.400	-0.590	-0.550	0.100	0.100	0.000
TV	-0.450	0.770	0.400	0.350	0.900	2.550
TV	0.350	-0.640	-0.610	0.140	0.700	-1.500
TV	-0.930	0.880	1.300	0.290	0.350	0.000
TV	-0.880	0.880	0.820	0.130	0.450	-1.200
TV	-0.600	-0.300	0.400	0.100	0.150	1.600
TV	0.620	-0.700	-0.620	0.550	0.900	-0.230
TV	1.100	-0.300	-0.310	0.150	0.450	0.000

Results

The estimated means are close to the true values, indicating that the algorithm successfully converges to the data-generating parameters. When the candidate model fitted to the simulated data coincides with the data-generating model, the resulting bias and RMSE are low. This pattern is consistent across all models and aligns with the original simulation setup. Variations in the cure rate do not affect algorithm convergence or the recorded bias and RMSE. Coverage probabilities were calculated

using confidence intervals constructed from the standard errors derived from the Hessian matrix, with the interval containing the true value for each simulated dataset. With a nominal confidence level of 95%, the observed coverage probabilities are close to 95%.

When a candidate model differs from the data-generating model, the maximum likelihood estimates obtained via the stochastic EM algorithm do not necessarily converge to the true values, resulting in higher bias and RMSE. In these cases, the observed coverage probabilities are lower than the nominal 95% level. The true parameter values, along with the corresponding estimates, biases, and RMSEs obtained from fitting the true models, are presented in the following tables.

When we calculated the cure rate under Model 1, as shown in Table 2.10, the p-values of the difference between the true cure rate (0.611, 0.646) and estimated cure rate (0.628, 0.649) of our model under the low cure rate case are equal to 0.211 and 0.438, respectively. These correspond to the cases when the categorical variable, x_i , is equal to 1 and 0, respectively. As the obtained p-values are considered larger than 0.05, we do not see a statistically significant difference between the true cure rate and the estimated cure rate, and so we concluded that the difference observed is due to random chance. When the true model is the same as the fitted model, the results shown in Table 2.8 reveal that the true and estimated cure rates possess a small difference in values, and all cases share similar characteristics in Table 2.10. The difference of true cure and estimated cure rates result in p-values being larger than 0.05. In short, we conclude the model we have proposed has successfully revealed the true cure rate in the simulation study.

Table 2.8: Simulation results of estimated means, Bias, and RMSE, for different choices of baseline shape parameter, γ , and true model and fitted model are the same, based on 1200 iterations.

True: Bernoulli							
Fitted: Bernoulli							
Baseline: $\gamma = 0$							
n	Censoring (high)	Cure Rate (low)	Parameter	T.V.	Estimate	Bias	RMSE
1000	0.748	0.330	b_0	0.400	0.519	0.119	0.154
			b_1	-0.590	-0.540	0.050	0.098
			β_1	-0.550	-0.564	-0.014	0.112
			μ	0.100	0.072	-0.028	0.032
			σ	0.100	0.064	-0.036	0.039
			γ	0.000	-0.013	-0.013	0.014
True: Poisson							
Fitted: Poisson							
Baseline: $\gamma = 0$							
n	Censoring (high)	Cure Rate (low to mid)	Parameter	T.V.	Estimate	Bias	RMSE
1000	0.743	0.434	b_0	-0.700	-0.651	0.049	0.069
			b_1	0.520	0.536	0.016	0.042
			β_1	0.640	0.642	0.002	0.010
			μ	0.200	0.162	-0.038	0.050
			σ	0.170	0.152	-0.018	0.042
			γ	0.000	-0.032	-0.031	0.033
True: Bernoulli							
Fitted: Bernoulli							
Baseline: $\gamma = 2.55$							
n	Censoring (mid)	Cure Rate (mid)	Parameter	TV	Estimate	Bias	RMSE
1000	0.529	0.560	b_0	-0.450	-0.385	0.065	0.093
			b_1	0.770	0.844	0.074	0.107
			β_1	0.400	0.447	0.047	0.113
			μ	0.350	0.331	-0.019	0.053
			σ	0.900	0.812	-0.088	0.144
			γ	2.550	2.456	-0.094	0.152

True: Bernoulli							
Fitted: Bernoulli							
Baseline: $\gamma = -1.5$							
n	Censoring (high)	Cure Rate (low)	Parameter	TV	Estimate	Bias	RMSE
1000	0.781	0.334	b_0	0.350	0.336	-0.014	0.060
			b_1	-0.640	-0.693	-0.053	0.074
			β_1	-0.610	-0.633	-0.023	0.185
			μ	0.140	0.228	0.088	0.130
			σ	0.700	0.607	-0.093	0.105
			γ	-1.500	-1.514	-0.014	0.104
True: Poisson							
Fitted: Poisson							
Baseline: $\gamma = -1.1$							
n	Censoring (high)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
1000	0.652	0.6136	b_0	-0.890	-0.880	0.010	0.048
			b_1	0.530	0.567	0.037	0.043
			β_1	0.875	0.878	0.003	0.011
			μ	0.140	0.147	0.007	0.042
			σ	0.440	0.397	-0.043	0.068
			γ	-1.100	-1.119	-0.019	0.036
True: Geometric							
Fitted: Geometric							
Baseline: $\gamma = -1.2$							
n	Censoring (high)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
1000	0.626	0.643	b_0	-0.880	-0.868	0.012	0.119
			b_1	0.880	0.893	0.013	0.024
			β_1	0.820	0.865	0.045	0.061
			μ	0.130	0.143	0.013	0.051
			σ	0.450	0.431	-0.019	0.071
			γ	-1.200	-1.200	0.000	0.184

True: Geometric							
Fitted: Geometric							
Baseline: $\gamma = 0$							
n	Censoring (high)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
1000	0.643	0.672	b_0	-0.930	-0.952	-0.022	0.026
			b_1	0.880	0.881	0.001	0.017
			β_1	1.300	1.317	0.017	0.018
			μ	0.290	0.316	0.026	0.039
			σ	0.350	0.330	-0.020	0.041
			γ	0.000	-0.071	-0.071	0.305
True: Logarithmic							
Fitted: Logarithmic							
Baseline: $\gamma = -0.23$							
n	Censoring (high)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
1000	0.619	0.569	b_0	0.620	0.629	0.009	0.075
			b_1	-0.700	-0.706	-0.006	0.061
			β_1	-0.62	-0.630	-0.010	0.076
			μ	0.550	0.546	-0.004	0.040
			σ	0.900	0.913	0.013	0.093
			γ	-0.230	-0.217	0.013	0.029
True: Logarithmic							
Fitted: Logarithmic							
Baseline: $\gamma = 0$							
n	Censoring (high)	Cure Rate (moderate)	Parameter	TV	Estimate	Bias	RMSE
1000	0.696	0.497	b_0	1.100	1.196	0.096	0.103
			b_1	-0.300	-0.280	0.020	0.035
			β_1	-0.310	-0.297	0.013	0.024
			μ	0.150	0.215	0.065	0.074
			σ	0.450	0.424	-0.026	0.027
			γ	0.000	-0.078	-0.078	0.085

True: Bernoulli							
Fitted: Bernoulli							
Baseline: $\gamma = 2.55$							
n	Censoring (moderate)	Cure Rate (moderate)	Parameter	TV	Estimate	Bias	RMSE
500	0.588	0.535	b_0	-0.450	-0.430	0.020	0.088
			b_1	0.770	0.653	-0.117	0.180
			β_1	0.400	0.388	-0.012	0.137
			μ	0.350	0.317	-0.033	0.071
			σ	0.900	0.808	-0.092	0.191
			γ	2.550	2.536	-0.014	0.131
True: Poisson							
Fitted: Poisson							
Baseline: $\gamma = -1.1$							
n	Censoring (high)	Cure Rate (moderate)	Parameter	TV	Estimate	Bias	RMSE
500	0.686	0.589	b_0	-0.890	-0.864	0.026	0.193
			b_1	0.530	0.522	-0.008	0.043
			β_1	0.875	0.881	0.006	0.062
			μ	0.140	0.154	0.014	0.036
			σ	0.440	0.322	-0.118	0.135
			γ	-1.100	-1.025	0.075	0.300
True: Geometric							
Fitted: Geometric							
Baseline: $\gamma = 0$							
n	Censoring (moderate)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
500	0.626	0.683	b_0	-0.930	-0.950	-0.020	0.025
			b_1	0.880	0.888	0.008	0.010
			β_1	1.300	1.316	0.016	0.017
			μ	0.290	0.323	0.033	0.050
			σ	0.350	0.357	0.007	0.048
			γ	0.000	-0.075	-0.075	0.078

True: Logarithmic							
Fitted: Logarithmic							
Baseline: $\gamma = 0$							
n	Censoring (high)	Cure Rate (high)	Parameter	TV	Estimate	Bias	RMSE
500	0.626	0.683	b_0	0.620	0.642	0.022	0.129
			b_1	-0.700	-0.707	-0.007	0.089
			β_1	-0.620	-0.629	-0.009	0.132
			μ	0.550	0.546	-0.004	0.034
			σ	0.900	0.912	0.012	0.074
			γ	-0.230	-0.212	0.018	0.027

Table 2.9: Simulation results of mean estimates, bias, and root mean square error (RMSE) based on 1200 iterations. ($n = 1000$, censored proportion = (Case 1: 0.68, Case 2: 0.73), moderate and high cure rate = (0.611, 0.711))

True M_{ij} : Geometric								
Fitted M_{ij}	n	Censoring	Cure Rate	Par.	T.V.	Estimate	Bias	RMSE
Geometric	1000	Case 1	moderate	b_0	-0.600	-0.526	0.074	0.129
				b_1	0.150	0.127	-0.023	0.066
				β_1	0.300	0.277	-0.023	0.280
				μ	0.100	0.053	-0.047	0.048
				σ	0.150	0.093	-0.057	0.060
				γ	1.600	1.660	0.060	0.130
Geometric	1000	Case 2	High	b_0	-0.600	-0.619	-0.019	0.040
				b_1	-0.300	-0.254	0.046	0.064
				β_1	0.400	0.372	-0.028	0.151
				μ	0.100	0.051	-0.049	0.049
				σ	0.150	0.089	-0.061	0.062
				γ	1.600	1.661	0.061	0.096

Table 2.10: Cure Probabilities: $n = 1000$

Geometric		
censored:0.682, lower cure rate		
Cure Rate	$x_i = 1$	$x_i = 0$
True	0.61064	0.64566
Estimated	0.62803	0.64897

2.6 Analysis of smoking cessation data

Signs of cure fraction for b_1 , b_2 , b_3 , and b_4 , corresponding to gender, duration, treatment, and consumption, are the same for all selected models except the intercept, b_0 . Estimated empirical means of $\mathbf{b}$ have the lowest standard errors under the promotion time cure model (Model 2), where $\mathbf{b}$ is related to the cure fraction. We may therefore obtain the inferences on the cure rate via survival plots as shown in Figures 2.4 and 2.5.

Signs of β_2 and β_4 , covariates for the survival function, are the same for all selected models. β_1 has a negative sign under the promotion time cure model (Model 2) and logarithmic cure model (Model 4), and a positive sign under others. β_3 is negative under Bernoulli (Model 1), and positive under others. Estimate means of β have the lowest standard errors under the promotion time cure model (Model 2). The means of parameters for baseline distribution μ (mean), σ (scale), γ (location), are all close values for all the models.

The $\mathbf{b}$ parameters are related to the cure fraction and so we may obtain inferences on the cure rate. In Figure 2.4, for a patient who has smoking duration of 30 years, and consumption of 6 cigarettes per day, the survival function of a male patient with

smoking intervention (blue curve) has lower probability of survival than a female patient (red curve) who received smoking intervention and same smoking habits. We may conclude that for this particular population where smokers received smoking intervention, the women smokers have higher probability of quitting than men smokers. The patient who received smoking intervention, i.e., treatment, has higher probability of quitting than the one who received usual care, as shows in Figure 2.5. The smoker who has higher level of cigarette consumption has lower probability of quitting than the ones who smoke less.

Recalling that the confidence interval is calculated using the formula $(\hat{\xi} - z_{1-\delta/2}se(\hat{\xi}), \hat{\xi} + z_{1-\delta/2}se(\hat{\xi}))$, we calculated the 95% confidence interval and have presented them in Table 2.11 and 2.12.

Table 2.13 presents negative log-likelihood, AIC, BIC, and AICc values for Models 1 to 4. The promotion time cure model (Model 2) is seen to give the lowest AIC, BIC, AICc values among all.

Based on all the results, we observe that the model that has the lowest AIC and BIC scores, among all the models considered for the smoking cessation data, is the promotion time cure rate model (Model 2) with Type II of GEV as baseline. We visualize the difference of the survival function based on the optimized estimates of the parameters for Model 2 in Figures 2.4 and 2.5. Figure 2.6 is an example of the parameter evolution graphs for the 11 parameters of Model 2 and it shows that the estimates of the model parameters oscillate around the mean after the burn-in period of 500. Thus, we note that our algorithm has converged. (Additional graphs are presented in Figure A.18 in Appendix A.)

Finally, the cure rates and survival probabilities of all 223 patients are predicted,

and the results have been visualized using maps in Figure 2.3. The darker pink/purple colours are associated with high cure rates and survival probabilities, and the lighter pink colour indicates lower cure rates and lower survival probabilities. The patients are stratified by the zip code they belong to. For regions with more than one patients, we use the average cure rate of the region and the average survival probability of the region for demonstration and colouring purposes. The maps indicate patients from unlike zip codes have cure rate and survival probabilities that fluctuate. For these reasons, in the following Chapters, we introduce spatial frailties to our models, and so spatial survival analysis is designed and conducted in the subsequent chapters.

Figure 2.3: Maps of cure rate and survival probabilities stratified by zip codes.

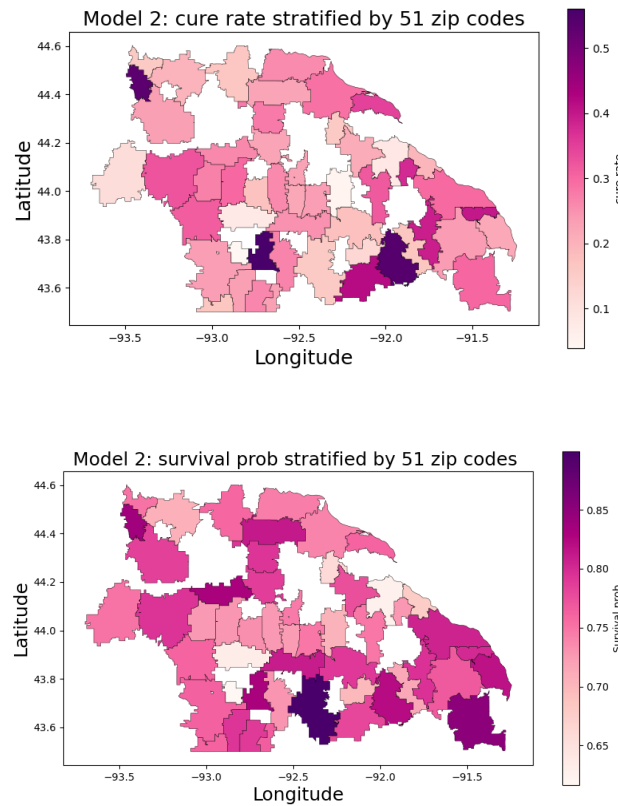


Table 2.11: Estimated mean, standard error, and 95% CI when competing causes follow Bernoulli and Poisson (Model 1 and Model 2) for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Model 1				
Parameter	Mean	SE	2.50%	97.50%
b_0 (Intercept)	-0.373	0.788	-1.917	1.171
b_1 (Gender)	0.681	0.313	0.067	1.295
b_2 (Duration)	-0.015	0.021	-0.057	0.027
b_3 (Treatment)	-0.567	0.347	-1.248	0.115
b_4 (Consumption)	0.066	0.016	0.036	0.097
β_1 (Gender)	0.321	0.261	-0.190	0.833
β_2 (Duration)	-0.001	0.012	-0.024	0.021
β_3 (Treatment)	-0.179	0.280	-0.728	0.371
β_4 (Consumption)	0.000	0.010	-0.020	0.020
μ	3.234	0.665	1.931	4.537
σ	6.982	2.058	2.948	11.015
γ	2.805	0.232	2.351	3.259
Model 2				
Parameter	Mean	SE	2.50%	97.50%
b_0 (Intercept)	-0.187	0.446	-1.062	0.687
b_1 (Gender)	0.684	0.202	0.288	1.080
b_2 (Duration)	-0.014	0.012	-0.038	0.011
b_3 (Treatment)	-0.433	0.211	-0.846	-0.020
b_4 (Consumption)	0.038	0.009	0.021	0.056
β_1 (Gender)	-0.252	0.305	-0.850	0.346
β_2 (Duration)	-0.002	0.013	-0.027	0.023
β_3 (Treatment)	0.054	0.324	-0.582	0.690
β_4 (Consumption)	-0.025	0.013	-0.050	-0.001
μ	3.274	0.659	1.982	4.567
σ	7.038	2.042	3.034	11.041
γ	2.777	0.231	2.324	3.230

Table 2.12: Estimated mean, standard error, and 95% CI when M_{ij} follow geometric and logarithmic models (Model 3 and Model 4) for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Model 3				
Parameter	Mean	SE	2.50%	97.50%
b_0 (Intercept)	0.670	0.666	-0.636	1.976
b_1 (Gender)	0.555	0.301	-0.034	1.144
b_2 (Duration)	-0.025	0.019	-0.061	0.012
b_3 (Treatment)	-0.547	0.351	-1.235	0.141
b_4 (Consumption)	0.024	0.013	0.000	0.049
β_1 (Gender)	0.019	0.345	-0.657	0.694
β_2 (Duration)	-0.009	0.015	-0.039	0.021
β_3 (Treatment)	0.046	0.388	-0.714	0.807
β_4 (Consumption)	-0.008	0.014	-0.035	0.019
μ	3.900	0.936	2.066	5.735
σ	8.408	2.865	2.794	14.023
γ	2.656	0.214	2.237	3.076
Model 4				
Parameter	Mean	SE	2.50%	97.50%
b_0 (Intercept)	2.072	1.250	-0.378	4.522
b_1 (Gender)	0.770	0.551	-0.310	1.850
b_2 (Duration)	-0.069	0.036	-0.140	0.002
b_3 (Treatment)	-0.815	0.641	-2.072	0.441
b_4 (Consumption)	0.081	0.025	0.032	0.130
β_1 (Gender)	-0.015	0.396	-0.791	0.761
β_2 (Duration)	-0.003	0.019	-0.040	0.033
β_3 (Treatment)	0.135	0.459	-0.764	1.034
β_4 (Consumption)	-0.019	0.018	-0.054	0.016
μ	4.198	1.352	1.548	6.849
σ	7.683	4.031	-0.219	15.584
γ	2.151	0.212	1.735	2.567

Table 2.13: Negative log-likelihood values, AIC, BIC, AICc for Model 1, Model 2, Model 3, and Model 4 for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Activation	Distribution of M_{ij}	-ll	AIC	BIC	AICc
Random	Model 1	295.368	614.735	655.621	616.221
First	Model 2	292.638	609.277	650.163	610.762
	Model 3	296.250	616.500	657.386	617.985
	Model 4	294.602	613.205	654.091	614.690

Survival plots

Figure 2.4: Survival Plots (C.P.): Female vs Male w/ SI

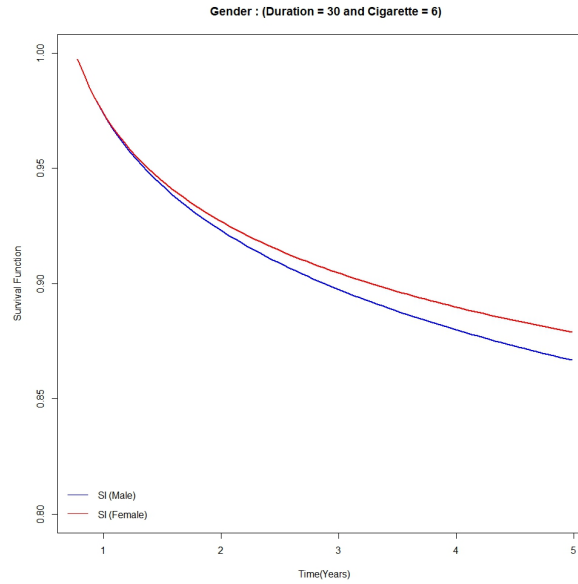


Figure 2.5: Treatment: Female, Consump:6, Duration: 30

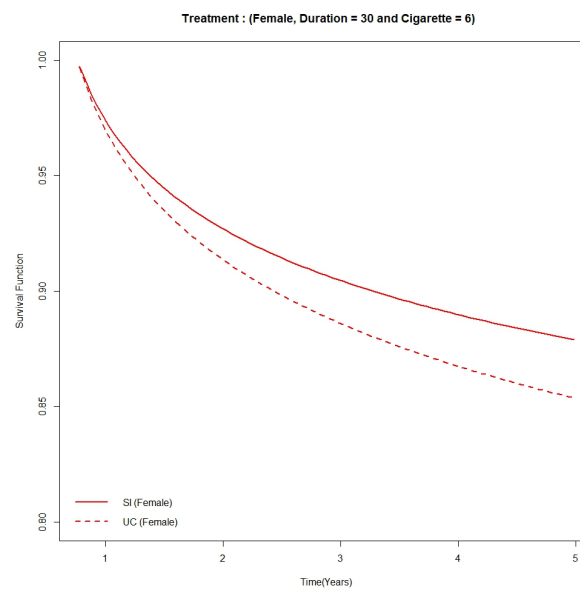
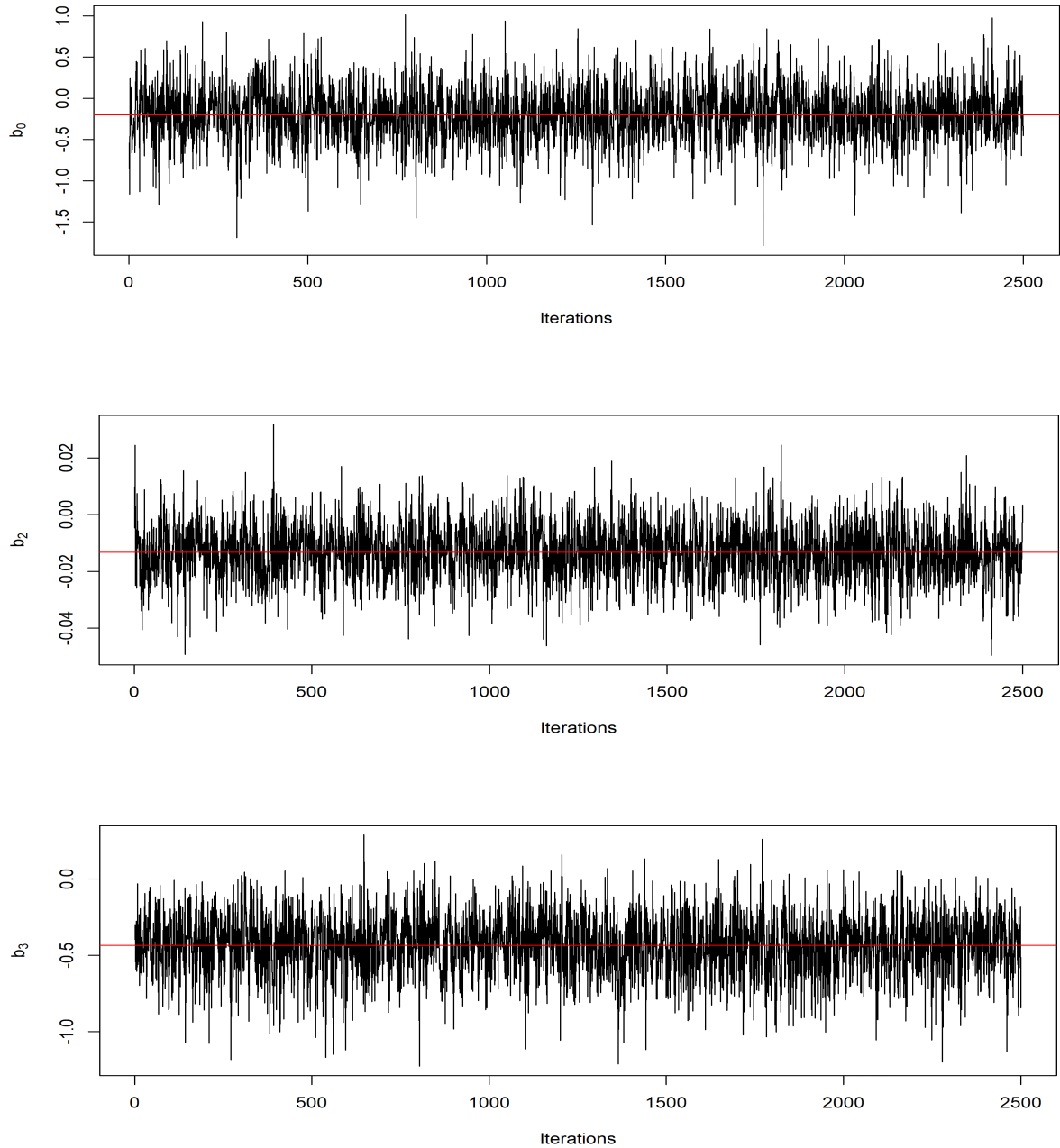


Figure 2.6: Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm when M_{ij} follow Model 2 for the smoking cessation dataset. (2500 iterations after burn-in period of 500)



Chapter 3

Stochastic EM-based Likelihood Inference for First Activation Scheme of PS Cure Rate Model with Gaussian Spatial Frailties

3.1 Introduction

In this chapter, we consider spatial component that exists in the data, and reveal the impact of location of a patient on the occurrence of event of interest. We will thus extend survival analysis carried out in the last chapter to spatial survival analysis. We incorporate spatial frailties into cure rate models, assuming that the event of interest occurs when the first possible competing cause is present (first activation scheme) and when some competing causes are present (random activation scheme).

Recall that we assume in the i th region ($i = 1, \dots, I$), there are n_i number of subjects indexed by $j = 1, \dots, n_i$. The observed event time for the (i, j) th subject is denoted by T_{ij} .

The competing cause M_{ij} is assumed to follow power series cure rate model with p.m.f. as in (1.3). Given that the number of competing causes for the (i, j) th individual is $M_{ij} = m$, let random variable W_{k^*ij} be the time-to-event due to the k^* th latent risk, with distribution function $F(\cdot) = 1 - S(\cdot)$ and survival function $S(\cdot)$, for $k^* = 1, 2, \dots, m$.

The reality is that the competing cause M_{ij} and the lifetime related to a specific cause, W_{kij} , are not observable. Under the first activation scheme, assuming that any competing cause will eventually trigger the event, the time to the event of interest is denoted by the random variable $T_{ij} = \min\{W_{1ij}, \dots, W_{m_{ij}}\}$ for $M_{ij} \geq 1$, and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P[Y_{ij} = \infty \mid M_{ij} = 0] = 1$. Under the random activation scheme, the event occurs when some possible competing causes are presented, given the number of competing causes M_{ij} , the number of activated competing causes is then a random variable with the discrete uniform distribution on $\{1, \dots, M_{ij}\}$.

The spatial frailties are constructed by using the geospatial information (longitude and latitude) and the Gaussian process. In Section 3.2, the Gaussian spatial effects are added to cure rate models, which results in Bernoulli mixture cure rate model with spatial frailties (Model 5), promotion time cure rate model with spatial frailties (Model 6), geometric cure rate model with spatial frailties (Model 7), and logarithmic cure rate model with spatial frailties (Model 8). The details of the developed models are presented in Section 3.2. The steps for the derivation of the log-likelihood function and the estimation technique using SEM algorithm are developed and presented in

Sections 3.3 and 3.4, respectively. In Section 3.5, an extensive simulation study is conducted with consideration to different sample sizes and parameter settings for evaluating the performance of Models 5 - 8. Model selection and model discrimination are performed and the obtained results are presented in Section 3.5. In Section 3.6, the proposed models are illustrated with a real life dataset. The effect of spatial frailties are captured by the proposed models, and the different spatial effects on cure rate and survival probabilities are demonstrated using heat maps. The obtained results are visualized by survival plots as well. The parameter evolution plots for the our model parameters are provided for ensuring the convergence of the SEM algorithm. In addition, the results of the likelihood ratio test do confirm the necessity of adding spatial effect to the proposed cure rate models.

3.2 Cure models

As discussed in Section 1.10, the spatial component is added to our model in linear combination shown in Eqs. (1.8) and (1.9), where $\phi_{ij}^* = \mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i$ and $\alpha_{ij}^* = \mathbf{x}_{ij}^T \mathbf{b} + u_i$, for $i = 1, \dots, I$, $j = 1, \dots, n_i$.

When the competing causes M_{ij} follow Bernoulli or Poisson distributions, the PS cure rate model reduces to a Bernoulli mixture model or a promotion time cure rate model, respectively, with the power parameter given by $\theta_{ij} = \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)$.

Bernoulli cure rate with spatial frailties (Model 5)

In the case where we assume that the latent competing causes M_{ij} follow a Bernoulli distribution, a Bernoulli mixture cure model with spatial frailties (Model 5) is obtained from the PS cure rate model, with the spatial effect modeled through a Gaussian process. In this setting, the series function is $G(\theta_{ij}) = 1 + \theta_{ij}$. From (1.4), the mass function of M_{ij} takes the form

$$p_{m_{ij}} = P(M_{ij} = m_{ij}) = \left(\frac{1}{1 + \theta_{ij}} \right)^{1-m_{ij}} \left(\frac{\theta_{ij}}{1 + \theta_{ij}} \right)^{m_{ij}}, \quad (3.70)$$

for $\theta_{ij} > 0$, $m_{ij} = 0, 1$. The cure probability is

$$p_{0ij} = P(M_{ij} = 0) = \frac{1}{1 + \theta_{ij}}. \quad (3.71)$$

The population survival function and probability density function for the population are

$$\begin{aligned} S_p(t_{ij}) &= \sum_{m_{ij}=0}^1 \left(\frac{1}{1 + \theta_{ij}} \right)^{1-m_{ij}} \left(\frac{\theta_{ij}}{1 + \theta_{ij}} \right)^{m_{ij}} S(t_{ij})^{m_{ij}} \\ &= p_{0ij} + (1 - p_{0ij})S(t_{ij}), \end{aligned} \quad (3.72)$$

$$f_p(t_{ij}) = \frac{\theta_{ij}}{(1 + \theta_{ij})} f(t_{ij}), \quad (3.73)$$

respectively.

Poisson cure rate with spatial frailties (Model 6)

When M_{ij} follow a Poisson distribution, the promotion time cure model with spatial frailties (Model 6) is obtained from the PS cure rate model, with the spatial effect incorporated through a Gaussian process. In this case, the series function has the form $G(\theta_{ij}) = \exp(\theta_{ij})$. From (1.4), the mass function of M_{ij} is

$$p_{m_{ij}} = P(M_{ij} = m_{ij}) = \frac{\theta_{ij}^{m_{ij}}}{m_{ij}! \exp(\theta_{ij})}, \quad (3.74)$$

for $\theta_{ij} > 0$, $m_{ij} = 0, 1, \dots$. The cure probability is

$$p_{0ij} = P(M_{ij} = 0) = \exp(-\theta_{ij}). \quad (3.75)$$

The population survival function and probability density function are given by

$$\begin{aligned} S_p(t_{ij}) &= \sum_{m_{ij}=0}^{\infty} \frac{\theta_{ij}^{m_{ij}}}{m_{ij}! \exp(\theta_{ij})} S(t_{ij})^{m_{ij}} \\ &= \exp(-\theta_{ij}) \sum_{m_{ij}=0}^{\infty} \frac{[\theta_{ij} S(t_{ij})]^{m_{ij}}}{m_{ij}!} \\ &= \exp(-\theta_{ij}) \exp(\theta_{ij} S(t_{ij})) \\ &= \exp(-\theta_{ij} F(t_{ij})), \end{aligned} \quad (3.76)$$

$$f_p(t_{ij}) = \theta_{ij} f(t_{ij}) \exp(-\theta_{ij} F(t_{ij})), \quad (3.77)$$

respectively, where $f(t_{ij}) = \lambda(t_{ij}) * S(t_{ij})$.

When the PS cure rate model changes to a geometric cure rate model and a

logarithmic cure rate model, we have the corresponding power parameter to be

$$\theta_{ij} = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \quad (3.78)$$

Geometric cure rate with spatial frailties (Model 7)

When M_{ij} follow a geometric distribution, we obtain the geometric cure rate model with spatial frailties (Model 7). The series function in this case has the form $G(\theta_{ij}) = \frac{1}{1-\theta_{ij}}$. From (1.4), the mass function of M_{ij} is

$$p_{m_{ij}} = P(M_{ij} = m_{ij}) = (1 - \theta_{ij})\theta_{ij}^{m_{ij}}, \quad (3.79)$$

for $0 < \theta_{ij} < 1$, $m_{ij} = 0, 1, \dots$. The cure probability is

$$p_{0ij} = P(M_{ij} = 0) = 1 - \theta_{ij}. \quad (3.80)$$

The population survival function and probability density function are

$$\begin{aligned} S_p(t_{ij}) &= \sum_{m_{ij}=0}^{\infty} (1 - \theta_{ij})\theta_{ij}^{m_{ij}} S(t_{ij})^{m_{ij}} \\ &= \frac{1 - \theta_{ij}}{1 - \theta_{ij}S(t_{ij})}, \end{aligned} \quad (3.81)$$

$$f_p(t_{ij}) = \theta_{ij}(1 - \theta_{ij})f(t_{ij})[1 - \theta_{ij}S(t_{ij})]^{-2}, \quad (3.82)$$

respectively.

Logarithmic cure rate with spatial frailties (Model 8)

Similarly, when M_{ij} follow a logarithmic distribution, the logarithmic cure rate model with spatial frailties (Model 8) is obtained, with the series function, cure probability, survival function, and probability density function is given by $G(\theta_{ij}) = \frac{-\log(1-\theta_{ij})}{\theta_{ij}}$,

$$p_{0ij} = \frac{-\theta_{ij}}{\log(1 - \theta_{ij})} \quad (3.83)$$

$$S_p(c_{ij}) = \frac{\log(1 - \theta_{ij}S(c_{ij}))}{S(c_{ij})\log(1 - \theta_{ij})}, \quad (3.84)$$

$$f_p(c_{ij}) = -\frac{f(c_{ij})}{S(c_{ij})\log(1 - \theta_{ij})} \left[\frac{\log(1 - \theta_{ij}S(c_{ij}))}{S(c_{ij})} + \frac{\theta_{ij}}{1 - \theta_{ij}S(c_{ij})} \right]. \quad (3.85)$$

3.3 The log-likelihood function

3.3.1 Bernoulli mixture cure rate with spatial frailties (Model 5)

When we incorporate spatial frailties, the complete log-likelihood function under the Bernoulli mixture cure rate model (with M_{ij} following a Bernoulli distribution) is given by

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))} \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{3.86}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$, λ_0 is the baseline hazard function as in (1.15), S_0 is the baseline survival function as in (1.14), u_i is the Gaussian spatial frailty corresponding to location i , and f_U is the density function of spatial component as in (1.7).

Recall that the baseline survival function and baseline hazard function has two forms depending on parameter γ . When $\gamma = 0$, the log-likelihood function is given

by

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))} \frac{1}{\sigma t_{ij}} \exp\left(-\frac{\log(t_{ij}) - \mu}{\sigma}\right) \right. \\
&\quad \times \left[\exp \left\{ \exp\left(-\frac{\log(t_{ij}) - \mu}{\sigma}\right) \right\} - 1 \right]^{-1} \\
&\quad \times \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \left[1 - \exp \left\{ - \exp\left(\frac{\log(t_{ij}) - \mu}{\sigma}\right) \right\} \right]^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta})} \Big\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \right. \\
&\quad \times \left. \left[1 - \exp \left\{ - \exp\left(\frac{\log(t_{ij}) - \mu}{\sigma}\right) \right\} \right]^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{3.87}$$

When $\gamma \neq 0$, the log-likelihood function has a general form

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))} \frac{1}{\sigma t_{ij}} \left(1 + \gamma \frac{\log(t_{ij}) - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}-1} \right. \\
&\quad \times \left[\exp \left\{ \left(1 + \gamma \frac{\log(t_{ij}) - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}} \right\} - 1 \right]^{-1} \\
&\quad \times \exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i) \left[1 - \exp \left\{ - \left(1 + \gamma \frac{\log(t_{ij}) - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}} \right\} \right]^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \Big\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right) \right. \\
&\quad \times \left[1 - \exp \left\{ - \left(1 + \gamma \frac{\log(t_{ij}) - \mu}{\sigma}\right)_+^{\frac{-1}{\gamma}} \right\} \right]^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \Big\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{3.88}$$

which can be further written into cases with $\gamma > 0$ and $\gamma < 0$ corresponding to Fréchet and Weibull baseline distributions, respectively.

3.3.2 Promotion time cure rate Poisson model with spatial frailties (Model 6)

When M_{ij} follows a Poisson distribution and the the event of interest occurs when the first competing cause presents, when combining the Eqs (2.41), (2.42) and 2.43, the complete log-likelihood function is obtained as

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\
&\quad \left. \times \exp[-\exp(\mathbf{x}_i^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{ \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_i^* \log \left\{ \exp[-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})] \right. \\
&\quad \left. - \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{3.89}$$

where the power parameter $\theta_{ij} > 0$.

3.3.3 Geometric cure rate model with spatial frailties (Model 7)

When M_{ij} follows a geometric distribution, the complete log-likelihood function is given by

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_i) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \times S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right]^{-2} \Big\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} J_{ij}^* \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} - \left[\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right] \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{3.90}$$

where the power parameter $0 < \theta_{ij} < 1$.

3.3.4 Logarithmic cure rate model with spatial frailties (Model 8)

When M_{ij} follow a logarithmic distribution, the logarithmic cure rate model with spatial frailties (Model 8) is obtained. In this case, the complete log-likelihood function

is given by

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} \right] \Bigg\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} J_{ij}^* \log \left\{ \frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} - \left(\frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right) \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{3.91}$$

where the power parameter $0 < \theta_{ij} < 1$.

All pertinent details for the above cases are provided in Appendix B.

Though the complete log-likelihood functions for the proposed cure rate models are presented, we do not have complete information for individuals who are right censored. The survival status remains unknown for censored subjects since they can be either cured or susceptible. We overcome this difficulty by implementing the stochastic step (S-step), explained in the next section.

3.4 Stochastic EM

Recall that for the j th individual in the i th region, if the individual's lifetime is observed, we record the lifetime t_{ij} and set the censoring indicator $\delta_{ij} = 1$, which implies that the subject is susceptible and thus the cure status is $J_{ij}^* = 1$. If the lifetime is unobserved, the individual may or may not be cured from the event of interest. In this case, we have $\delta_{ij} = 0$ and $t_{ij} = c_{ij}$, so the lifetime status of the subject is unknown. Consequently, there are two possible outcomes: $J_{ij}^* = 0$ if the individual is cured, or $J_{ij}^* = 1$ if the individual is uncured.

Steps for Stochastic EM algorithm

Step 1

We initialize the parameter $\xi^{(0)} = (\mathbf{b}^{(0)'}, \boldsymbol{\beta}^{(0)'}, \mu^{(0)}, \sigma^{(0)}, \gamma^{(0)}, \boldsymbol{\theta}_s^{(0)'})'$ using the grid search method. The set of parameter values in the parameter space that maximize the complete log-likelihood function is set to be the initial value, $\xi^{(0)}$, for the Stochastic EM algorithm.

Step 2: Stochastic step (S-step)

Recall that for censored subject ij , we label their censored lifetime by $\delta_{ij} = 0$. Among them, the indicator of the cure status, J_{ij}^* , is obtained from a Bernoulli distribution

with probability computed from (1.21), and we can express it as

$$\begin{aligned}
p_{0ij}^{(0)} &= P[J_{ij}^* = 1 | T_{ij} > c_{ij}; \boldsymbol{\xi}^{(0)}] = \frac{P[T_{ij} > c_{ij} | J_{ij}^* = 1] P[J_{ij}^* = 1]}{P[T_{ij} > c_{ij}]} \Big|_{\boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}} \\
&= \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})} = \frac{p_{0ij} + (1 - p_{0ij})S(c_{ij}) - p_{0ij}}{p_{0ij} + (1 - p_{0ij})S(c_{ij})} \\
&= \frac{(1 - p_{0ij})S(c_{ij})}{p_{0ij} + (1 - p_{0ij})S(c_{ij})} = \frac{\theta_{ij}S(c_{ij})/(1 + \theta_{ij})}{(1 + \theta_{ij}S(c_{ij}))/ (1 + \theta_{ij})} = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}},
\end{aligned}$$

where p_{0ij} and S_p are cure probabilities and survival function from (3.72), S_0 is the baseline survival function, and u_i is spatial frailty.

Now, suppose the computed cure status for the first round of censored subject ij is denoted by $\pi_{ij}^{(0)}$. We then replace J_{ij}^* in the complete log-likelihood functions in (3.86), (3.89), (3.90), and (3.91) when assuming M_{ij} follow Bernoulli, Poisson, geometric, and logarithmic distributions, respectively. Now, combining the Gaussian spatial effect, their associated cure rate models are given by Model 5 - 8, respectively.

Model 5

J_{ij}^* in (3.86) is replaced by $\pi_{ij}^{(0)}$. The complete log-likelihood function of the Bernoulli mixture cure rate along with spatial frailties (Model 5) can be rewritten as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_i=1} \sum_{\delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i))} \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} \sum_{\delta_i=0} (1 - \pi_{ij}^{(0)}) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} \sum_{\delta_i=0} \pi_{ij}^{(0)} \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i)} \right) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{3.92}$$

Model 6

J_{ij}^* in (3.89) are replaced by $\pi_{ij}^{(0)}$. The complete log-likelihood function of promotion time cure rate along with spatial frailties (Model 6) can be rewritten as

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\
&\quad \left. \times \exp \left[-\exp(\mathbf{x}_i^T \mathbf{b}) (1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}) \right] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - \pi_{ij}^{(0)}) \log \{ \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \pi_{ij}^{(0)} \log \left\{ \exp \left[-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) (1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}) \right] \right. \\
&\quad \left. - \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{3.93}$$

Model 7

J_{ij}^* in (3.90) is replaced by $\pi_{ij}^{(0)}$. The complete log-likelihood function of the geometric cure rate along with spatial frailties (Model 7) can be rewritten as

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_i) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \times S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right]^{-2} \Big\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum (1 - \pi_{ij}^{(0)}) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum \pi_{ij}^{(0)} \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} - \left[\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right] \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{3.94}$$

Model 8

J_{ij}^* in (3.91) is replaced by $\pi_{ij}^{(0)}$. The complete log-likelihood function of the logarithmic cure rate along with spatial frailties (Model 8) can be rewritten as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}) S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right] \Bigg\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - \pi_{ij}^{(0)}) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} \pi_{ij}^{(0)} \log \left\{ \frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}) S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} - \left(\frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right) \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{3.95}$$

Step 3

If the censored subject is susceptible, i.e., $\pi_{ij}^{(0)} = 1$, the complete lifetime y_{ij}^* is computed from the truncated distribution with density function

$$f_T(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = \frac{f_p(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}, \tag{3.96}$$

where $c_{ij} < y_{ij}^* < \infty$, and the cdf is

$$F_T(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) - S_p(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})}, \quad (3.97)$$

where $c_{ij} < y_{ij}^* < \infty$. Eq. (3.97) is not a proper cdf since

$$\lim_{y_{ij}^* \rightarrow \infty} F_T(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})} \neq 1. \quad (3.98)$$

The cured/immunized subject with $\pi_{ij}^{(0)} = 0$ is treated as long term survivor and its lifetime is infinite with respect to the event of interest. Hence, it takes the form of $\lim_{y_{ij}^* \rightarrow \infty} S_p(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) = p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$. To generate y_{ij}^* from (3.96) under the susceptible scenario, we adopt inverse transformation sampling technique. It is easy to show that $F_T(y_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ follows an $\text{Uniform}(a^*, b^*)$ distribution with parameters $a^* = 0$ and $b^* = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})}$. $S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ and $p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ have different forms when we consider Bernoulli mixture cure rate with spatial frailties (Model 5), Poisson cure rate with spatial frailties (Model 6), Geometric cure rate with spatial frailties (Model 7), and Logarithmic cure rate with spatial frailties (Model 8). The form of b^* takes different forms under different assumptions for the latent cause, M_{ij} .

Model 5

For the Bernoulli mixture cure rate along with spatial frailties (Model 5), $S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ and $p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ are replaced by (3.72) and (3.71). In this case, b^* is given by

$$b^* = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}. \quad (3.99)$$

Model 6

For the promotion time cure rate model with spatial frailties (Model 6), $S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ and $p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ are replaced by (3.76) and (3.75). In this case, b^* is given by

$$\begin{aligned} b^* &= \frac{\exp(-\theta_{ij}) \exp(\theta_{ij} S(c_{ij})) - \exp(-\theta_{ij})}{\exp(-\theta_{ij}) \exp(\theta_{ij} S(c_{ij}))} = \frac{\exp(\theta_{ij} S(c_{ij})) - 1}{\exp(\theta_{ij} S(c_{ij}))} \\ &= 1 - \exp\{-\theta_{ij} S(c_{ij})\} = 1 - \exp\{-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}\}. \end{aligned} \quad (3.100)$$

Model 7

Under the geometric mixture cure rate with spatial frailties (Model 7), $S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ and $p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ are replaced by (3.81) and (3.80). In this case, b^* is given by

$$\begin{aligned} b^* &= \frac{\frac{1-\theta_{ij}}{1-\theta_{ij} S(c_{ij})} - (1-\theta_{ij})}{\frac{1-\theta_{ij}}{1-\theta_{ij} S(c_{ij})}} = \theta_{ij} S(c_{ij}) \\ &= \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} / \{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)\}. \end{aligned} \quad (3.101)$$

Model 8

When assuming latent cause, M_{ij} , to follow logarithmic, we obtain the logarithmic cure rate model with spatial effect (Model 8). In this case, $S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ and

$p_{0ij}(\mathbf{x}_{ij}|u_i; \boldsymbol{\xi}^{(0)})$ are replaced by (3.84) and (3.83), and the b^* is given by

$$\begin{aligned} b^* &= \frac{\frac{\log(1 - \theta_{ij}S(t_{ij}))}{S(c_{ij})\log(1 - \theta_{ij})} - \frac{-\theta_{ij}}{\log(1 - \theta_{ij})}}{\frac{\log(1 - \theta_{ij}S(c_{ij}))}{S(c_{ij})\log(1 - \theta_{ij})}} = 1 + \frac{\theta_{ij}S(c_{ij})}{\log(1 - \theta_{ij}S(c_{ij}))} \\ &= \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{\left[1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)\right] \log \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(c_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right]}. \end{aligned} \quad (3.102)$$

Step 4: Maximization step (M-step)

We fill the censored data with the generated data from Step 3. Now, the improved estimate of $\boldsymbol{\xi}$ can be found using the pseudo-complete data as

$$\begin{aligned} \boldsymbol{\xi}^{(1)} &= (\mathbf{b}^{(1)'}, \boldsymbol{\beta}^{(1)'}, \mu^{(1)}, \sigma^{(1)}, \gamma^{(1)}, \boldsymbol{\theta}_s^{(1)'})' \\ &= \arg \max_{\boldsymbol{\xi}^{(1)}} \log L_c(\boldsymbol{\xi}; (\mathbf{t}, \mathbf{J}^*), (\mathbf{t}^*, \mathbf{J}^{**}), \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s), \end{aligned}$$

where $\mathbf{t}^*$ and $\mathbf{J}^{**}$ are vectors of t_{ij}^* and J_{ij}^{**} , respectively. The optimal value of $\boldsymbol{\xi}$ is obtained using the ‘L-BFGS-B’ package in R software, where the algorithm is set to be converged when the desired tolerance level, i.e., $|\hat{\xi}_{r+1} - \hat{\xi}_r| < 10^{-6}$, is achieved.

Step 5: Iterative step

Using the estimate $\hat{\boldsymbol{\xi}}^{(1)} = (\hat{\mathbf{b}}^{(1)'}, \hat{\boldsymbol{\beta}}^{(1)'}, \hat{\mu}^{(1)}, \hat{\sigma}^{(1)}, \hat{\gamma}^{(1)}, \hat{\boldsymbol{\theta}}_s^{(1)'})'$ that we obtained in Step 4, repeat Steps 2 - 4 R times, to generate $\hat{\boldsymbol{\xi}}^{(r)} = (\hat{\mathbf{b}}^{(r)'}, \hat{\boldsymbol{\beta}}^{(r)'}, \hat{\mu}^{(r)}, \hat{\sigma}^{(r)}, \hat{\gamma}^{(r)}, \hat{\boldsymbol{\theta}}_s^{(r)'})'$, $r = 1, \dots, R$. The results are a sequence of estimates as a Markov Chain, and instead of converging to a single value, it converges to a stationary distribution under the standard conditions as discussed in Diebolt and Celeux (1993) and Diebolt and Ip

(1996).

Step 6: Burn-in and the MLE step

To obtain the stationary distribution, we discard the first r^* iterations as a burn-in, and then compute the estimates by averaging every third of the remaining iterates to avoid auto-correlation. By adopting the burn-in period, the random perturbations of the Markov chains preclude the influence of local maximum, so that the estimates become more reliable.

3.5 Simulation Study for Spatially Correlated Data

As discussed in Section 1.8.2, we fix the total number of individuals to be 500 and 1000. We further assume the patients resident in the 5 regions/5 zip codes, in Minnesota, US. The latitude, longitude, zip codes and cities are listed in Table 1.2. In each region, we have 100 and 200 patients, for the two sample sizes accordingly.

The competing cause M_{ij} , with spatial frailties, is assumed to follow Bernoulli, Poisson, geometric, or logarithmic distributions. Let the covariate x_{ij} be a categorical variable taking values 0 or 1, generated from a Bernoulli distribution with success probability 0.6. Combining this with the geographic information for each individual, we construct Models 6–9.

For Models 6 and 7, the power parameter θ_{ij} is taken as

$$\theta_{ij} = \exp(b_0 + b_1 x_{ij} + u_i).$$

For Models 8 and 9, θ_{ij} is given by

$$\theta_{ij} = \frac{\exp(b_0 + b_1 x_{ij} + u_i)}{1 + \exp(b_0 + b_1 x_{ij} + u_i)}.$$

Here, $i = 1, \dots, 5$ and $j = 1, \dots, 100$ (or 200).

Censored time, C_{ij} , is set to follow exponential with parameter cc . We generate Y_{ij} from the quantile function of the GEV distribution with parameters, μ , σ , and γ . If $Y_{ij} \leq C_{ij}$, we set $\delta_{ij} = 1$, and $\delta_{ij} = 0$ otherwise.

We run the algorithm 1200 times and discard the first 200 as the burn-in period, and set the spacing to be 3.

Table 3.14: Some of the true values selected for the model

Parameter:	b_0	b_1	β_1	μ	σ	γ	μ_s	σ_s^2	ϕ_s
True Value	0.500	0.400	0.400	0.100	0.100	1.600	-0.800	0.200	0.500
True Value	0.400	-0.590	-0.550	0.100	0.100	0.000	-0.800	0.400	0.800
True Value	0.400	0.900	0.900	0.200	0.600	2.500	-0.800	0.400	1.200
True Value	0.720	0.550	0.730	0.150	0.470	0.000	-1.300	0.600	1.100
True Value	0.880	0.880	1.300	0.150	0.350	0.000	-0.990	0.750	2.000
True Value	0.400	-0.590	-0.550	0.200	0.780	-1.500	-0.800	0.400	0.800
True Value	-0.650	0.820	0.850	0.250	0.450	-1.300	-0.800	0.400	1.000
True Value	0.880	0.820	0.820	0.330	0.450	-1.200	-0.990	0.750	2.000

Table 3.15: The setup of the spatial correlation, ϕ_s , considered for the simulated dataset

ϕ_s
0.500
0.800
1.000
1.100
1.200
2.000

Table 3.16: The settings of the shape parameter, γ , considered for the baseline function

γ			
Type I	0.000	Type III	-1.200
Type II	1.600		-1.300
	2.500		-1.500

The mean parameter estimates were obtained under the Bernoulli (Model 5), Poisson (Model 6), geometric (Model 7), and logarithmic (Model 8) cure rate models with spatial frailties. Overall, the estimates are close to the true values, particularly for larger sample sizes.

When the candidate model fitted to the simulated data is the same as the model used for data generation, the estimates exhibit minimal bias and RMSE, with coverage probabilities close to their nominal levels. Bias and RMSE generally decrease, and coverage probabilities increase, as the sample size grows from 500 to 1000, reflecting the additional information available with larger datasets. Coverage probabilities may fall below nominal levels for some parameters when the fitted model differs from the true model.

The spatial dependence parameter ϕ_s is consistently well-estimated across all models and settings, even under model misspecification, with convergence probabilities close to the nominal 95% level.

3.5.1 Model discrimination

In this section, we further investigate the performance of the proposed models. We generate 50 datasets under the choice of setting corresponding to the true model being Bernoulli with spatial frailty (Model 5), Poisson with spatial frailty (Model 6),

Geometric with spatial frailty (Model 7), and logarithmic with spatial frailty (Model 8), and baseline being Type I, Type II, and Type III of the GEV distribution. The sample size of our dataset is fixed to be 500. We then fitted the proposed Models 5-8 to the datasets. The results obtained were then compared by information-based criteria such as AIC, BIC, and AICc as discussed in Chapter 1.

The selection rates based on AIC are presented in Table 3.19. The correct models had the highest selection rates among the candidate models. Under Models 5-8, the selection rates for choosing the correct model are 0.92, 0.94, 0.90, and 0.92, respectively. Overall, the performance of the purposed models is seen to be very good.

Table 3.17: Simulation results on mean estimates, bias, and root mean square error (RMSE) under different choices of baseline, $\gamma = (-1.3, 0, 1.6, 2.5)$, and different spatial dependencies $\phi_s = (0.5, 0.8, 1, 1.1, 1.2)$, based on 1200 iterations

moderate censoring = 0.58							
baseline: $\gamma = 2.5$, spatial: $\phi_s = 1.2$							
low to mid cure rate = (0.27, 0.37, 0.38, 0.40, 0.44)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 5	Model 5	b_0	0.400	0.434	0.034	0.303
			b_1	0.900	0.917	0.017	0.523
			β_1	0.900	0.907	0.007	0.208
			μ	0.200	0.238	0.038	0.069
			σ	0.600	0.564	-0.036	0.153
			γ	2.500	2.351	-0.149	0.262
			μ_s	-0.800	-0.803	-0.003	0.084
			σ_s^2	0.400	0.398	-0.002	0.048
			ϕ_s	1.200	1.201	0.001	0.021
low censoring = 0.35							
baseline: $\gamma = 0$, spatial: $\phi_s = 0.8$							
mid to high cure rate = (0.50, 0.51, 0.53, 0.58, 0.64))							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 5	Model 5	b_0	0.400	0.477	0.077	0.144
			b_1	-0.590	-0.605	-0.015	0.091
			β_1	-0.550	-0.564	-0.014	0.084
			μ	0.100	0.008	-0.092	0.098
			σ	0.100	0.048	-0.052	0.130
			γ	0.000	-0.006	-0.006	0.012
			μ_s	-0.800	-0.795	0.004	0.005
			σ_s^2	0.400	0.395	-0.005	0.006
			ϕ_s	0.800	0.810	0.010	0.010
moderate to high censoring = 0.57							
baseline: $\gamma = -1.3$, spatial: $\phi_s = 1$							
moderate to high cure =(0.53, 0.54, 0.55, 0.56, 0.64)							
n	True Model	Fitted Model	Parameter	T.V.	Estimate	Bias	RMSE
1000	Model 6	Model 6	b_0	-0.650	-0.686	-0.036	0.408
			b_1	0.820	0.833	0.013	0.386
			β_1	0.850	0.832	-0.018	0.170
			μ	0.250	0.343	0.093	0.100
			σ	0.450	0.452	0.002	0.071
			γ	-1.300	-1.279	0.021	0.426
			μ_s	-0.800	-0.800	0.001	0.009
			σ_s^2	0.400	0.412	0.012	0.018
			ϕ_s	1.000	0.990	-0.010	0.011

high level censoring = 0.686							
baseline: $\gamma = 0$, spatial: $\phi_s = 1.1$							
low to mid cure rate = (0.18, 0.20, 0.21, 0.32, 0.4)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 6	Model 6	b_0	0.720	0.706	-0.014	0.027
			b_1	0.550	0.568	0.018	0.031
			β_1	0.730	0.716	-0.014	0.015
			μ	0.150	0.111	-0.039	0.040
			σ	0.470	0.471	0.001	0.013
			γ	0.000	-0.097	-0.097	0.097
			μ_s	-1.300	-1.310	-0.010	0.010
			σ_s^2	0.600	0.601	0.001	0.004
			ϕ_s	1.100	1.100	0.000	0.002
low censoring = 0.283							
baseline: $\gamma = 0$, spatial: $\phi_s = 2$							
high cure rate = (0.79, 0.83, 0.85, 0.85, 0.88)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 7	Model 7	b_0	0.880	0.942	0.062	0.355
			b_1	0.880	0.905	0.025	0.138
			β_1	1.300	1.317	0.017	0.084
			μ	0.150	0.138	-0.012	0.052
			σ	0.350	0.362	0.012	0.125
			γ	0.000	-0.003	-0.003	0.005
			μ_s	-0.990	-0.990	0.000	0.001
			σ_s^2	0.750	0.750	0.000	0.002
			ϕ_s	2.000	2.000	0.000	0.001
high level censoring = 0.604							
baseline: $\gamma = 1.6$, spatial: $\phi_s = 0.5$							
mid to high cure rate = (0.43, 0.52, 0.55, 0.60, 0.67)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 7	Model 7	b_0	0.50	0.316	-0.184	0.218
			b_1	0.40	0.324	-0.076	0.143
			β_1	0.40	0.450	0.050	0.084
			μ	0.10	0.089	-0.011	0.023
			σ	0.10	0.192	0.092	0.104
			γ	1.60	2.100	0.500	0.517
			μ_s	-0.80	-0.800	0.000	0.012
			σ_s^2	0.20	0.199	-0.001	0.030
			ϕ_s	0.50	0.501	0.001	0.010

moderate censoring = 0.564							
baseline: $\gamma = -1.2$, spatial: $\phi_s = 2$							
moderate cure rate = (0.47, 0.51, 0.54, 0.54, 0.57)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 7	Model 7	b_0	0.880	0.921	0.041	0.350
			b_1	0.820	0.822	0.002	0.076
			β_1	0.820	0.826	0.006	0.046
			μ	0.330	0.212	-0.118	0.625
			σ	0.450	0.510	0.060	0.266
			γ	-1.200	-1.183	0.017	0.257
			μ_s	-0.990	-0.980	0.010	0.055
			σ_s^2	0.750	0.741	-0.009	0.052
			ϕ_s	2.000	2.020	0.020	0.107
moderate censoring = 0.564							
baseline: $\gamma = 0$, spatial: $\phi_s = 0.8$							
high cure rate = (0.628, 0.587, 0.658, 0.626, 0.69)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
1000	Model 8	Model 8	b_0	0.350	0.390	0.040	0.215
			b_1	-0.880	-0.848	0.032	0.159
			β_1	-0.850	-0.802	0.048	0.161
			μ	0.850	0.793	-0.057	0.191
			σ	1.000	0.904	-0.096	0.236
			γ	0.000	-0.001	-0.001	0.003
			μ_s	-0.400	-0.410	-0.010	0.028
			σ_s^2	0.400	0.410	0.010	0.029
			ϕ_s	0.800	0.810	0.010	0.028
low to moderate censoring = 0.494							
baseline: $\gamma = 0$, spatial: $\phi_s = 0.8$							
low to moderate cure rate = (0.36, 0.44, 0.43, 0.45, 0.55)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
500	Model 5	Model 5	b_0	0.400	0.306	-0.094	0.118
			b_1	-0.590	-0.682	-0.092	0.111
			β_1	-0.550	-0.633	-0.083	0.119
			μ	0.100	0.020	-0.080	0.085
			σ	0.100	0.088	-0.012	0.158
			γ	0.000	-0.023	-0.023	0.041
			μ_s	-0.800	-0.795	0.005	0.006
			σ_s^2	0.400	0.396	-0.004	0.008
			ϕ_s	0.800	0.810	0.010	0.011

low to moderate censoring = 0.394							
baseline: $\gamma = -1.3$, spatial: $\phi_s = 1$							
high cure rate = (0.63, 0.64, 0.68, 0.69, 0.72)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
500	Model 6	Model 6	b_0	-0.650	-0.553	0.097	0.314
			b_1	0.820	0.853	0.033	0.101
			β_1	0.850	0.857	0.007	0.028
			μ	0.250	0.164	-0.086	0.101
			σ	0.450	0.481	0.031	0.107
			γ	-1.300	-1.218	0.082	0.310
			μ_s	-0.800	-0.800	0.000	0.006
			σ_s^2	0.400	0.411	0.011	0.015
			ϕ_s	1.000	0.990	-0.010	0.010
high censoring = 0.596							
baseline: $\gamma = -1.2$, spatial: $\phi_s = 2$							
low to moderate cure rate = (0.44, 0.42, 0.40, 0.48 ,0.43)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
500	Model 7	Model 7	b_0	0.880	0.882	0.002	0.163
			b_1	0.820	0.789	-0.031	0.080
			β_1	0.820	0.803	-0.017	0.041
			μ	0.330	0.318	-0.012	0.027
			σ	0.450	0.398	-0.052	0.064
			γ	-1.200	-1.158	0.042	0.189
			μ_s	-0.990	-0.980	0.010	0.012
			σ_s^2	0.750	0.741	-0.009	0.014
			ϕ_s	2.000	2.020	0.020	0.021
low censoring = 0.312							
baseline: $\gamma = 0$, spatial: $\phi_s = 0.8$							
high cure rate = (0.72, 0.62, 0.63, 0.58, 0.61)							
n	True Model	Fitted Model	Parameter	TV	Estimate	Bias	RMSE
500	Model 8	Model 8	b_0	0.350	0.446	0.096	0.349
			b_1	-0.880	-0.813	0.067	0.241
			β_1	-0.850	-0.776	0.074	0.221
			μ	0.850	0.784	-0.066	0.235
			σ	1.000	0.924	-0.076	0.267
			γ	0.000	-0.007	-0.007	0.037
			μ_s	-0.400	-0.410	-0.010	0.011
			σ_s^2	0.400	0.410	0.010	0.013
			ϕ_s	0.800	0.810	0.010	0.011

Table 3.18: Comparison of simulation results on mean estimates, bias, and root mean square error (RMSE) for different fitted models based on 1200 iterations.

censoring = 0.604						
moderate to high cure = (0.43, 0.52, 0.55, 0.60, 0.67)						
n	True Model	Fitted Model	Parameter	Estimate	Bias	RMSE
1000	Model 7	Model 5	b_0	0.458	-0.042	0.105
			b_1	0.432	0.032	0.105
			β_1	0.524	0.124	0.195
			μ	0.065	-0.035	0.043
			σ	0.162	0.062	0.091
			γ	2.454	0.854	0.903
			μ_s	-0.800	0.000	0.007
			σ_s^2	0.198	-0.002	0.020
			ϕ_s	0.501	0.001	0.006
1000	Model 7	Model 6	b_0	-0.058	-0.558	0.570
			b_1	0.336	-0.064	0.092
			β_1	0.346	-0.054	0.159
			μ	0.101	0.001	0.030
			σ	0.238	0.138	0.163
			γ	2.283	0.683	0.718
			μ_s	-0.799	0.001	0.009
			σ_s^2	0.199	-0.001	0.027
			ϕ_s	0.500	0.000	0.009
1000	Model 7	Model 8	b_0	1.008	0.508	0.534
			b_1	0.555	0.155	0.191
			β_1	0.477	0.077	0.147
			μ	0.087	-0.013	0.031
			σ	0.169	0.069	0.081
			γ	1.909	0.309	0.333
			μ_s	-0.500	0.300	0.301
			σ_s^2	0.101	-0.099	0.123
			ϕ_s	0.500	0.000	0.012

Table 3.19: Selection rates based on AIC ($n=500$)

	B (Model 5)	P (Model 6)	G(Model 7)	L (Model 8)
B (Model 5)	0.92	0.04	0.02	0.02
P (Model 6)	0.04	0.94	0.02	0.00
G (Model 7)	0.02	0.06	0.90	0.02
L (Model 8)	0.00	0.02	0.06	0.92

3.6 Analysis of smoking cessation data

We applied the proposed models to the smoking cessation dataset, using the Stochastic EM algorithm to obtain the optimal estimates of the model parameters. The results indicate that the coefficients for b_1 (Gender) are all positive, for b_3 (Treatment) are all negative, for b_4 (Consumption) are all positive, and for β_2 (Duration) are all negative. The coefficients for β_4 (Consumption) are all positive. For b_2 (Duration), the sign is positive for the Bernoulli mixture model and negative for the other models. The coefficients for b_0 (Intercept) and β_3 (Treatment) are negative for the Bernoulli and promotion time cure models, and positive for the others. The coefficient for β_1 (Gender) is negative only for the Poisson cure model.

The mean estimates of the baseline distribution parameters, μ (mean), σ (scale), and γ (shape), as well as the spatial effect parameters, μ_s , σ_s^2 , and ϕ_s , are similar across all models considered, with $\phi_s \approx 0.6$ for all models. This indicates that the spatial effect is present and has been successfully captured by the models.

Also, $\mathbf{b}$ parameters are related to the cure fraction and so we may obtain inferences on the cure rate. If we keep all other variables as fixed values, the women smokers have a lower probability of quitting than men smokers (Gender: 1= female). The smokers who received smoking intervention (Treatment) have a higher probability of quitting than those who received the usual care. The smoker who has a higher level of cigarette consumption has a lower probability of quitting than those who smoke less. We further visualize the impact of location, as a spatial frailty component, on top of survival time for different cohorts of smokers using survival plots. Plots for survival functions stratified by location, treatment, gender, and consumption are shown in Figures 3.11 - ??.

Confidence intervals for the model parameters were calculated using (1.33) and are reported in Table 3.20, while Table 3.22 presents the negative log-likelihood, AIC, BIC, and AICc values. Under the Poisson assumption for competing causes (promotion time cure model with Gaussian spatial frailties, Model 6) and a Type II GEV baseline, the lowest AIC, BIC, and AICc values were obtained: 285.260, 600.521, 605.745, and 602.839, respectively.

Likelihood ratio tests comparing models with and without spatial frailties (Models 5–8 vs. Models 1–4) yielded p-values of 0.006, 0.002, 0.004, and 0.014, indicating that models with spatial frailties provide a significantly better fit.

Goodness-of-fit was assessed using normalized random quantile residuals from the SEM estimates (R. and J., 1968). QQ plots for Model 6 (Figure 3.10) and the Kolmogorov–Smirnov test ($p = 0.194$) support the normality of residuals, confirming an adequate fit.

Spatial frailties across 51 zip codes (Figure 3.7) reveal regional variation, with darker purple indicating stronger positive and lighter purple stronger negative frailties. Maps of differences in cure rates and survival probabilities (Figure 3.8) show that areas with stronger negative frailties tend to increase outcomes, whereas areas with stronger positive frailties decrease them. Cure rates and survival probabilities stratified by zip code (Figure 3.9) are consistent with these patterns.

Parameter evolution plots (Figure 3.12) confirm convergence of the SEM algorithm, with no systematic upward or downward trends. Additional evaluation plots are provided in Appendix B.

Figure 3.7: Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota

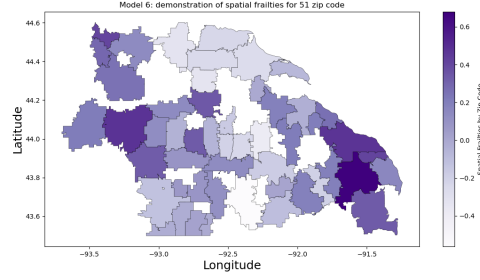


Figure 3.8: Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota

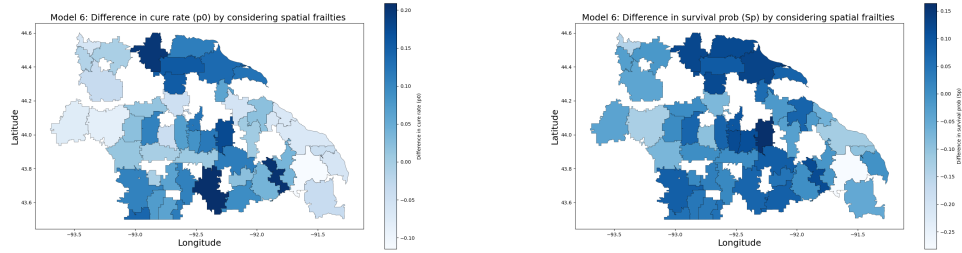


Figure 3.9: Demonstration of different spatial frailties for patients from 51 zip codes in Minnesota

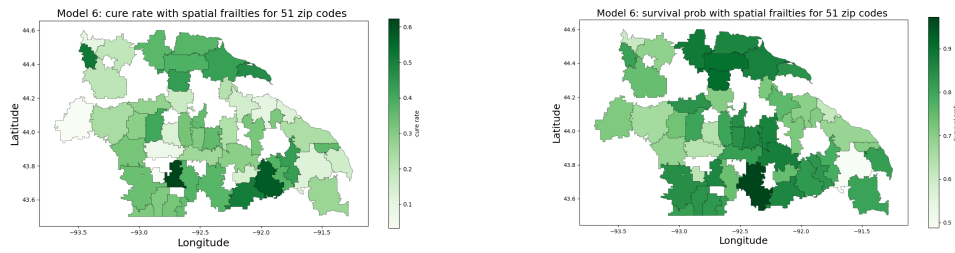


Table 3.20: Mean estimates, standard error, and 95% CI for model parameters for the smoking cessation data assuming M_{ij} follow Models 5 and 6 (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Model 5				
Parameter	Mean	SE	2.5%	97.5%
b_0 (Intercept)	-1.596	0.810	-3.185	-0.008
b_1 (Gender)	0.368	0.325	-0.270	1.006
b_2 (Duration)	0.010	0.022	-0.032	0.053
b_3 (Treatment)	-0.400	0.374	-1.134	0.334
b_4 (Consumption)	0.065	0.017	0.032	0.097
β_1 (Gender)	0.463	0.245	-0.018	0.944
β_2 (Duration)	-0.025	0.011	-0.046	-0.003
β_3 (Treatment)	-0.340	0.269	-0.867	0.186
β_4 (Consumption)	0.001	0.010	-0.020	0.022
μ	3.474	0.658	2.184	4.764
σ	7.739	1.989	3.840	11.639
γ	2.838	0.225	2.396	3.279
μ_s	0.692	0.106	0.485	0.899
σ_s^2	0.106	0.030	0.047	0.165
ϕ_s	0.570	0.156	0.264	0.876
Model 6				
Parameter	Mean	SE	2.5%	97.5%
b_0 (Intercept)	-0.390	0.451	-1.274	0.494
b_1 (Gender)	0.682	0.188	0.314	1.050
b_2 (Duration)	-0.010	0.012	-0.034	0.013
b_3 (Treatment)	-0.412	0.210	-0.823	-0.001
b_4 (Consumption)	0.038	0.009	0.021	0.056
β_1 (Gender)	-0.225	0.304	-0.822	0.371
β_2 (Duration)	-0.007	0.013	-0.032	0.019
β_3 (Treatment)	0.000	0.323	-0.634	0.634
β_4 (Consumption)	-0.024	0.012	-0.049	0.000
μ	3.402	0.695	2.040	4.764
σ	7.450	2.052	3.427	11.472
γ	2.803	0.220	2.371	3.235
μ_s	0.088	0.118	-0.144	0.320
σ_s^2	0.084	0.037	0.011	0.156
ϕ_s	0.626	0.294	0.050	1.203

Table 3.21: Estimates of mean, standard error, and 95% CI for model parameters for the smoking cessation data assuming M_{ij} follow Models 7 and 8 (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Model 7				
Parameter	Mean	SE	2.5%	97.5%
b_0 (Intercept)	0.474	0.690	-0.878	1.826
b_1 (Gender)	0.528	0.291	-0.043	1.099
b_2 (Duration)	-0.022	0.019	-0.059	0.015
b_3 (Treatment)	-0.503	0.325	-1.139	0.134
b_4 (Consumption)	0.022	0.012	-0.001	0.046
β_1 (Gender)	0.069	0.336	-0.589	0.728
β_2 (Duration)	-0.012	0.015	-0.041	0.017
β_3 (Treatment)	-0.011	0.348	-0.692	0.670
β_4 (Consumption)	-0.004	0.013	-0.030	0.022
μ	4.243	1.186	1.918	6.568
σ	9.454	3.604	2.390	16.517
γ	2.697	0.196	2.313	3.081
μ_s	0.091	0.118	-0.140	0.322
σ_s^2	0.084	0.029	0.027	0.141
ϕ_s	0.631	0.280	0.082	1.181
Model 8				
Parameter	Mean	SE	2.5%	97.5%
b_0 (Intercept)	1.797	1.202	-0.558	4.152
b_1 (Gender)	0.740	0.542	-0.321	1.802
b_2 (Duration)	-0.062	0.035	-0.130	0.005
b_3 (Treatment)	-0.772	0.619	-1.984	0.441
b_4 (Consumption)	0.070	0.024	0.023	0.118
β_1 (Gender)	0.023	0.409	-0.779	0.825
β_2 (Duration)	-0.003	0.018	-0.038	0.033
β_3 (Treatment)	0.083	0.447	-0.792	0.958
β_4 (Consumption)	-0.014	0.017	-0.047	0.019
μ	4.477	1.396	1.740	7.214
σ	8.293	4.069	0.317	16.268
γ	2.121	0.202	1.725	2.518
μ_s	0.084	0.213	-0.334	0.502
σ_s^2	0.162	0.079	0.007	0.316
ϕ_s	0.675	0.281	0.124	1.225

Table 3.22: Negative log-likelihood, AIC, BIC, and AICc values for the smoking cessation data assuming M_{ij} follow Models 5 - 8 (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Activation	Distribution of M_{ij}	-ll	AIC	BIC	AICc
Random	Bernoulli (Model 5)	289.060	608.120	613.345	610.439
First	Poisson (Model 6)	285.260	600.520	605.745	602.839
	Geometric (Model 7)	289.604	609.209	614.433	611.528
	Logarithmic (Model 8)	289.331	608.662	613.887	610.981

Figure 3.10: QQ plot

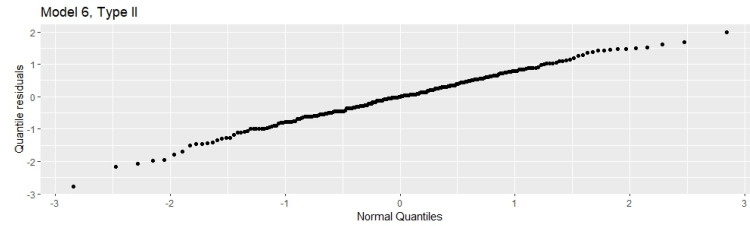
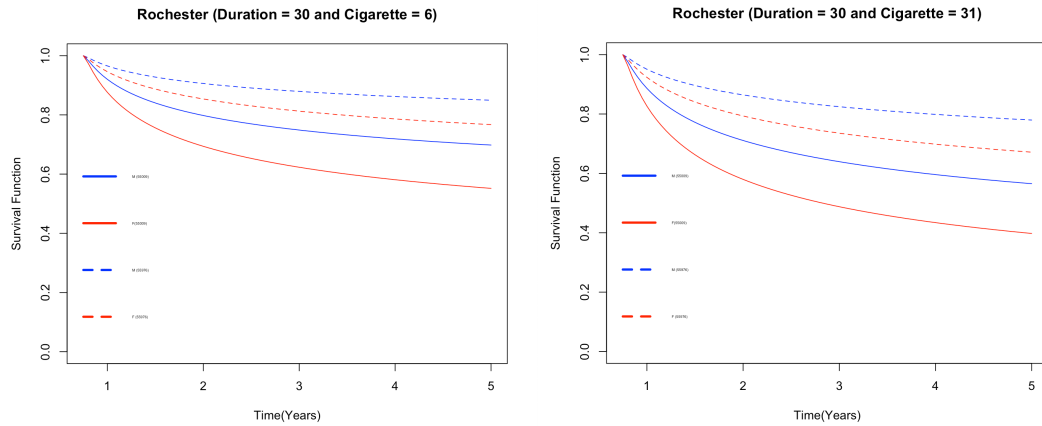
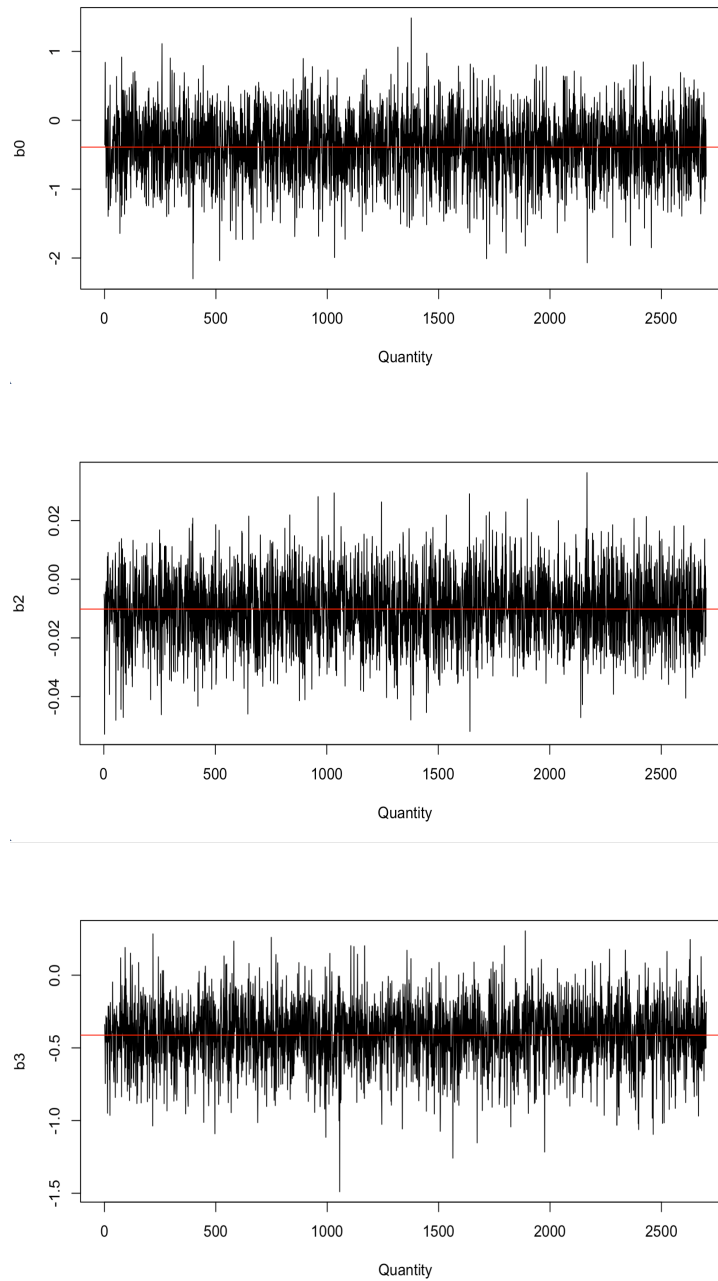


Figure 3.11: Surviving function: stratified by location (Cannon Falls, Minnesota; Stewartville, Minnesota) and gender with Duration (30 years-mean) and cigarette consumption: 6 and 31.



The blue line indicates male smoker, and the red line indicates a female smoker. Blue: Male (zip code: 55009 - Cannon Falls), Red: Female (zip code: 55009 - Cannon Falls), Blue dashed: Male (zip code: 55976 - Stewartville), Red dashed: Female (zip code: 55976 - Stewartville)

Figure 3.12: Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)



Chapter 4

Stochastic EM-based Likelihood Inference for Last Activation Scheme of PS Cure Rate Model with Gaussian Spatial Frailties

4.1 Introduction

In this Chapter, we extend the cure rate models proposed in Chapter 3 to models that focus on the cases when the event of interest only takes place when all of the competing causes are initiated. It is also called last activation scheme of PS cure rate model (Noack (1950)), introduced earlier in Section 1.1.3.

Given that the number of competing causes for the ij th individual is $M_{ij} = m$, let random variable W_{k^*ij} be the time-to-event due to the k^* th latent risk, with

distribution function $F(\cdot) = 1 - S(\cdot)$ and survival function $S(\cdot)$, for $k^* = 1, 2, \dots, m$.

Under the last activation scheme, the time to the event of interest is denoted by the random variable $T_{ij} = \max\{W_{1ij}, \dots, W_{M_{ij}}\}$ for $M_{ij} \geq 1$, and $T_{ij} = \infty$ if $M_{ij} = 0$ with $P[T_{ij} = \infty \mid M_{ij} = 0] = 1$.

In addition, the spatial effect on survival time of susceptible patients is also incorporated. Spatial component is assumed to follow a Gaussian process as defined in Section 1.4 (Li and Ryan (2002)). The spatial frailties are once again developed using the geographic information on individuals (i.e, longitude and latitude) in the spatial-referred data.

In this Chapter, by adjusting the power parameter and series function, we restrict the last activation scheme to result in complementary promotion time cure rate model, complementary geometric cure rate model, and complementary logarithmic cure rate model. The Gaussian spatial effect is then added to the proposed models and produces complementary promotion time cure rate model with spatial effect (Model 9), complementary geometric cure rate model with spatial effect (Model 10), and complementary logarithmic cure rate model with spatial effect (Model 11). The flexible baseline function that follows a GEV distribution is adopted, and with different values of the shape parameter γ , the baseline function changes to Type I ($\gamma = 0$), Type II ($\gamma > 0$), and Type III of GEV ($\gamma < 0$).

The relative functions such as survival functions, and density functions of all the models are presented in Section 4.2. The likelihood function and expression of SEM are described in Sections 4.3 and 4.4, respectively. In Section 4.5, a simulation study is conducted with the consideration to multiple sample sizes and different parameter settings. Different spatial associations to the event of interests are also considered.

In addition, the effectiveness of capturing spatial frailty is demonstrated in the simulation studies. The performance of Models 9 -11 are then evaluated through a model discrimination.

In addition, in Section 4.6, the proposed models are fitted to a real-life dataset on smoking cessation. The spatial frailty is captured effectively, and its impact on survival is then visualized with heat maps.

4.2 Cure models

For the Bernoulli mixture cure model with spatial frailties (Model 5) and the complementary promotion time Poisson cure model with spatial frailties (Model 9), we take

$$\theta_{ij} = \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i),$$

where u_i is the spatial frailty for region i , and $\mathbf{b}$ is the vector of covariate effects on the cure probability p_{0ij} .

As discussed in Section 3.2, when M_{ij} follows a Bernoulli distribution and is combined with the survival and hazard functions conditional on spatial frailty, we obtain the Bernoulli mixture cure model with spatial frailties (Model 5). Its series function is

$$G(\theta_{ij}) = 1 + \theta_{ij},$$

and the corresponding cure probability is

$$p_{0ij} = P(M_{ij} = 0) = \frac{1}{1 + \theta_{ij}}.$$

In this case, the population survival and density functions are given by

$$S_p(t_{ij}) = p_{0ij} + (1 - p_{0ij})S(t_{ij}), \quad (4.103)$$

$$f_p(t_{ij}) = \frac{\theta_{ij}}{(1 + \theta_{ij})} f(t_{ij}), \quad (4.104)$$

where $S(t_{ij}) = S_0(t_{ij})^{\exp(z_{ij}^T \beta + u_i)}$, $S_0(t_{ij})$ is the baseline survival function in (1.14), and u_i is the spatial frailty from a Gaussian process, as described in Section 1.4. The density function is $f(t_{ij}) = \lambda_0(t_{ij})S(t_{ij})$, where $\lambda_0(t_{ij})$ is the baseline hazard in (1.15).

Complementary promotion time cure with spatial frailties (Model 9)

The series function is $G(\theta_{ij}) = \exp(\theta_{ij})$, when the competing cause, M_{ij} follow a Poisson distribution. Now taking into account the survival and hazard function conditioned on spatial effect, along with last activation scheme of the PS cure rate model, we obtained the complementary promotion time cure Poisson model with spatial frailties (Model 9). The population survival function and probability density function are given by

$$p_{0ij} = \exp(-\theta_{ij}) \quad (4.105)$$

$$S_p(t_{ij}) = 1 + \exp(-\theta_{ij}) - \exp(-\theta_{ij}S(t_{ij})), \quad (4.106)$$

$$f_p(t_{ij}) = \theta_{ij}f(t_{ij})\exp(-\theta_{ij}S(t_{ij})), \quad (4.107)$$

where $\theta_{ij} > 0$.

For the complementary geometric cure rate model with spatial frailties (Model 10) and complementary logarithmic cure rate model with spatial frailties (Model 11),

we similarly have

$$\theta_{ij} = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}, \quad (4.108)$$

where u_i is spatial frailty corresponding to region i , and $\mathbf{b}$ is the vector representing the effect of covariates on the cure probability p_{0ij} .

Complementary geometric cure model with spatial frailties (Model 10)

If the series function of the PS distribution is $G(\theta_{ij}) = \frac{1}{1-\theta_{ij}}$, and the competing cause, M_{ij} , follows a geometric distribution, the complementary geometric cure model with spatial frailties (Model 10) is obtained utilizing the last activation scheme of the PS cure rate model and Gaussian spatial frailty. The associated cure probability is $p_{0ij} = 1 - \theta_{ij}$. The population survival function and probability density function, conditioned on spatial frailties, and are given by

$$S_p(t_{ij}) = 1 + (1 - \theta_{ij}) - \frac{(1 - \theta_{ij})}{1 - \theta_{ij}F(t_{ij})}, \quad (4.109)$$

$$f_p(t_{ij}) = \theta_{ij}(1 - \theta_{ij})f(t_{ij})[1 - \theta_{ij}F(t_{ij})]^{-2}, \quad (4.110)$$

where $F(t_{ij}) = 1 - S(t_{ij})$, and $0 < \theta_{ij} < 1$.

Complementary logarithmic cure rate model with spatial frailties (Model 11)

If $G(\theta_{ij}) = \frac{-\log(1-\theta_{ij})}{\theta_{ij}}$, and the competing cause, M_{ij} , follows a logarithmic distribution, combining it with the last activation scheme of PS cure rate model together with

spatial component, a complementary logarithmic cure rate model with spatial frailties (Model 11) is obtained. Its cure probability is $p_{0ij} = \frac{-\theta_{ij}}{\log(1-\theta_{ij})}$. The population survival function and population probability density function are given by

$$S_p(t_{ij}) = 1 - \frac{\theta_{ij}}{\log(1 - \theta_{ij})} - \frac{\log(1 - \theta_{ij}F(t_{ij}))}{F(t_{ij})\log(1 - \theta_{ij})}, \quad (4.111)$$

$$f_p(t_{ij}) = -\frac{f(t_{ij})}{F(t_{ij})\log(1 - \theta_{ij})} \left[\frac{\log(1 - \theta_{ij}F(t_{ij}))}{F(t_{ij})} + \frac{\theta_{ij}}{1 - \theta_{ij}F(t_{ij})} \right], \quad (4.112)$$

where $0 < \theta_{ij} < 1$.

4.3 The likelihood function

4.3.1 Model 9

Combining the Eqs. (4.105) - (4.107), the complete log-likelihood function of Model 9 is obtained as

$$\begin{aligned} l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\ &\quad \times \exp \left[- \exp(\mathbf{x}_i^T \mathbf{b} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right] \left. \right\} + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{ \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \} \\ &\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_i^* \log \left\{ 1 - \exp \left[- \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right] \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s), \end{aligned} \quad (4.113)$$

where the baseline survival function, $S_0(t_{ij})$, and the baseline hazard function, $\lambda_0(t_{ij})$, are conditioned on spatial frailty, u_i . Baseline changes to Type I, Type II, and Type

III of GEV depending on parameter γ . f_U is the density function of the spatial effect and it follows a Gaussian process as explained in Section 1.4. The detailed steps of obtaining the log-likelihood function for Models 9 - 11 are provided in Appendix C.

4.3.2 Model 10

The complete log-likelihood function of Model 10 is given by

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \left. \times S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b})} \left(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right) \right]^{-2} \right\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&+ \sum_{(i,j): \delta_{ij}=0} \sum J_i^* \log \left\{ 1 - \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} (1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})} \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{4.114}$$

4.3.3 Model 11

The log-likelihood function of Model 11 is given by

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij}) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)}{(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)) \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log\left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right)}{1 - S_0(t_{ij}) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))} \right] \Big\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_i^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_i^* \log \left\{ \frac{\log\left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right)}{(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{4.115}$$

However, we do not have complete information on subjects who are right censored. The survival status remains unknown these censored subjects, since they can be either cured or susceptible. We overcome this difficulties by implementing the stochastic step (S-step) described in the following section.

4.4 Stochastic EM

Step 1: We find the set of initial values via grid search in the parameter space;

Step 2: Recall that for censored subject ij , we have $\delta_{ij} = 0$, and J_{ij}^* can be generated from a Bernoulli distribution with conditional probability of success as

$$\begin{aligned} p_{0ij}^{(0)} &= P[J_{ij}^* = 1 | T_{ij} > c_{ij}; \boldsymbol{\xi}^{(0)}] = \frac{P[T_{ij} > c_{ij} | J_{ij}^* = 1]P[J_{ij}^* = 1]}{P[T_{ij} > c_{ij}]} \Bigg|_{\boldsymbol{\xi} = \boldsymbol{\xi}^{(0)}} \\ &= \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}, \end{aligned}$$

where p_{0ij} and S_p are cure probabilities and survival function from (3.71) and (3.72), respectively. J_{ij}^* is then replaced with $\pi_{ij}^{(1)}$ in the complete log-likelihood functions in (3.86), (4.113), (4.114), and (4.115) when assuming Model 5, Model 9, Model 10, and Model 11, respectively. $\pi_{ij}^{(1)}$ is the generated cure status for ij th individual with censored lifetime for the first round of iteration.

The case of Model 5 has been discussed in detail in Chapters, and so we will omit it here for the sake of brevity.

Model 9: The J_{ij}^* in (4.113) are replaced by $\pi_{ij}^{(1)}$, and the complete log-likelihood

function of Model 9 can then be rewritten as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\
&\quad \times \exp[-\exp(\mathbf{x}_i^T \mathbf{b}) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}] \left. \right\} + \sum_{(i,j): \delta_{ij}=0} \sum (1 - \pi_{ij}^{(1)}) \log \{ \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum \pi_{ij}^{(1)} \log \left\{ 1 - \exp[-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}] \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{4.116}$$

Model 10: The J_{ij}^* in (4.114) are replaced by $\pi_{ij}^{(1)}$, and the complete log-likelihood function of Model 10 can then be rewritten as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_i) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \times S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right]^{-2} \left. \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum (1 - \pi_{ij}^{(1)}) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum \pi_{ij}^{(1)} \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} - \left[\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right] \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{4.117}$$

Model 11: The J_{ij}^* in (4.115) is replaced by $\pi_{ij}^{(1)}$, and the complete log-likelihood

function of Model 11 can be rewritten as

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) &= \sum_{(i,j): \delta_{ij}=1} \sum_{\delta_{ij}=1} \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)}}{(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}) \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log\left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right)}{1 - S_0(t_{ij})^{\exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)}} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})} \right] \Bigg\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} (1 - \pi_{ij}^{(1)}) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum_{\delta_{ij}=0} \pi_{ij}^{(1)} \log \left\{ \frac{\log\left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right)}{(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{4.118}$$

Step 3: If the censored subject is susceptible, i.e., $\pi_{ij}^{(1)} = 1$, the complete lifetime t_{ij}^* is from the truncated distribution with density function

$$f_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = f_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) / S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}), \tag{4.119}$$

where $c_{ij} < t_{ij}^* < \infty$. The corresponding cumulative function is given by $F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = \frac{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - S_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})}$, where $c_{ij} < t_{ij}^* < \infty$, which is not a proper cdf since $\lim_{t_{ij}^* \rightarrow \infty} F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) \neq 1$.

The cured/immunized subject with $\pi_{ij}^{(1)} = 0$ is treated as long term survivor and its lifetime is infinite with respect to the event of interest. Hence, it takes the form of

$\lim_{t_{ij}^* \rightarrow \infty} S_p(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) = p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)})$. To generate t_{ij}^* from (4.119) under the susceptible scenario, we adopt inverse transformation sampling techniques. In short, $F_T(t_{ij}^*, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)})$ follows an $\text{Uniform}(a^* = 0, b^* = S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) - p_{0ij}(\mathbf{x}_{ij} | u_i; \boldsymbol{\xi}^{(0)}) / S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij} | u_i; \boldsymbol{\xi}^{(0)}))$ distribution.

Model 9: Under Model 9, we rewrite b^* as

$$b^* = \frac{1 + \exp(-\theta_{ij}) - \exp(-\theta_{ij}S(c_{ij})) - \exp(-\theta_{ij})}{1 + \exp(-\theta_{ij}) - \exp(-\theta_{ij}S(c_{ij}))} = \frac{1 - \exp(-\theta_{ij}S(c_{ij}))}{1 + \exp(-\theta_{ij}) - \exp(-\theta_{ij}S(c_{ij}))}, \quad (4.120)$$

where $\theta_{ij} = \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)$, and $S(c_{ij})$ is constructed using baseline survival function as in (1.10). Also, c_{ij} is the censored lifetime of j th individual in region i , and u_i is spatial effect associated with region i .

Model 8: Under Model 8, we rewrite b^* as

$$b^* = \frac{1 + (1 - \theta_{ij}) - \frac{(1 - \theta_{ij})}{1 - \theta_{ij}F(c_{ij})} - (1 - \theta_{ij})}{1 + (1 - \theta_{ij}) - \frac{(1 - \theta_{ij})}{1 - \theta_{ij}F(c_{ij})}} = \frac{\theta_{ij}(1 - F(c_{ij}))}{(\theta_{ij}F(c_{ij}) - 1)^2}, \quad (4.121)$$

where $\theta_{ij} = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}$, $F(c_{ij}) = 1 - S(c_{ij})$, $S(c_{ij})$ is constructed using baseline survival function as (1.10), c_{ij} is the censored lifetime of j th individual in region i , and u_i is spatial effect associated with region i .

Model 11: Under Model 11, we rewrite b^* as

$$b^* = \frac{1 - \frac{\theta_{ij}}{\log(1 - \theta_{ij})} - \frac{\log(1 - \theta_{ij}F(c_{ij}))}{F(c_{ij})\log(1 - \theta_{ij})} - \frac{-\theta_{ij}}{\log(1 - \theta_{ij})}}{1 - \frac{\theta_{ij}}{\log(1 - \theta_{ij})} - \frac{\log(1 - \theta_{ij}F(c_{ij}))}{F(c_{ij})\log(1 - \theta_{ij})}}, \quad (4.122)$$

where $\theta_{ij} = \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}$, $F(c_{ij}) = 1 - S(c_{ij})$, and $S(c_{ij})$ is constructed using baseline

survival function as (1.10).

Step 4: Maximization step (M-step)

We fill the censored data with the generated data from Step 3. Now, the improved estimate of ξ can be found using the pseudo-complete data as

$$\xi^{(1)} = (\mathbf{b}^{(1)'}, \beta^{(1)'}, \mu^{(1)}, \sigma^{(1)}, \gamma^{(1)}, \boldsymbol{\theta}_s^{(1)'})' = \arg \max_{\xi^{(1)}} \log L_c(\xi; (\mathbf{t}, \mathbf{J}^*), (\mathbf{t}^*, \boldsymbol{\pi}^{(1)}), \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s),$$

where $\mathbf{t}^*$ and $\boldsymbol{\pi}^{(1)}$ are vectors of t_{ij}^* (generated lifetime) and $\pi_{ij}^{(1)}$ (generated cure status), respectively. The optimal values of ξ is obtained using the ‘L-BFGS-B’ package in R software, where the algorithm is set to be converged when the desired tolerance level, i.e., $|\hat{\xi}_{r+1} - \hat{\xi}_r| < 10^{-6}$, is achieved.

Step 5: Iterative step

Using the estimate $\hat{\xi}^{(1)} = (\hat{\mathbf{b}}^{(1)'}, \hat{\beta}^{(1)'}, \hat{\mu}^{(1)}, \hat{\sigma}^{(1)}, \hat{\gamma}^{(1)}, \hat{\boldsymbol{\theta}}_s^{(1)'})'$ that we obtained in Step 4, we repeat Steps 2 - 4 R times, to obtain $\hat{\xi}^{(r)} = (\hat{\mathbf{b}}^{(r)'}, \hat{\beta}^{(r)'}, \hat{\mu}^{(r)}, \hat{\sigma}^{(r)}, \hat{\gamma}^{(r)}, \hat{\boldsymbol{\theta}}_s^{(r)'})'$, $r = 1, \dots, R$. The results are a sequence of estimates as a Markov Chain, which, instead of converging to a single value, it converges to a stationary distribution under the standard conditions as discussed in Diebolt and Celeux (1993) and Diebolt and Ip (1996).

Step 6: Burn-in and MLE step

To obtain the stationary distribution, we discard the first r^* iterations as a burn-in, and compute the estimates by averaging every third of the remaining iterates to avoid auto-correlation. By adopting the burn-in period, the random perturbations of the Markov chains preclude the influence of local maximum, so that the estimates are more reliable.

4.5 Simulation study

As discussed earlier in Section 1.8.2, in this simulation study, we fix the total number of patients to be 500 and 1000. We further assume the patients reside in the 5 regions, 5 zip code, in Minnesota, US. The latitude, longitude, zipcode and cities are listed in Table 1.2. In each region, we have 100 and 200 patients, for the two sample sizes, accordingly.

In this simulation study, we consider a covariate, x_{ij} , to be a categorical variable which takes values of 0 or 1, and can be generated from a Bernoulli distribution with success probability as 0.6.

The competing cause, M_{ij} , follows a class of distributions from a power series cure rate family with spatial frailties, which are Models 5-8 described earlier in 4.2. Under the simulation setup, θ_{ij} for Models 5 and 6 can be expressed as

$$\theta_{ij} = \exp(b_0 + b_1 * x_{ij} + u_i), \quad (4.123)$$

where $i = 1, \dots, 5$, $j = 1, \dots, 100(200)$. Under Models 7 and 8, θ_{ij} can expressed as

$$\theta_{ij} = \exp(b_0 + b_1 * x_{ij} + u_i) / (1 + \exp(b_0 + b_1 * x_{ij} + u_i)), \quad (4.124)$$

where $i = 1, \dots, 5$, $j = 1, \dots, 100(200)$.

We adopt several parameter settings as listed in Table 4.26. In Table 4.25, we record the values we fixed for the spatial dependence parameter, ϕ_s . The larger the value of ϕ_s , the stronger the association between the regions. The choice of shape parameter γ , listed in Table 4.24, is associated with the three different types of baseline distribution. We considered various values in each type category ($\gamma = 0$,

Table 4.23: Examples of the cure rates and the levels of censoring for the simulated datasets

	cure rate		censoring
high cure rate	0.68, 0.67, 0.71, 0.71, 0.74	low censoring	0.39
mid to high cure rate	0.59, 0.60, 0.64, 0.65, 0.67	moderate censoring	0.49
low to mid cure rate	0.27, 0.31, 0.34, 0.37, 0.46	high censoring	0.58
low to mid cure rate	0.27, 0.37, 0.38, 0.40, 0.44	high censoring	0.61

Table 4.24: The settings of the shape parameter, γ , considered for the baseline function

γ	0.000				
Baseline (GEV)	Type I				
γ	1.6	2.000	2.100	2.500	2.700
Baseline (GEV)	Type II	Type II	Type II	Type II	Type II
γ	-1.200	-1.250	-1.500		
Baseline (GEV)	Type III	Type III	Type III		

$\gamma > 0$, $\gamma < 0$) to mimic the possible shape of the baseline distribution.

Censoring time, C_{ij} , is set to follow exponential distribution with parameter cc . We generate Y_{ij} from the quantile function of the GEV distribution with the choice of baseline parameters, μ , σ , and γ . If $Y_{ij} \leq C_{ij}$, we set $\delta_{ij} = 1$, and $\delta_{ij} = 0$ otherwise.

The levels of censoring and cure rate for the five regions are summarized in Table 4.23. The cure rates considered range from low to moderate, moderate to high, and high, and the censoring levels range from low to moderate to high.

To obtain suitable initial values, a grid search based on the observed log-likelihood was conducted. The Stochastic EM algorithm was then used to compute the MLEs of the model parameters, with 1200 iterations performed, the first 200 discarded as burn-in, and a thinning interval of 3 applied.

Table 4.25: The settings of different spatial correlation, ϕ_s , considered for the simulated dataset

ϕ_s	0.500	0.800	1.000	1.200	1.500	2.000
----------	-------	-------	-------	-------	-------	-------

Table 4.26: The different settings of the true values of the model parameters

Parameter	T.V.	T.V.	T.V.	T.V.	T.V.
b_0	0.600	0.880	0.350	0.400	0.400
b_1	0.400	0.880	1.200	0.900	-0.590
β_1	0.600	1.300	0.990	0.900	-0.550
μ	0.200	0.150	0.190	0.200	0.100
σ	0.400	0.350	0.450	0.600	0.100
γ	2.000	2.100	2.700	2.500	0.000
μ_s	-0.800	-0.990	-0.800	-0.800	-0.800
σ_s^2	0.200	0.750	0.600	0.400	0.400
ϕ_s	0.500	2.000	1.500	1.200	0.800
Parameter	T.V.	T.V.	T.V.	T.V.	
b_0	-0.900	0.880	0.880	0.400	
b_1	0.820	0.880	1.000	-0.590	
β_1	0.850	1.300	1.000	-0.550	
μ	0.250	0.150	0.250	0.200	
σ	0.670	0.350	0.600	0.780	
γ	-1.250	0.000	-1.200	-1.500	
μ_s	-0.800	-0.990	-0.990	-0.800	
σ_s^2	0.400	0.750	0.750	0.400	
ϕ_s	1.000	2.000	2.000	0.800	

Results

When the fitted candidate models match the models used to generate the simulated data, the parameter estimates are close to the pre-specified values used in the simulation. In these cases, the results also gave the lowest bias and lowest RMSE compared to the other candidate models fitted to the corresponding simulated datasets. The coverage probabilities of the confidence intervals based on the asymptotic normality of the MLEs are quite close to the nominal levels of 95%. When we fitted other candidate

models that are different from the true models, the obtained coverage probabilities occasionally are lower than the nominal level.

By considering low level of censoring, moderate level of censoring and high level censoring, as well as low cure rate, moderate cure rate and high cure rate, the proposed model performed equally well in all cases. Thus, we conclude that the censoring level and the cure rate do not have much impact our algorithm.

The model parameter, ϕ_s , that associates with spatial correlation, has been captured successfully under the cases of low ($\phi_{ij} = 0.5$, $\phi_{ij} = 0.8$), moderate ($\phi_{ij} = 1$, $\phi_{ij} = 1.2$ and $\phi_{ij} = 1.5$) and high level of ($\phi_{ij} = 2$) spatial correlation setups.

When comparing the result of sample size of 500 and 1000, with similar censoring and level of cure rate for 5 regions, the decrease in bias and RMSE are observed. Thus, for fixed censoring proportion and cure rate, the Bias and RMSE decrease when sample size increases.

When the fitted model differs from the data-generating model, the parameter estimates exhibit slight variation, and the associated biases are modestly higher than those obtained when fitting the true cure model. Nonetheless, the signs of the maximum likelihood estimates remain consistent across all candidate models.

In this section, we focus primarily on results obtained when the fitted model coincides with the data-generating model under various simulation settings. Results from fitting alternative candidate models are presented selectively for illustrative purposes, with the remaining results omitted due to their similarity.

In addition, model discrimination is performed for evaluating the performance of the proposed models. 50 datasets were generated for the choice of parameter settings, where the choice of settings take different levels of strength of spatial effects and three

baseline distributions. Sample size was set to be 1000. As shown in Table 4.29, by using AIC, the selection rate for choosing the correct models are 0.98, 0.94, 0.90 under Models 9 - 11, respectively. The rates of selecting the correct models all being high demonstrate that the proposed models are performed very well.

Table 4.27: The estimated means, bias, and RMSE for selected models, with different choices of shape parameter, $\gamma = (2, 2.1, 2.7)$, and spatial dependency, $\phi_s = (0.5, 1.5, 2)$, when the true model of M_{ij} and the fitted model are the same.

True Model: Model 9						Baseline: $\gamma = 2$	
Fitted Model: Model 9						Spatial: $\phi_s = 0.5$	
n	Censoring (high)	Cure Rate (low to mid cure)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.614	(0.27, 0.31, 0.34, 0.37, 0.46)	b_0	0.600	0.728	0.128	0.182
			b_1	0.400	0.465	0.065	0.197
			β_1	0.600	0.522	-0.078	0.093
			μ	0.200	0.242	0.042	0.046
			σ	0.400	0.505	0.105	0.109
			γ	2.000	2.066	0.066	0.160
			μ_s	-0.800	-0.799	0.001	0.096
			σ_s^2	0.200	0.188	-0.012	0.075
			ϕ_s	0.500	0.503	0.003	0.053
True Model: Model 10						Baseline: $\gamma = 2.1$	
Fitted Model: Model 10						Spatial: $\phi_s = 2$	
n	Censoring (low)	Cure Rate (high cure)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.389	(0.68, 0.67, 0.71, 0.71, 0.74)	b_0	0.880	0.912	0.032	0.070
			b_1	0.880	0.859	-0.021	0.026
			β_1	1.300	1.321	0.021	0.080
			μ	0.150	0.133	-0.017	0.039
			σ	0.350	0.278	-0.072	0.100
			γ	2.100	2.081	-0.019	0.070
			μ_s	-0.990	-1.000	-0.010	0.011
			σ_s^2	0.750	0.740	-0.010	0.011
			ϕ_s	2.000	1.990	-0.010	0.011
True Model: Model 11						Baseline: $\gamma = 2.7$	
Fitted Model: Model 11						Spatial: $\phi_s = 1.5$	
n	Censoring (moderate)	Cure Rate (high cure)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.488	(0.59, 0.6, 0.64, 0.65, 0.67)	b_0	0.350	0.334	-0.016	0.024
			b_1	1.200	1.208	0.008	0.015
			β_1	0.990	1.032	0.042	0.051
			μ	0.190	0.191	0.001	0.039
			σ	0.450	0.486	0.036	0.089
			γ	2.700	2.551	-0.149	0.179
			μ_s	-0.800	-0.799	0.001	0.002
			σ_s^2	0.600	0.599	-0.001	0.004
			ϕ_s	1.500	1.490	-0.010	0.010

Table 4.28: The estimated means, bias, and RMSE for selected models, with different choices of shape parameter, $\gamma = (-1.25, 0, 2.5)$, and spatial dependency, $\phi_s = (0.8, 1, 1.2)$, assuming the true models of M_{ij} and the fitted models are the same.

True Model: Model 5						Baseline: $\gamma = 2.5$	
Fitted Model: Model 5						Spatial: $\phi_s = 1.2$	
n	Censoring	Cure Rate	Par.	T.V.	Estimate	Bias	RMSE
1000	0.584	(0.27, 0.37, 0.38, 0.40, 0.44)	b_0	0.400	0.434	0.034	0.303
	moderate	low to moderate cure	b_1	0.900	0.917	0.017	0.523
			β_1	0.900	0.907	0.007	0.208
			μ	0.200	0.238	0.038	0.069
			σ	0.600	0.564	-0.036	0.153
			γ	2.500	2.351	-0.149	0.262
			μ_s	-0.800	-0.803	-0.003	0.084
			σ_s^2	0.400	0.398	-0.002	0.048
			ϕ_s	1.200	1.201	0.001	0.021
True Model: Model 5						Baseline: $\gamma = 0$	
Fitted Model: Model 5						Spatial: $\phi_s = 0.8$	
n	Censoring	Cure Rate	Par.	T.V.	Estimate	Bias	RMSE
	(low)	(moderate to high)					
1000	0.350	0.50, 0.51, 0.53, 0.58, 0.64	b_0	0.400	0.477	0.077	0.144
			b_1	-0.590	-0.605	-0.015	0.091
			β_1	-0.550	-0.564	-0.014	0.084
			μ	0.100	0.008	-0.092	0.098
			σ	0.100	0.048	-0.052	0.130
			γ	0.000	-0.006	-0.006	0.012
			μ_s	-0.800	-0.795	0.004	0.005
			σ_s^2	0.400	0.395	-0.005	0.006
			ϕ_s	0.800	0.810	0.010	0.010
True Model: Model 9						Baseline: $\gamma = -1.25$	
Fitted Model: Model 9						Spatial: $\phi_s = 1$	
n	Censoring	Cure Rate	Par.	T.V.	Estimate	Bias	RMSE
	(moderate)	(moderate to high)					
1000	0.504	(0.54, 0.54, 0.58, 0.60, 0.63)	b_0	-0.900	-0.874	0.026	0.227
			b_1	0.820	0.815	-0.005	0.150
			β_1	0.850	0.830	-0.020	0.104
			μ	0.250	0.201	-0.049	0.104
			σ	0.670	0.529	-0.141	0.208
			γ	-1.250	-1.217	0.033	0.316
			μ_s	-0.800	-0.800	0.000	0.006
			σ_s^2	0.400	0.401	0.001	0.012
			ϕ_s	1.000	1.000	0.000	0.004

True Model: Model 10						Baseline: $\gamma = 0$	
Fitted Model: Model 10						Spatial: $\phi_s = 2$	
n	Censoring (high)	Cure Rate (moderate)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.599	0.41, 0.433, 0.50, 0.51, 0.54	b_0	0.880	0.896	0.016	0.048
			b_1	0.880	0.956	0.076	0.113
			β_1	1.300	1.250	-0.050	0.207
			μ	0.150	0.122	-0.028	0.078
			σ	0.350	0.368	0.018	0.292
			γ	0.000	-0.077	-0.077	0.098
			μ_s	-0.990	-1.000	-0.010	0.029
			σ_s^2	0.750	0.741	-0.009	0.028
			ϕ_s	2.000	1.990	-0.010	0.026
True Model: Model 10						Baseline: $\gamma = -1.2$	
Fitted Model: Model 10						Spatial: $\phi_s = 2$	
n	Censoring (high)	Cure Rate (moderate)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.662	0.34, 0.35, 0.39, 0.4, 0.45	b_0	0.880	0.915	0.035	0.049
			b_1	1.000	0.994	-0.006	0.008
			β_1	1.000	1.175	0.175	0.201
			μ	0.250	0.248	-0.002	0.010
			σ	0.600	0.554	-0.046	0.001
			γ	-1.200	-1.159	0.041	0.529
			μ_s	-0.990	-1.030	-0.040	0.179
			σ_s^2	0.750	0.741	-0.009	0.043
			ϕ_s	2.000	2.050	0.050	0.224
True Model: Model 11						Baseline: $\gamma = 0$	
Fitted Model: Model 11						Spatial: $\phi_s = 0.8$	
n	Censoring (high)	Cure Rate (moderate)	Par.	T.V.	Estimate	Bias	RMSE
1000	0.409	0.48, 0.50, 0.53, 0.53, 0.54	b_0	0.200	0.187	-0.013	0.099
			b_1	0.880	0.897	0.017	0.017
			β_1	0.880	0.882	0.002	0.104
			μ	0.850	0.844	-0.006	0.079
			σ	1.000	0.996	-0.004	0.078
			γ	0.000	-0.001	-0.001	0.003
			μ_s	-0.400	-0.410	-0.010	0.045
			σ_s^2	0.400	0.410	0.010	0.029
			ϕ_s	0.800	0.810	0.010	0.056

True Model: Model 9										Baseline: $\gamma = -1.25$		
Fitted Model: Model 9										Spatial: $\phi_s = 1$		
n	Censoring (low)	Cure Rate (moderate to high)					Par.	T.V.	Estimate	Bias	RMSE	
500	0.456	0.53	0.58	0.51	0.60	0.67	b_0	-0.900	-0.701	0.199	0.233	
							b_1	0.820	0.862	0.042	0.076	
							β_1	0.850	0.794	-0.056	0.092	
							μ	0.250	0.191	-0.059	0.117	
							σ	0.670	0.496	-0.174	0.210	
							γ	-1.250	-1.356	-0.106	0.140	
							μ_s	-0.800	-0.799	0.001	0.010	
							σ_s^2	0.400	0.404	0.004	0.019	
							ϕ_s	1.000	1.000	0.000	0.006	
True Model: Model 10										Baseline: $\gamma = 2.1$		
Fitted Model: Model 10										Spatial: $\phi_s = 2$		
n	Censoring (moderate)	Cure Rate (moderate to high)					Par.	T.V.	Estimate	Bias	RMSE	
500	0.512	0.52,	0.58,	0.58,	0.58,	0.7	b_0	0.880	0.855	-0.025	0.111	
							b_1	0.880	0.861	-0.019	0.058	
							β_1	1.300	1.353	0.053	0.117	
							μ	0.150	0.138	-0.012	0.035	
							σ	0.350	0.281	-0.069	0.092	
							γ	2.100	2.047	-0.053	0.094	
							μ_s	-0.990	-0.999	-0.009	0.011	
							σ_s^2	0.750	0.741	-0.009	0.014	
							ϕ_s	2.000	1.990	-0.010	0.011	
True Model: Model 11										Baseline: $\gamma = 0$		
Fitted Model: Model 11										Spatial: $\phi_s = 0.8$		
n	Censoring (moderate)	Cure Rate (moderate to high)					Par.	T.V.	Estimate	Bias	RMSE	
500	0.374	0.75	0.66	0.72	0.76	0.75	b_0	0.200	0.234	0.034	0.243	
							b_1	0.880	0.899	0.019	0.051	
							β_1	0.880	0.810	-0.070	0.377	
							μ	0.850	0.853	0.003	0.169	
							σ	1.000	0.971	-0.029	0.163	
							γ	0.000	-0.006	-0.006	0.007	
							μ_s	-0.400	-0.408	-0.008	0.014	
							σ_s^2	0.400	0.412	0.012	0.023	
							ϕ_s	0.800	0.809	0.009	0.015	

Table 4.29: The selection rates for Models 9 - 11

(n = 1000)	P (Model 9)	G (Model 10)	L (Model 11)
P (Model 9)	0.98	0.02	0.00
G (Model 10)	0.04	0.94	0.02
L (Model 11)	0.08	0.02	0.90

4.6 Analysis of smoking cessation data

Signs of the intercept b_0 , gender b_1 , and treatment b_3 are the same for the Bernoulli mixture model with spatial effect (Model 5), the complementary promotion time model with spatial effect (Model 9), the complementary geometric cure rate model with spatial effect (Model 10), and the complementary logarithmic cure rate model with spatial effect (Model 11). The sign of b_2 is positive under Models 5–11 and negative under Model 9.

The estimated empirical means of $\mathbf{b}$ have the lowest standard errors under Model 9. Since $\mathbf{b}$ is related to the cure fraction, these estimates allow more precise inference on the cure rate (e.g., through survival plots).

Signs of β , the covariates for the survival function, are the same across all selected models, and their mean estimates also have the lowest standard errors under Model 9. The estimated means of the baseline parameters and spatial frailties are all close in value, especially for the three complementary models. Overall, the spatial effect is present and is captured successfully.

AIC, BIC, and AICc values are listed in Table 4.32. The complementary promotion time cure rate model with spatial frailties (Model 9) yields the lowest values for all three criteria. Using the optimized parameters from Model 9, we then visualize the spatial impact on the event of interest.

A likelihood ratio test was conducted to compare the null hypothesis, Model 2 (restricted model with M_{ij} following Poisson), against the alternative hypothesis, Model 9 (full model with M_{ij} following Poisson and including spatial effects). The obtained p-value is 0.01, well below the nominal 0.05 threshold, indicating that the model with spatial effects is preferred.

The spatial frailties are visualized in Figure 4.13, where darker red indicates higher positive spatial frailties for patients in those zip codes, and lighter red indicates lower spatial frailties. The top and bottom maps in Figure 4.14 show the differences in cure rate and survival probability when spatial frailties are considered. Darker blue corresponds to a higher positive effect of spatial frailties, while lighter blue corresponds to a higher negative effect on the cure rate and survival probability.

In Figure 4.15, the left and right side of regions are shaded with darker purple colour, which represents high cure rate and high survival rate. The central regions are shaded with light purple, which indicates these regions have lower cure rates and lower survival probabilities. It is clear that the impact of spatial frailties, as shown in Figure 4.15, is consistent with the change of cure rate and survival shown in Figure 4.14.

In conclusion, the central regions exhibit higher positive spatial frailties, which lead to stronger negative differences in cure rates and survival probabilities, resulting in lower values for these outcomes in those regions. In contrast, the lower spatial frailties in the left and right regions correspond to positive differences in cure rates and survival probabilities, leading to higher values for the event of interest in these areas.

For illustration, the averages of cure rates and survival probabilities were used for regions with more than one subject. Parameter evaluation plots for the best model, where the competing cause follows the complementary promotion time cure model with spatial frailties (Model 9), are presented in Appendix C. No upward or downward trends are observed, indicating that the algorithm converged successfully.

Figure 4.13: Demonstration of different spatial frailties for subjects from 51 zip codes in Minnesota

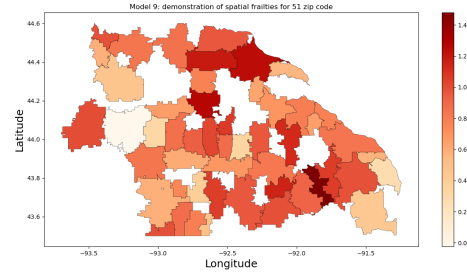


Figure 4.14: Demonstration of different spatial frailties for subjects from 51 zip codes in Minnesota

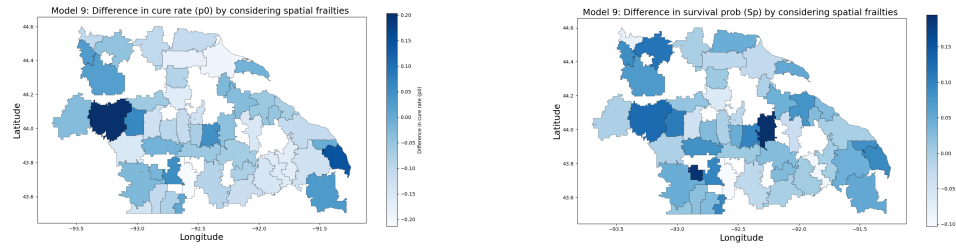


Figure 4.15: Cure rate and survival probabilities when considering spatial frailties for subjects from 51 zip codes in Minnesota

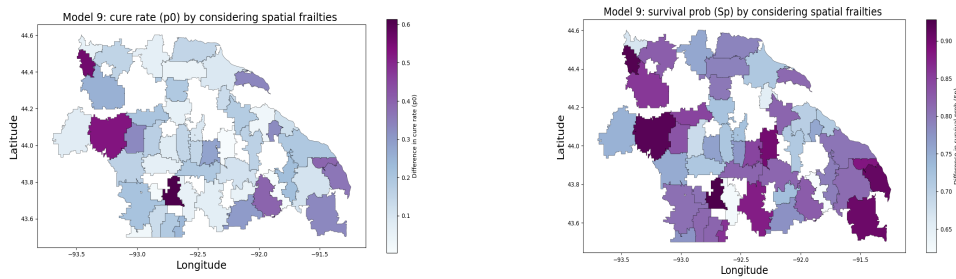


Table 4.30: Estimated mean, standard error, and 95% CI for the case when competing cause follow Bernoulli mixture model (Model 5) and complementary promotion time cure Poisson model (Model 9) with spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

M_{ij}	Parameter	Mean	SE	2.50%	97.50%
Model 5	b_0 (Intercept)	-1.596	0.810	-3.185	-0.008
	b_1 (Gender)	0.368	0.325	-0.270	1.006
	b_2 (Duration)	0.010	0.022	-0.032	0.053
	b_3 (Treatment)	-0.400	0.374	-1.134	0.334
	b_4 (Consumption)	0.065	0.017	0.032	0.097
	β_1 (Gender)	0.463	0.245	-0.018	0.944
	β_2 (Duration)	-0.025	0.011	-0.046	-0.003
	β_3 (Treatment)	-0.340	0.269	-0.867	0.186
	β_4 (Consumption)	0.001	0.010	-0.020	0.022
	μ	3.474	0.658	2.184	4.764
	σ	7.739	1.989	3.840	11.639
	γ	2.838	0.225	2.396	3.279
	μ_s	0.692	0.106	0.485	0.899
	σ_s^2	0.106	0.030	0.047	0.165
	ϕ_s	0.570	0.156	0.264	0.876
Model 9	b_0 (Intercept)	-0.474	0.393	-1.244	0.296
	b_1 (Gender)	0.567	0.194	0.187	0.947
	b_2 (Duration)	-0.038	0.012	-0.062	-0.013
	b_3 (Treatment)	-0.278	0.229	-0.726	0.171
	b_4 (Consumption)	0.050	0.009	0.033	0.067
	β_1 (Gender)	0.361	0.180	0.008	0.713
	β_2 (Duration)	-0.046	0.008	-0.061	-0.030
	β_3 (Treatment)	-0.493	0.185	-0.856	-0.130
	β_4 (Consumption)	0.024	0.007	0.010	0.038
	μ	1.346	0.001	1.345	1.348
	σ	1.989	0.251	1.497	2.481
	γ	3.439	0.441	2.574	4.303
	μ_s	1.058	0.153	0.759	1.358
	σ_s^2	0.190	0.079	0.035	0.345
	ϕ_s	0.360	0.176	0.015	0.706

Table 4.31: Estimated mean, standard error, and 95% CI for the cases when M_{ij} follow the complementary geometric cure rate model (Model 10) and complementary logarithmic cure rate model (Model 11) with spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

M_{ij}	Parameter	Mean	SE	2.50%	97.50%
Model 10	b_0 (Intercept)	-0.978	0.673	-2.296	0.341
	b_1 (Gender)	0.341	0.320	-0.286	0.967
	b_2 (Duration)	0.026	0.020	-0.013	0.066
	b_3 (Treatment)	-0.759	0.392	-1.528	0.009
	b_4 (Consumption)	0.068	0.016	0.036	0.100
	β_1 (Gender)	0.281	0.189	-0.090	0.652
	β_2 (Duration)	-0.007	0.008	-0.023	0.010
	β_3 (Treatment)	-0.550	0.208	-0.958	-0.141
	β_4 (Consumption)	0.026	0.008	0.011	0.041
	μ	1.612	0.010	1.592	1.632
	σ	2.805	0.431	1.960	3.651
	γ	3.321	0.530	2.281	4.360
	μ_s	0.146	0.134	-0.117	0.409
	σ_s^2	0.237	0.063	0.114	0.360
	ϕ_s	0.182	0.070	0.045	0.319
Model 11	b_0 (Intercept)	-1.502	2.453	-6.311	3.306
	b_1 (Gender)	0.867	1.599	-2.268	4.002
	b_2 (Duration)	0.032	0.076	-0.117	0.181
	b_3 (Treatment)	-0.552	1.751	-3.985	2.880
	b_4 (Consumption)	0.227	0.081	0.069	0.386
	β_1 (Gender)	0.277	0.265	-0.242	0.795
	β_2 (Duration)	-0.013	0.010	-0.033	0.007
	β_3 (Treatment)	-0.304	0.289	-0.870	0.261
	β_4 (Consumption)	0.030	0.010	0.011	0.049
	μ	2.218	0.051	2.119	2.317
	σ	4.789	0.546	3.720	5.859
	γ	3.297	0.459	2.398	4.195
	μ_s	0.688	0.110	0.473	0.903
	σ_s^2	0.139	0.042	0.057	0.220
	ϕ_s	0.248	0.124	0.006	0.491

Table 4.32: Negative log-likelihood, AIC, BIC, and AICc, values for the distribution of the competing causes, M_{ij} , following the selected models under the consideration of all possible underlying competing causes and spatial frailties for the smoking cessation data (Iterations: 3000, Burn-in: 500, and Spacing: 3).

Activation	M_{ij}	-ll	AIC	BIC	AICc
Last	Model 9	286.955	603.909	655.017	606.228
	Model 10	324.014	678.029	729.136	680.348
	Model 11	297.351	624.703	675.810	627.022
Random	Model 5	289.060	608.120	659.228	610.439

Survival plots

Figure 4.16: Survival plot of Model 9 stratified by treatments and gender with Duration: 30 years, cigarette consumption: 6/day and 31/day (Mean), and spatial frailty: Rochester, MN (Zip: 55066).

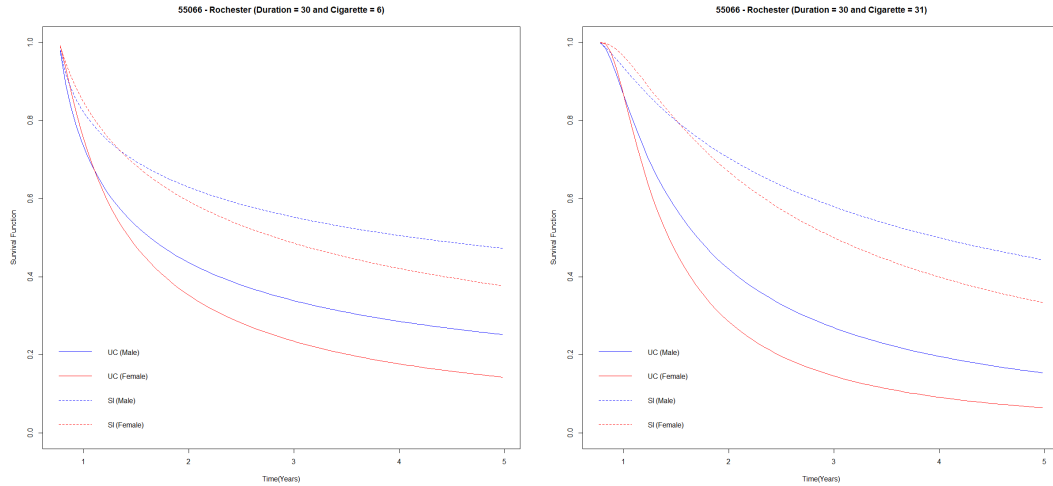
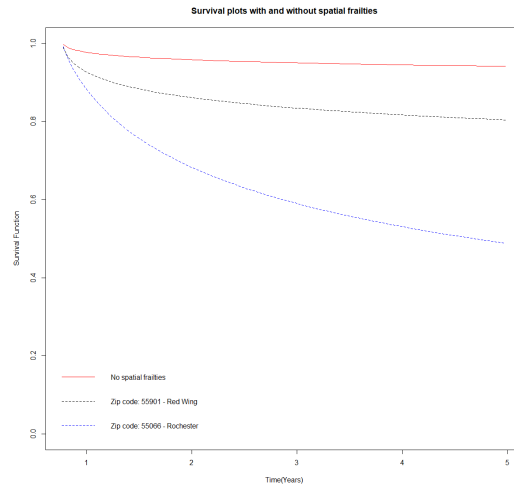


Figure 4.17: Survival plot of Model 9 for a male smoker with smoking habit of 33 yrs and consumption= 20 cigarettes/day



Chapter 5

Concluding Remarks

5.1 Summary of research

In Chapter 2, we adopted the stochastic EM algorithm to a family of cure rate models from a power series cure rate model. These are the Bernoulli cure rate model (Model 1), promotion time cure rate model (Model 2), geometric cure rate model (Model 3), and logarithmic cure rate model (Model 4), with the GEV distribution as baseline distribution. The Stochastic EM algorithm has been developed for finding the optimal estimates of the model parameters. The performance of the proposed models have been evaluated by means of a simulated study along with model discrimination using information-based criteria. The convergence of the developed algorithms for the said models under various settings has been demonstrated. Furthermore, the proposed cure rate models were applied to a real dataset. The resulting cure rates and survival probabilities have been visualized in heat maps. With variation of colours, the heat maps reveal high and low values of cure and survival probabilities among different regions. Next, in Chapter 3, we then extended this work to account for spatial frailties

and examined the spatial effect on cure rate and survival probabilities.

In Chapter 3, the spatial frailties were constructed using a spatial effect assumed to arise from a Gaussian random field (Li and Ryan (2002); Wilson and Wakefield (2020))). The spatial frailties got added to the cure model, which then formed the Bernoulli cure rate model with Gaussian spatial effect (Model 5), promotion time cure rate model with Gaussian spatial effect (Model 6), geometric cure rate model with Gaussian spatial effect (Model 7), and logarithmic cure rate model with Gaussian spatial effect (Model 8). The baseline of the proportional hazard function is fixed to follow the GEV distribution. The stochastic EM algorithm was then got developed and used to find the optimal estimates of the model parameters. The simulation studies for evaluating the performance had subjects coming from 5 specific regions in Minnesota, as shown in Figure 1.1 and Table 1.2. Model discrimination, using information-based criteria, illustrated that the model converge to the true values of parameters under high and low level of censoring, and high and low spatial correlation. The tail behaviour of the baseline got captured by the baseline model, and correct variations of baseline distribution (Type I, Type II, and Type III) get selected. The likelihood ratio test confirmed the improvement of the model with spatial frailties constructed from the spatial information of the subjects (i.e., longitude and latitude). The presence of spatial effect got successfully captured by the proposed spatial cure rate models, and the necessity of adding spatial effect to the model got visualized using heat maps.

While we constructed the cure rate model under the competing risks scenario, it is clear that the event of interest might occur after all of the competing causes have occurred. Unlike in Chapters 2 and 3, where we assumed the event of interest

takes place when the first possible competing cause has presented, in Chapter 4, we extended the work to the case when competing causes are presented, along with the consideration of spatial effect on cure rate and survival functions. Combining it with the spatial effect from a Gaussian random field, three additional spatial cure models were constructed, which are the complementary promotion time cure rate model with Gaussian spatial effect (Model 9), complementary geometric cure rate model with Gaussian spatial effect (Model 10), and complementary logarithmic cure rate model with Gaussian spatial effect (Model 11). Their MLEs of the model parameters were obtained using stochastic EM. The model performance was then evaluated through a simulation study, which was followed by model discrimination. The proposed models were then applied to real data on smoking cessation. The model performance got improved by adding the spatial frailties, and the improvement were demonstrated through model discrimination using information-based criteria. The difference in cure rates and survival probabilities was also demonstrated using maps. Given that the average computation time is approximately 3–5 seconds per iteration, the efficiency of the proposed Stochastic EM algorithm is also clearly demonstrated.

Last but not least, although techniques such as cross-validation can provide empirical support for model selection and validation, they were not explicitly employed in the present study.

5.2 Future work

In this thesis, we have studied the spatial survival for the right-censored cluster data. Nowadays, it is common for individuals to be examined periodically with regard to the recurrence of a certain disease. One may then consider developing spatial survival

analysis for such interval censored lifetime data, and develop suitable models and model fitting methodology in a cure rate setup.

We have demonstrated analysis using a parametric approach through MLEs by the SEM approach. A spatial cure rate model through a semi-parameter approach may be good to develop, where we do not assume any specific baseline hazard function. Then, compared to the work we have done now, where we have assumed the lifetime is from a particular parametric distribution, the semi-parametric approach would provide a more general approach for the analysis to be conducted.

Appendix A

Appendix A for Chapter 2

The log-likelihood functions for Models 1 - 4.

Model 1

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) &= \sum_{i: \delta_i=1} \log f_p(t_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) - p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{\theta_i}{(1 + \theta_i)} f(t_i) \right\} + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \theta_i} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \theta_i}\right) S(t_i) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{(1 + \exp(\mathbf{x}_i^T \mathbf{b}))} \lambda(t_i) S(t_i) \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}\right) S(t_i) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{(1 + \exp(\mathbf{x}_i^T \mathbf{b}))} \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}\right) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right\},
\end{aligned} \tag{A.125}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$ is the model parameter vector.

Model 2

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) &= \sum_{i: \delta_i=1} \log f_p(t_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) - p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \theta_i f(t_i) \exp[-\theta_i(1 - S(t_i))] \right\} + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{\exp(-\theta_i)\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \exp[-\theta_i(1 - S(t_i))] - \exp(-\theta_i) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \theta_i \lambda(t_i) S(t_i) \exp[-\theta_i(1 - S(t_i))] \right\} + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{\exp(-\theta_i)\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \exp[-\theta_i(1 - S(t_i))] - \exp(-\theta_i) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \exp(\mathbf{x}_i^T \mathbf{b}) \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right. \\
&\quad \left. \times \exp[-\exp(\mathbf{x}_i^T \mathbf{b})(1 - S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})})] \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{\exp(-\exp(\mathbf{x}_i^T \mathbf{b}))\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \exp[-\exp(\mathbf{x}_i^T \mathbf{b})(1 - S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})})] - \exp(-\exp(\mathbf{x}_i^T \mathbf{b})) \right\}.
\end{aligned}
\tag{A.126}$$

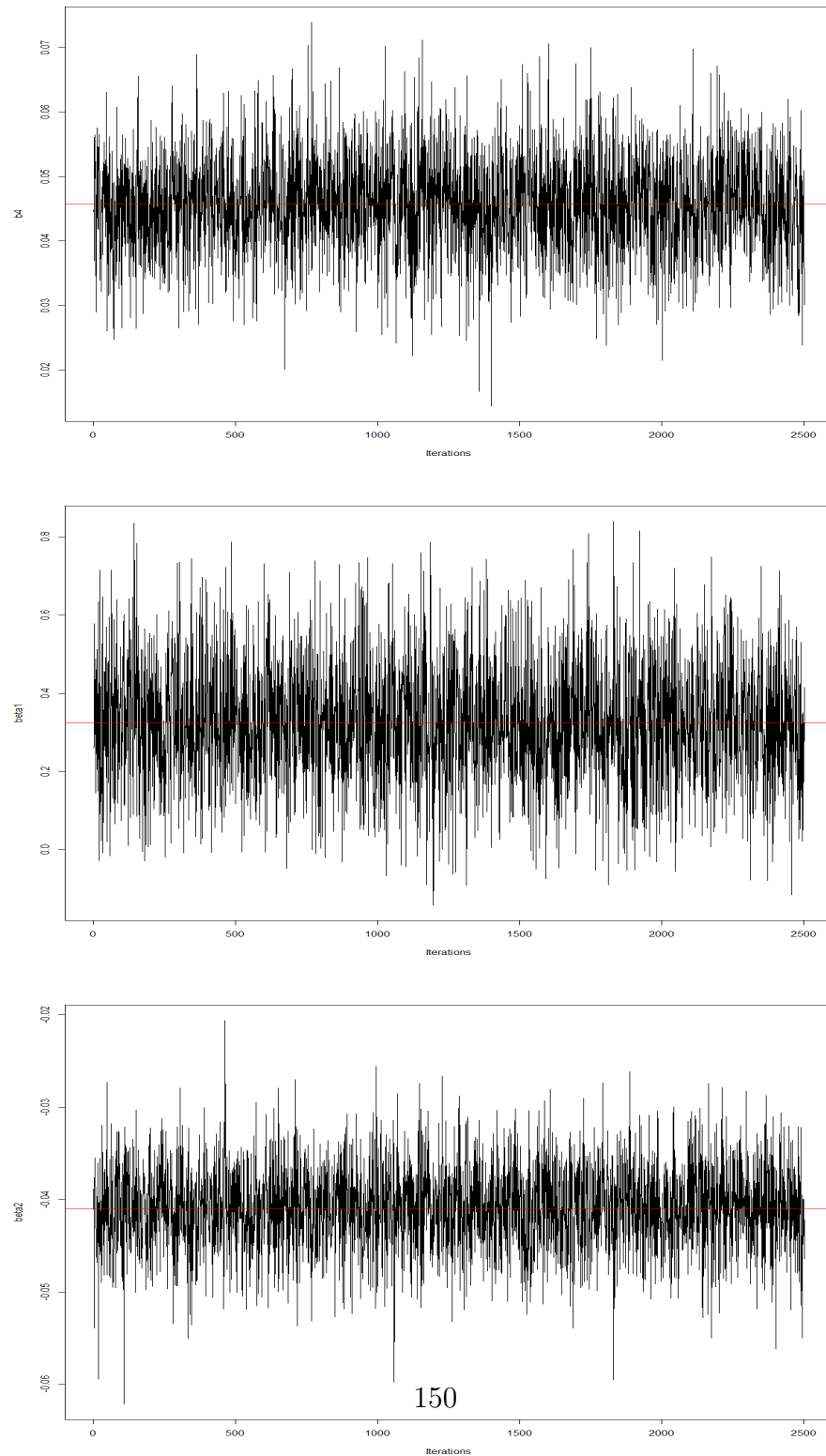
Model 3

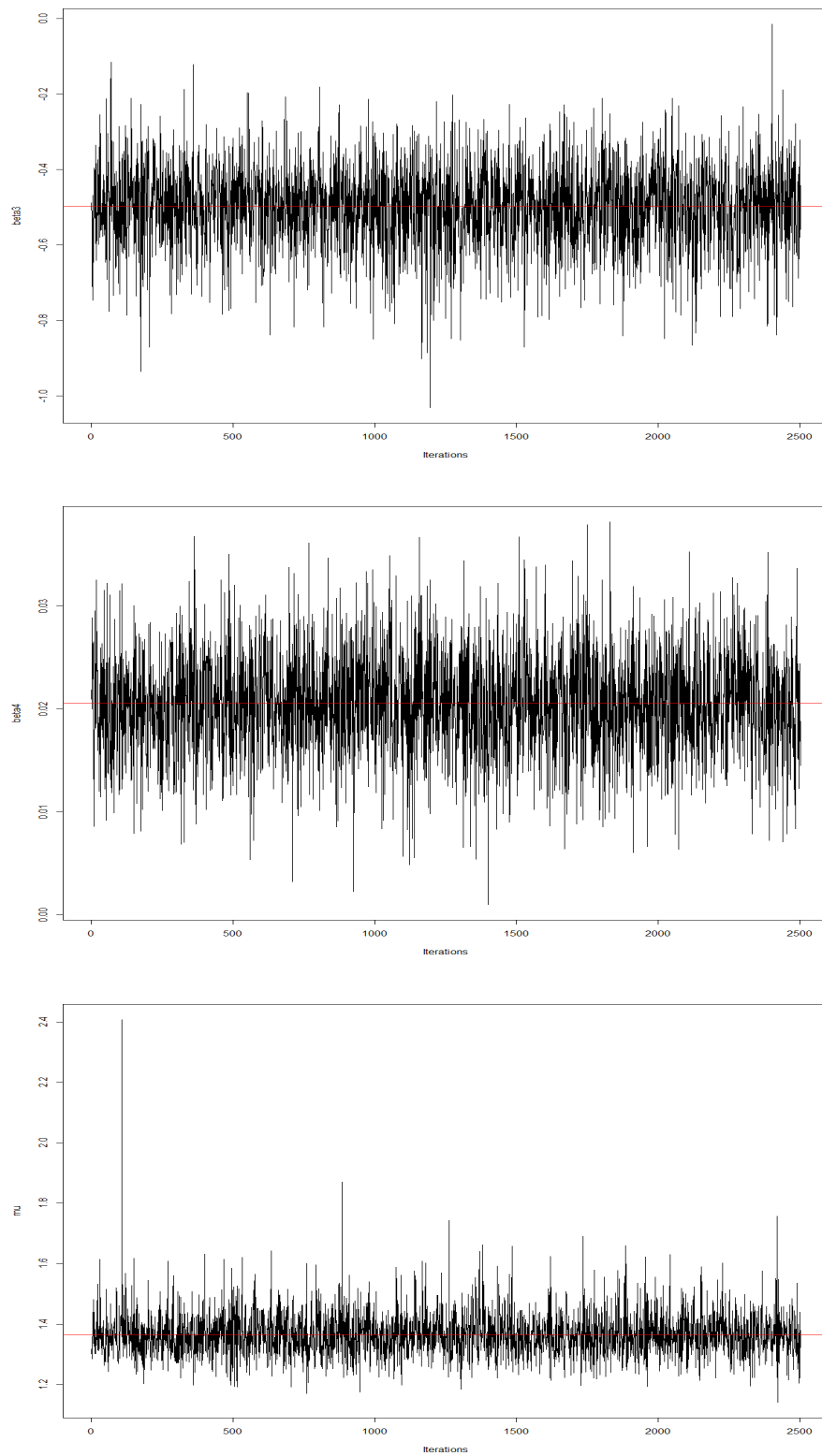
$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) &= \sum_{i: \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) - p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \theta_i (1 - \theta_i) f(t_i) [1 - \theta_i S(t_i)]^{-2} \right\} + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{1 - \theta_i\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{(1 - \theta_i)}{1 - \theta_i S(t_i)} - (1 - \theta_i) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \left(\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right) \lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \right. \\
&\quad \left. \times S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \left[1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right]^{-2} \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i)^{\exp(\mathbf{z}_i^T \boldsymbol{\beta})}} - \left[\frac{1}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} \right] \right\}.
\end{aligned} \tag{A.127}$$

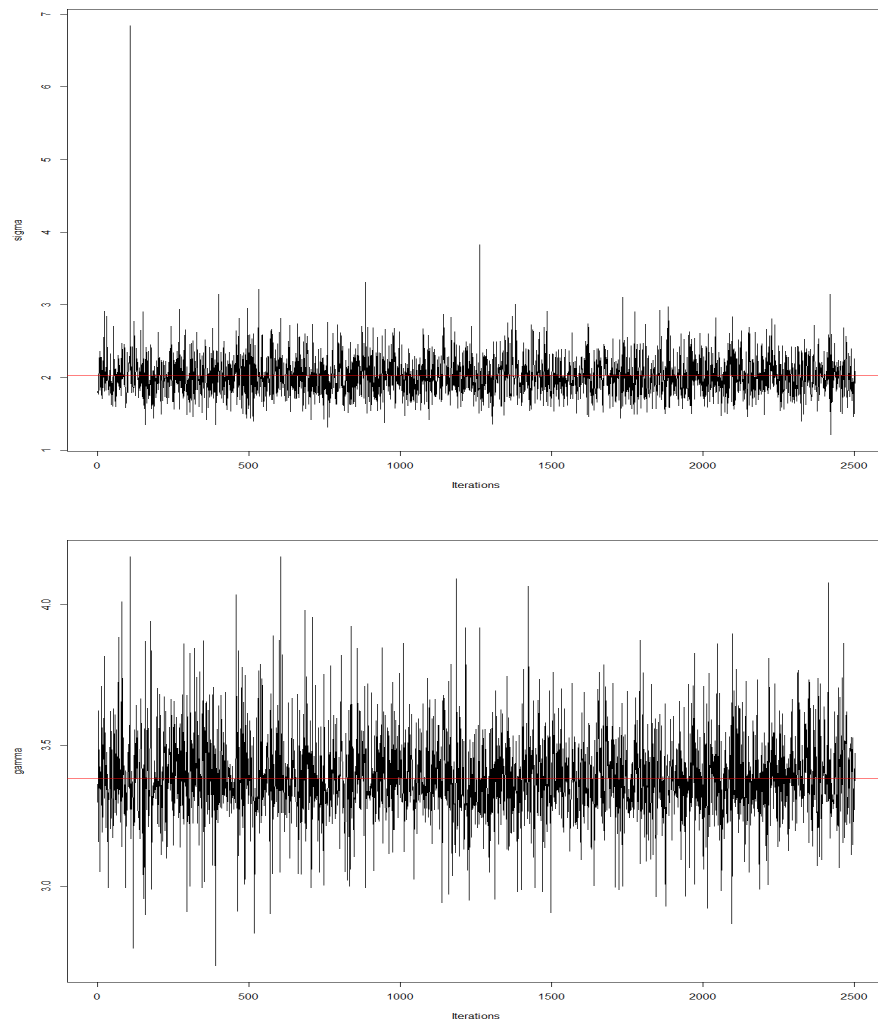
Model 4

$$\begin{aligned}
l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}) &= \sum_{i: \delta_i=1} \log f_p(t_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) + \sum_{i: \delta_i=0} (1 - J_i^*) \log \{p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \{S_p(c_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\xi}) - p_{0i}(\mathbf{x}_i; \mathbf{b})\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{-f(t_i)}{S(t_i) \log(1 - \theta_i)} \left[\frac{\log(1 - \theta_i S(t_i))}{S(t_i)} + \frac{\theta_i}{1 - \theta_i S(t_i)} \right] \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{-\theta_i}{\log(1 - \theta_i)} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{\log(1 - \theta_i S(t_i))}{S(t_i) \log(1 - \theta_i)} - \left(\frac{-\theta_i}{\log(1 - \theta_i)} \right) \right\} \\
&= \sum_{i: \delta_i=1} \log \left\{ \frac{-\lambda_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})}{S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right. \\
&\quad \times \left[\frac{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}))}{S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})} + \frac{\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta})} \right] \left. \right\} \\
&\quad + \sum_{i: \delta_i=0} (1 - J_i^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right\} \\
&\quad + \sum_{i: \delta_i=0} J_i^* \log \left\{ \frac{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})} S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}))}{S_0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} - \left(\frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b})}{1 + \exp(\mathbf{x}_i^T \mathbf{b})})} \right) \right\}.
\end{aligned}
\tag{A.128}$$

Figure A.18: Parameter evolution plots of the SEM algorithm when M_{ij} follow Model 2 for the smoking cessation dataset. (2500 iterations after burn-in period of 500)







Appendix B

Appendix B for Chapter 3

The log-likelihood functions for Models 5 - 8 are as follows.

Model 5

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{\theta_{ij}}{(1 + \theta_{ij})} f(t_{ij}) \right\} + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \theta_{ij}} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \theta_{ij}}\right) S(t_{ij}) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))} \lambda(t_{ij}) S(t_{ij}) \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right) S(t_{ij}) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_i=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{(1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))} \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_i=0} J_{ij}^* \log \left\{ \left(1 - \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}\right) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{B.129}$$

where $\boldsymbol{\xi} = (\mathbf{b}', \boldsymbol{\beta}', \mu, \sigma, \gamma)'$ is the model parameter vector, λ_0 is the baseline hazard function, and S_0 is the baseline survival function.

Model 6

$$\begin{aligned}
& l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \theta_{ij} f(t_{ij}) \exp[-\theta_{ij}(1 - S(t_{ij}))] \right\} + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{\exp(-\theta_{ij})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \exp[-\theta_{ij}(1 - S(t_{ij}))] - \exp(-\theta_{ij}) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \theta_{ij} \lambda(t_{ij}) S(t_{ij}) \exp[-\theta_{ij}(1 - S(t_{ij}))] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{\exp(-\theta_{ij})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \exp[-\theta_{ij}(1 - S(t_{ij}))] - \exp(-\theta_{ij}) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\
&\quad \left. \times \exp[-\exp(\mathbf{x}_i^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{\exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \exp[-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})] \right. \\
&\quad \left. - \exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{B.130}$$

where the power parameter $\theta_{ij} > 0$, S_0 is the baseline survival function, and λ_0 is the baseline hazard function. The baseline hazard and baseline survival functions takes different forms depends on the shape and scale parameters of the GEV distribution.

Model 7

$$l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s)$$

$$\begin{aligned}
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \theta_{ij} (1 - \theta_{ij}) f(t_{ij}) [1 - \theta_{ij} S(t_{ij})]^{-2} \right\} + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{1 - \theta_{ij}\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \frac{(1 - \theta_{ij})}{1 - \theta_{ij} S(t_{ij})} - (1 - \theta_{ij}) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_i) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \left. \times S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right]^{-2} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}} - \left[\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right] \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{B.131}$$

Model 8

$$l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s)$$

$$\begin{aligned}
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{-f(t_{ij})}{S(t_{ij}) \log(1 - \theta_{ij})} \left[\frac{\log(1 - \theta_{ij} S(t_{ij}))}{S(t_{ij})} + \frac{\theta_{ij}}{1 - \theta_{ij} S(t_{ij})} \right] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{-\theta_{ij}}{\log(1 - \theta_{ij})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \frac{\log(1 - \theta_{ij} S(t_{ij}))}{S(t_{ij}) \log(1 - \theta_{ij})} - \left(\frac{-\theta_{ij}}{\log(1 - \theta_{ij})} \right) \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b})}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_i) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right] \Big\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ \frac{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} - \left(\frac{-\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right) \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned}$$

(B.132)

Figure B.19: Parameter evolution plots for b_4 , β_1 and β_2 of the SEM algorithm, when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)

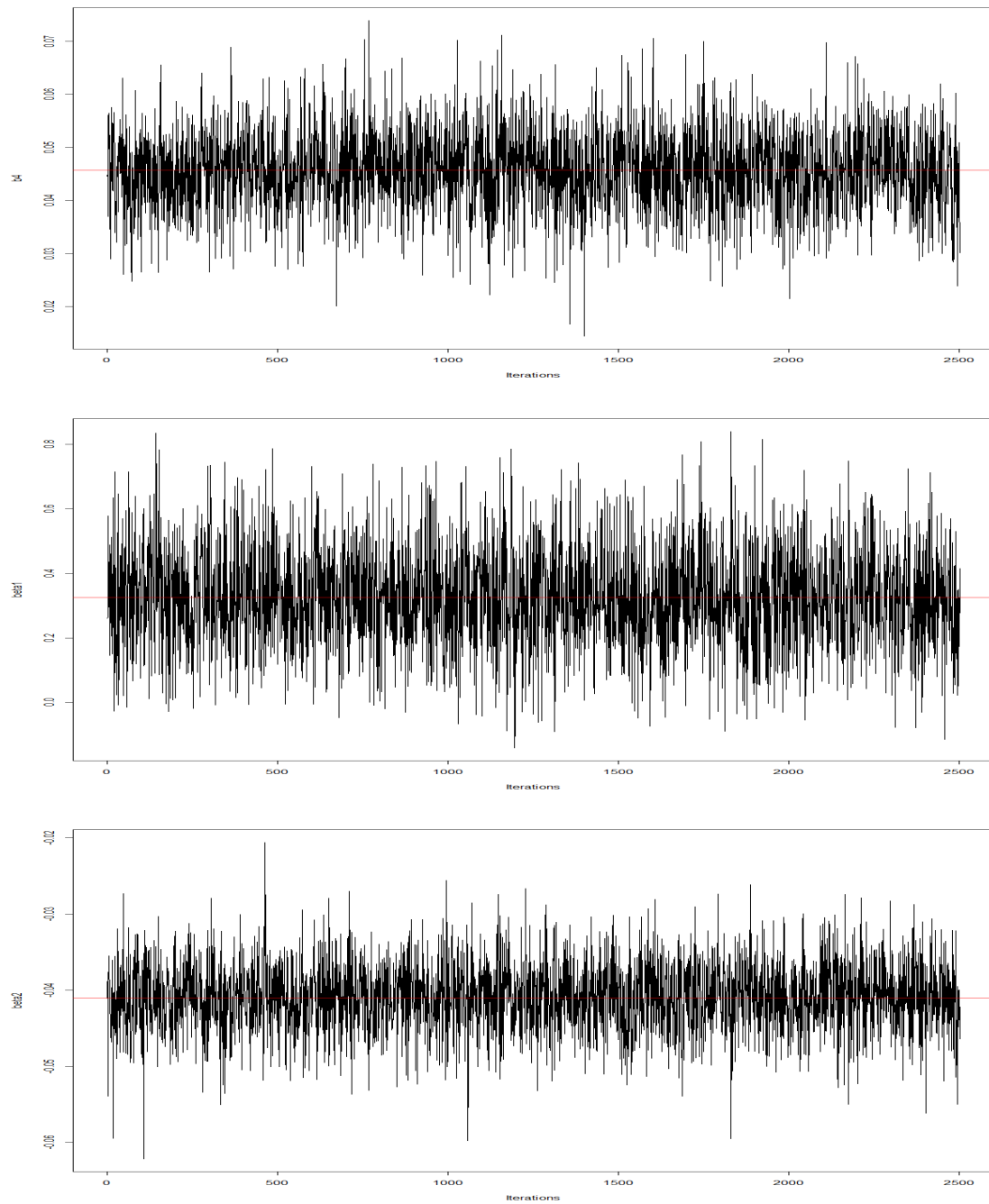


Figure B.20: Parameter evolution plots for β_3 , β_4 and μ of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)

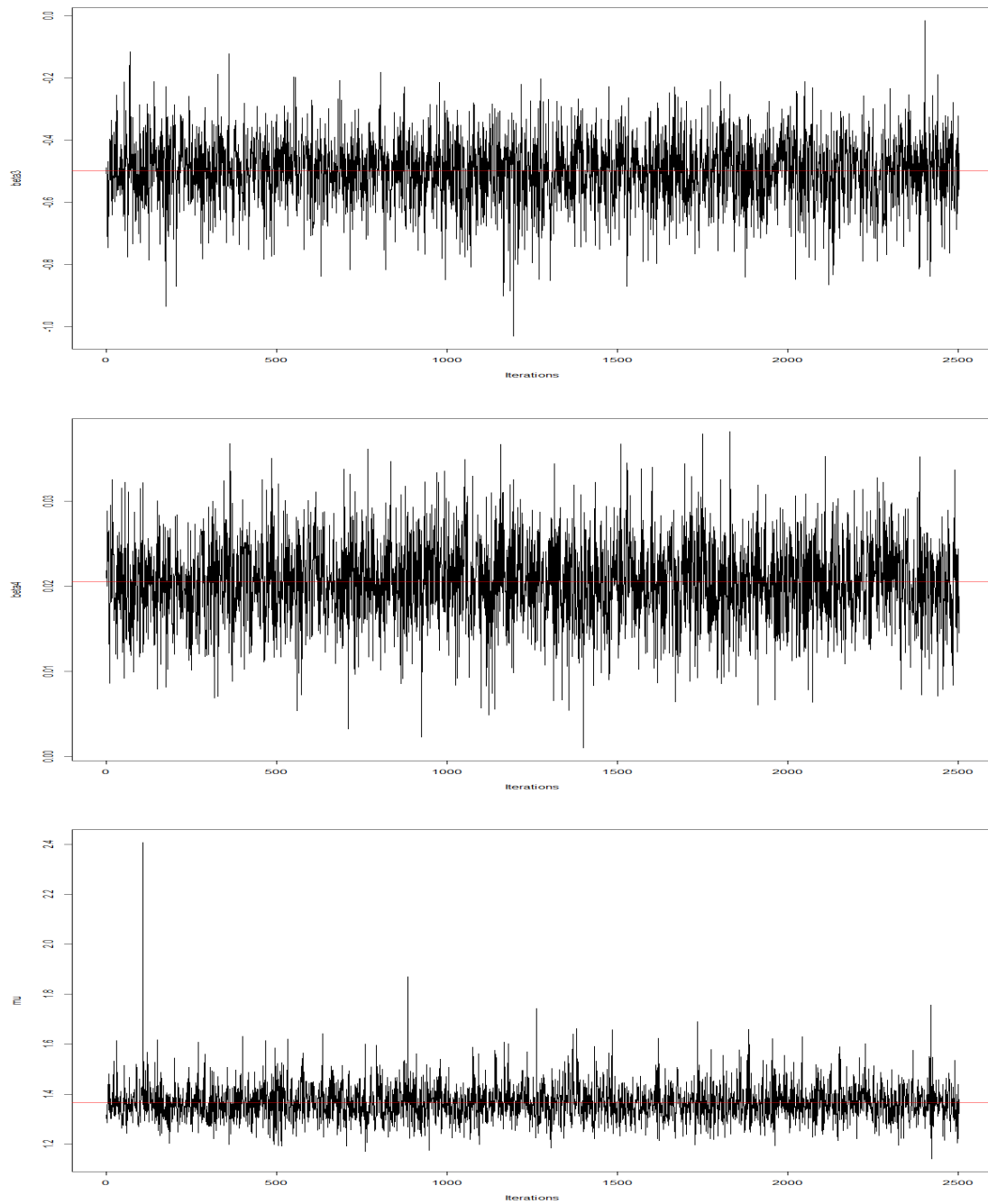
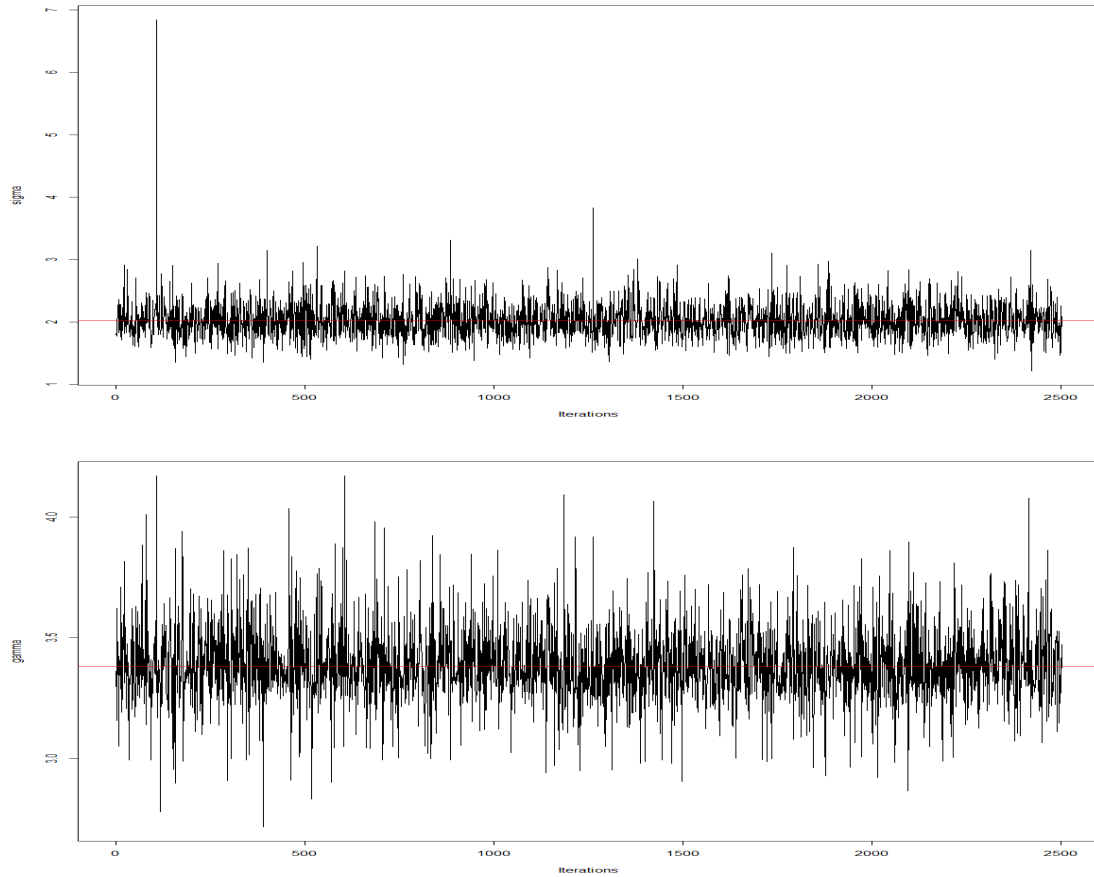


Figure B.21: Parameter evolution plots for σ , γ of the SEM algorithm when M_{ij} follow Model 6. (2500 iterations after the 500 burn-in period)



Appendix C

Appendix C for Chapter 4

The log-likelihood functions for Models 9 - 11 are as follows.

Model 9

$$l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s)$$

$$\begin{aligned}
&= \sum_{(i,j): \delta_{ij}=1} \sum \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \theta_i f(t_{ij}) \exp[-\theta_{ij} S(t_{ij})] \right\} + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{\exp(-\theta_{ij})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_{ij}^* \log \left\{ 1 - \exp[-\theta_{ij} S(t_{ij})] \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \theta_{ij} \lambda(t_{ij}) S(t_{ij}) \exp[-\theta_{ij} S(t_{ij})] \right\} + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{\exp(-\theta_{ij})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_{ij}^* \log \left\{ 1 - \exp[-\theta_{ij} S(t_{ij})] \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \sum \log \left\{ \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_i)^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right. \\
&\quad \left. \times \exp[-\exp(\mathbf{x}_i^T \mathbf{b} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}] \right\} + \sum_{(i,j): \delta_{ij}=0} \sum (1 - J_{ij}^*) \log \{\exp(-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i))\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} \sum J_{ij}^* \log \left\{ 1 - \exp[-\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)}] \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s),
\end{aligned} \tag{C.133}$$

where the baseline survival function, $S_0(t_{ij})$, and the baseline hazard function, $\lambda_0(t_{ij})$, changes to Type I, Type II, and Type III of GEV depending on the shape parameter γ .

Model 10

$$l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s)$$

$$\begin{aligned}
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \theta_i (1 - \theta_i) f(t_i) [1 - \theta_i F(t_i)]^{-2} \right\} + \sum_{(i,j): \delta_{ij}=0} (1 - J_i^*) \log \{1 - \theta_i\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_i^* \log \left\{ 1 - \frac{(1 - \theta_i)}{1 - \theta_i F(t_i)} \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right) \lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) \right. \\
&\quad \left. \times S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \left[1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \left(1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)} \right) \right]^{-2} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_i^*) \log \left\{ \frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_i^* \log \left\{ 1 - \frac{\frac{1}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} (1 - S_0(t_{ij})^{\exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)})} \right\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned}$$

(C.134)

Model 11

$$l_c(\boldsymbol{\xi}; \mathbf{t}, \mathbf{J}^*, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}_s)$$

$$\begin{aligned}
&= \sum_{(i,j): \delta_{ij}=1} \log f_p(t_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}; \boldsymbol{\xi}) + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \{p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \{S_p(c_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}|U_i; \boldsymbol{\xi}) - p_{0ij}(\mathbf{x}_{ij}|U_i; \mathbf{b})\} + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{-f(t_{ij})}{F(t_{ij}) \log(1 - \theta_{ij})} \left[\frac{\log(1 - \theta_{ij} F(t_{ij}))}{F(t_{ij})} + \frac{\theta_{ij}}{1 - \theta_{ij} F(t_{ij})} \right] \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_{ij}^*) \log \left\{ \frac{-\theta_{ij}}{\log(1 - \theta_{ij})} \right\} + \sum_{(i,j): \delta_{ij}=0} J_{ij}^* \log \left\{ 1 - \frac{\log(1 - \theta_{ij} F(t_{ij}))}{F(t_{ij}) \log(1 - \theta_{ij})} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s) \\
&= \sum_{(i,j): \delta_{ij}=1} \log \left\{ \frac{-\lambda_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i) S_0(t_{ij}) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)}{(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)) \log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right. \\
&\quad \times \left[\frac{\log \left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) (1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right)}{1 - S_0(t_{ij}) \exp(\mathbf{z}_i^T \boldsymbol{\beta} + u_i)} + \frac{\frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}}{1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} (1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))} \right] \left. \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} (1 - J_i^*) \log \left\{ \frac{-\frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)}}{\log(1 - \frac{\exp(\mathbf{x}_i^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_i^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{(i,j): \delta_{ij}=0} J_i^* \log \left\{ \frac{\log \left(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i) (1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i))}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)} \right)}{(1 - S_0(t_{ij}) \exp(\mathbf{z}_{ij}^T \boldsymbol{\beta} + u_i)) \log(1 - \frac{\exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)}{1 + \exp(\mathbf{x}_{ij}^T \mathbf{b} + u_i)})} \right\} \\
&\quad + \sum_{i=1}^I \log f_U(u_i; \boldsymbol{\theta}_s).
\end{aligned} \tag{C.135}$$

Figure C.22: Parameter evolution plots for b_0 , b_1 and b_2 of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period)

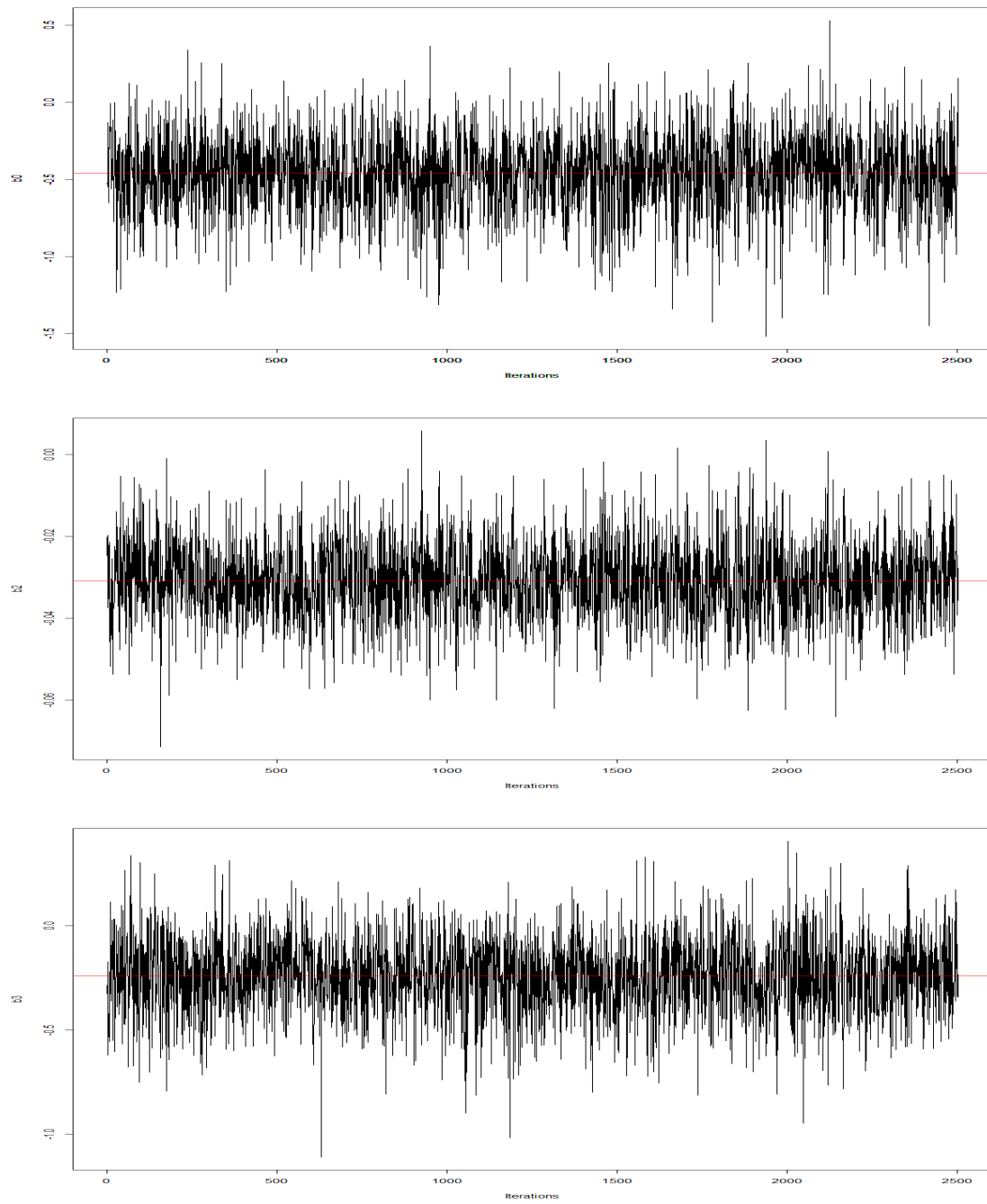


Figure C.23: Parameter evolution plots for b_4 , β_1 and β_2 of the SEM algorithm, under the scenario when all competing causes are present and follow the complementary promotion time cure rate model (Model 9). (2500 iterations after the 500 burn-in period)

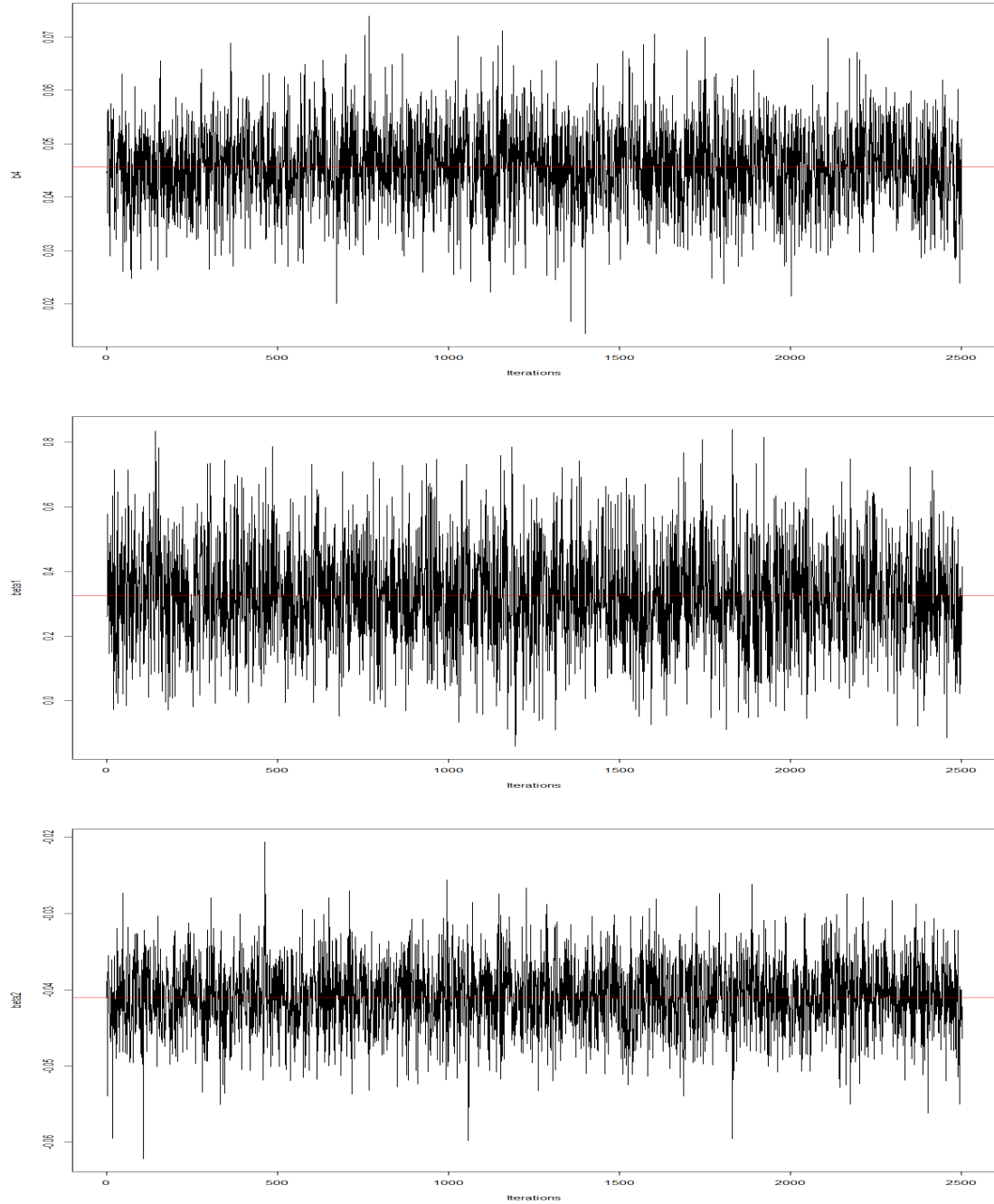


Figure C.24: Parameter evolution plots for β_3 , β_4 and μ of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period)

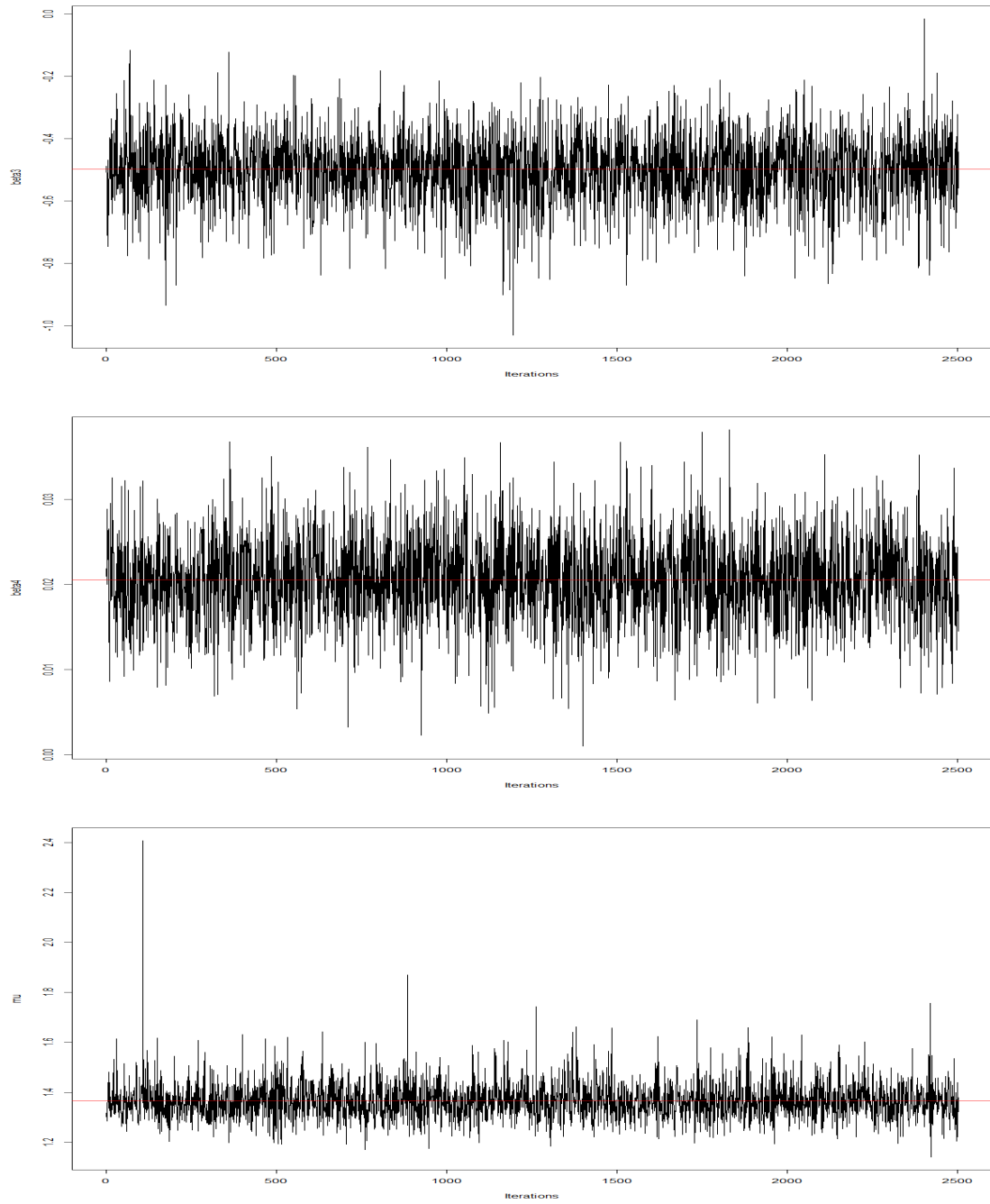
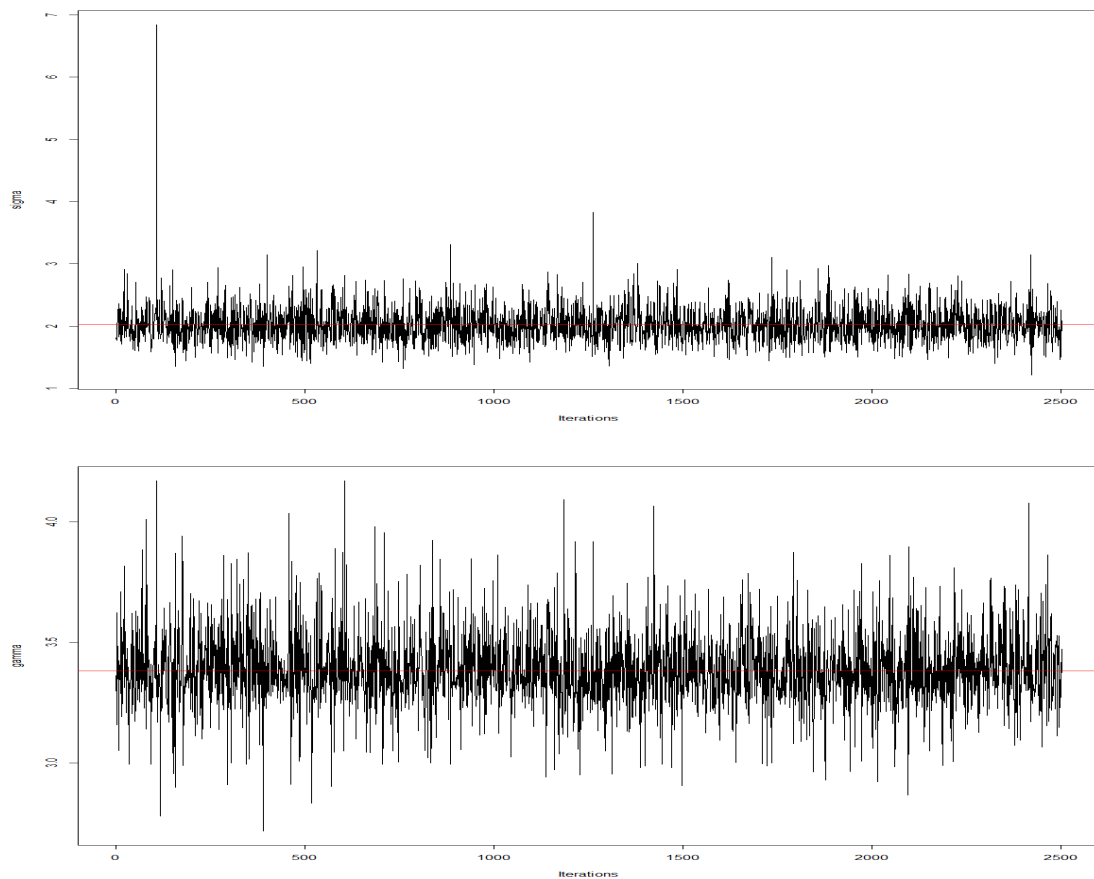


Figure C.25: Parameter evolution plots for σ , γ of the SEM algorithm, when all competing causes are present and follow the complementary promotion time cure rate model. (2500 iterations after the 500 burn-in period)



Bibliography

- Banerjee, S. and Dey, D. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Anal.*, **11** (2), 175–91.
- Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics*, **4** (1), 123–42.
- Bao, Y., Cancho, V. G., Dey, D. K., Balakrishnan, N., and Adriano, K. S. (2020). Power series cure rate model for spatially correlated interval-censored data based on generalized extreme value distribution. *Journal of Computational and Applied Mathematics*, **364**, 112362.
- Celeux, G. and Diebolt, J. (1985). The sem algorithm: A probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistics Quarterly*, **2**, 73–82.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag, London.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, **34**(2), 187–220.

- Diebolt, J. and Celeux, G. (1993). Asymptotic properties of a stochastic em algorithm for estimating mixing proportions. *Stochastic Models*, **9**, 599–613.
- Diebolt, J. and Ip, E. H. (1995). A stochastic em algorithm for approximating the maximum likelihood estimate. Technical report, Sandia National Lab. (SNL-CA), Livermore, CA, USA.
- Hanson, T. E., Jara, A., and Zhao, L. (2012). A Bayesian semiparametric temporally-stratified proportional hazards model with spatial frailties. *Bayesian Analysis*, **7**, 147–188.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. (2001). *Continuous Multivariate Distributions: Volume 1: Models and Applications*. John Wiley and Sons, New York.
- Li, D., Wang, X., and Dey, D. (2016). A flexible cure rate model for spatially correlated survival data based on generalized extreme value distribution and Gaussian process priors. *Biometrical Journal*, **58**, 1178–1197.
- Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, **101**, No. **474**, 591–603.
- Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, **58**, 287–297.
- Martins, R., Silva, G. L., and Andreozzi, V. (2016). Bayesian joint modeling of longitudinal and spatial survival aids data. *Statistics in Medicine*, **30** (19), 3368–84.

- Nielsen, S. F. (2000). The stochastic em algorithm: estimation and asymptotic results. *Bernoulli*, **6**(3), 457–489.
- Noack, A. (1950). A class of random variables with discrete distributions. *Annals of Mathematical Statistics*, **21**(1), 127–132.
- Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Computational Statistics and Data Analysis*, **74**, 198–209.
- R., C. D. and J., S. E. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B (Methodological)*, **30** (2), 248–265.
- Schnell, P., Bandyopadhyay, D., Reich, B. J., and Nunn, M. (2015). A marginal cure rate proportional hazards model for spatial survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **64** (4), 673–691.
- Sugiura, N. (1978). Further analysis of the data by Akaike’s information criterion and the finite corrections: Further analysis of the data by Akaike’ s. *Communications in Statistics - Theory and Methods*, **7**, 13–26.
- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Computational Statistics Quarterly*, **16**(3), 439–454.
- Wilson, K. and Wakefield, J. (2020). Pointless spatial modeling. *Biostatistics*, **21** (2), e17–e32.
- Yakovlev, A. Y., Tsodikov, A. D., and Asselain, B. (1996). *Stochastic Models Of Tumor Latency And Their Biostatistical Applications*. World Scientific, Singapore.

Zhou, H. and Hanson, T. (2018). A unified framework for fitting Bayesian semiparametric models to arbitrarily censored survival data, including spatially referenced data. *Journal of the American Statistical Association*, **113** (522), 571–581.

Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, **9**, 43–68.