



Two-Part Mixed Effects Mixture Model for Zero-Inflated Longitudinal Compositional Data

Viviana A. Rodriguez¹ · Rebecca N. Mahon² · Elisabeth Weiss² · Nitai D. Mukhopadhyay¹

Accepted: 31 March 2024
© The Author(s) 2024

Abstract

Compositional data (CD) is mostly analyzed using ratios of components and log-ratio transformations to apply known multivariable statistical methods. Therefore, CD where some components equal zero represents a problem. Furthermore, when the data is measured longitudinally, and appear to come from different sub-populations, the analysis becomes highly complex. Our objective is to build a statistical model addressing structural zeros in longitudinal CD and apply it to the analysis of radiation-induced lung damage (RILD) over time. We propose a two-part mixed-effects model extended to the case where the non-zero components of the vector might come from a two-component mixture population. Maximum likelihood estimates for fixed effects and variance components were calculated by an approximate Fisher scoring procedure base on sixth-order Laplace approximation. The expectation-maximization (EM) algorithm estimates the mixture model's probability. This model was used to analyze the radiation therapy effect on tissue change in one patient with non-small cell lung cancer (NSCLC), utilizing five CT scans over 24 months. Instead of using voxel-level data, voxels were grouped into larger subvolumes called patches. Each patch's data is a CD vector showing proportions of dense, hazy, or normal tissue. Proposed method performed reasonably for estimation of the fixed effects, and their variability. However, the model produced biased estimates of the nuisance parameters in the model.

Keywords Compositional data · Multivariate longitudinal data · Generalized linear mixed model · Laplace method · Approximate Fisher scoring algorithm

✉ Nitai D. Mukhopadhyay
ndmukhopadhy@vcu.edu

¹ Department of Biostatistics, Virginia Commonwealth University, Richmond, USA

² Department of Radiation Oncology, Virginia Commonwealth University, Richmond, USA

1 Introduction

Compositional data are vectors of non-negative real numbers that represent parts of some whole. Therefore, a composition $\mathbf{x} = [x_1, x_2, \dots, x_D]$ is subject to the constraint $\sum_{i=1}^D x_i = k$, with $x_i > 0$. Usually, the parts of a composition are presented as proportions or percentages (Aitchison, 1986). Common examples include the proportion of different cell types in a patient's blood, proportion of nutrients in a patient's diet, portions of the chemical elements in the air, the proportion of working time spent on different activities, etc.

Compositional data have been commonly analyzed using multivariate data analysis without any transformation. However, compositional data analysis using traditional multivariate data analysis can lead to paradoxes and misinterpretations of the results given the negative bias induced by the constant-sum constraint (Aitchison et al., 2005; Pawlowsky-Glahn et al., 2015). Aitchison (1986) proposed the first methodology with new geometry to analyze compositional data as relative data, using ratios of components and the log transformation that are more amenable to unrestricted multivariate statistical models. However, compositional data, where at least one of the components equals zero, cannot be passed through these ratio transformations. Some methods for handling this issue have been proposed (Aitchison, 1986; Fry et al., 2000; Bacon-Shone, 2003; Martín-Fernández et al., 2003; Aitchison and Kay, 2003; Martín-Fernández et al., 2012; Chen and Li, 2016). The modified Aitchison zero-replacement method is one of the most frequently implemented method in the analysis of compositional data, (Fry et al., 2000) and it works properly in the presence of rounding zeros or generated by detection limits in the measurement. Nonetheless, in the presence of structural zeros (true zeros), the idea of substituting them for a small amount does not seem entirely adequate.

We address the issue in the context of zero inflated longitudinal compositional data that naturally occurs in the lung tissue changes in early-stage lung cancer patients receiving radiation treatment (RT) as captured through CT scans. Lung tissue surrounding the tumor, observed as a composition of different levels of radiation damage over time, poses a challenging modelling problem. Recent work on a similar problem is addressed in Bear and Billheimer (2016), but the context is not longitudinal. Templ et al. (2017) developed a method for outlier and extreme value detection. Silverman et al. (2018) dealt with a very similar inference problem where they used longitudinal compositional data for microbiota using a dynamic linear model without any special accommodation for zero inflation. The context of lung tissue composition we are dealing with is quite different from other scenarios, and the abundance of zero values significantly restricts the direct application of standard methods. Also, the added complexity in this model is the heterogeneity in the longitudinal pattern depicting a split course of temporal change, possibly based on tissue tolerability of radiation therapy. A two-part mixed-effects mixture model for zero-inflated longitudinal compositional data is proposed to deal with abundance of zeros and heterogeneous temporal changes. Modeling and analyzing these changes

leads to our understanding of the responses of lung tissue to high dose radiotherapy, which in turn leads to long term clinical outcomes.

Section 2 briefly outlines the proposed two-part mixed-effect mixture model for zero-inflated longitudinal compositional data. In Section 3, we describe the radiographic lung change process following lung cancer radiotherapy, which partly motivated this research. Section 4 summarizes a simulation study to evaluate the model performance based on data described in Section 3, and the application of the method using the lung cancer data is presented in Section 5. Finally, an overall discussion of the model, its results, usefulness, and areas for further research are presented in Section 6.

2 Two-part Mixed Effects Mixture Model for Zero-Inflated Longitudinal Compositional Data

Consider a compositional data vector of three parts $p_{it} = (d_{it}, h_{it}, n_{it})$, where $i = 1, \dots, N$ (total number of observations), at t (different time points), d_{it} , h_{it} and n_{it} represent the proportion of each of the three parts in the observation i at time t . Further, $0 \leq d_{it} \leq 1$ and $0 \leq h_{it} \leq 1$, and $n_{it} = 1 - d_{it} - h_{it}$. This vector is in the form of compositional data measured over time. Now, let $\mathbf{Y}_{ij} = (d_{ij}/n_{ij}, h_{ij}/n_{ij})$ be a transformation of $p_{it} = (d_{it}, h_{it}, n_{it})$. Also, assume that a large proportion of d_{it} and h_{it} equal zero for most i and t . Hence, modeling this data necessitates a zero inflated longitudinal distribution. Also, the change over time of these parts is rather heterogeneous and thus requires a two-component mixture.

2.1 Model Formulation

Let \mathbf{Y}_{ij} a semicontinuous bivariate vector for the i -th ($i = 1, \dots, N$) observation at time t_{ij} ($j = 1, \dots, n_i$). Then, this outcome vector can be represented by two processes, an occurrence variable U_{ij} and an intensity vector \mathbf{V}_{ij} , where

$$U_{ij} = \begin{cases} 1 & \text{if } \mathbf{Y}_{ij} \neq \mathbf{0} \\ 0 & \text{otherwise} \end{cases}$$

and the intensity variable $\mathbf{V}_{ij} = \log(\mathbf{Y}_{ij})$, when $\mathbf{Y}_{ij} \neq \mathbf{0}$. The log transformation has been used to make \mathbf{V}_{ij} approximately normal with a observation-time-specific mean, and at the same time, handle compositional data. The joint distribution is specified by the distribution of the occurrence variable U_{ij} , and the conditional distribution of \mathbf{V}_{ij} . Suppose U_{ij} follows a random effects logistic regression model:

$$\begin{aligned} \text{logit}\{Pr(U_{ij} = 1 | \mathbf{c}_i)\} &= \text{logit}\{\pi_{ij} | \mathbf{c}_i\} \\ &= \eta_{ij} \\ &= \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{c}_i \end{aligned} \tag{1}$$

where \mathbf{X}_{ij} is a $1 \times p$ covariate vector for the fixed effects, $\boldsymbol{\beta}$ is a $p \times 1$ fixed effects regression coefficient vector, \mathbf{Z}_{ij} is a $1 \times q$ covariate vector for the random effects,

and \mathbf{c}_i is a $q \times 1$ observation-level vector of random intercept and random effects. The vector $\mathbf{V}_{ij} | \mathbf{Y}_{ij} \neq 0$ follows a bivariate linear mixed model:

$$\mathbf{V}_{ij} | (\mathbf{d}_i, \mathbf{Y}_{ij} \neq 0) = \mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i + \boldsymbol{\epsilon}_{ij} \tag{2}$$

where \mathbf{X}_{ij}^* and \mathbf{Z}_{ij}^* might have the same variables as \mathbf{X}_{ij} and \mathbf{Z}_{ij} in (1), but this is not required; $\boldsymbol{\gamma}$ is the regression coefficient matrix ($p^* \times 2$) for the fixed effects; and \mathbf{d}_i is the subject-level matrix with the random effects ($q^* \times 2$). Additionally, the error term

$$\boldsymbol{\epsilon}_{ij} \sim N \left(\mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{Y_1}^2 & \rho \sigma_{Y_1} \sigma_{Y_2} \\ \rho \sigma_{Y_1} \sigma_{Y_2} & \sigma_{Y_2}^2 \end{pmatrix} \right) \tag{3}$$

The random effects vector $(\mathbf{c}_i, \tilde{\mathbf{d}}_i)$, where $\tilde{\mathbf{d}}_i = \text{vec}(\mathbf{d}_i)$, is distributed by a mixture of 2 multivariate normal distributions. Correlation between random effects in (1) and (2) is allowed. Then,

$$\mathbf{b}_i = \begin{bmatrix} \mathbf{c}_i \\ \tilde{\mathbf{d}}_i \end{bmatrix} \sim \sum_{R=1}^2 m_R N \left(\boldsymbol{\mu}_R = \begin{pmatrix} \mathbf{0}_{q \times q} \\ \boldsymbol{\mu}_r \end{pmatrix}, \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{cc} & \boldsymbol{\Psi}_{cd} \\ \boldsymbol{\Psi}_{dc} & \boldsymbol{\Psi}_{dd} \end{pmatrix} \right) \tag{4}$$

with $\sum_{r=1}^2 m_r = 1$ and $\sum_{r=1}^2 m_r \boldsymbol{\mu}_r = \mathbf{0}$. Suppose $\Delta_i = 1$ if \mathbf{b}_i is sampled from the first component in the mixture, and 0 otherwise. Then,

$$\mathbf{c}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}_{cc}) \tag{5}$$

and

$$\tilde{\mathbf{d}}_i | \mathbf{c}_i \sim [\mathbf{m}_1 N(\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})]^{\Delta_i} [(I - \mathbf{m}_1) N(\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})]^{1-\Delta_i} \tag{6}$$

where $\mathbf{H} = \boldsymbol{\Psi}_{dd} - \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \boldsymbol{\Psi}_{cd}$, and $\boldsymbol{\mu}_2 = -\frac{m_1}{1-m_1} \boldsymbol{\mu}_1$.

2.2 The Likelihood

The estimation of the parameters in the model defined by (1) to (6) is based on the maximization of the likelihood, defined as

$$L \propto \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} f(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{c}_i, \tilde{\mathbf{d}}_i, \boldsymbol{\Sigma}) f(\mathbf{c}_i, \tilde{\mathbf{d}}_i | \mathbf{m}_1, \boldsymbol{\mu}_1, \boldsymbol{\Psi}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \tag{7}$$

which, with the two-parts model and the conditional distribution of the random effects, becomes

$$L \propto \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \prod_{j=1}^{n_i} (1 - \pi_{ij})^{(1-U_{ij})} \pi_{ij}^{U_{ij}} [f(\mathbf{V}_{ij} | \boldsymbol{\gamma}, \tilde{\mathbf{d}}_i, \boldsymbol{\Sigma})]^{U_{ij}} \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | \mathbf{m}_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \tag{8}$$

The double integral in (8) is intractable, which implies that the likelihood in the equation cannot be maximized analytically. Using Olsen and Schafer (2001)

approach with a sixth-order Laplace approximation, we obtained an approximate for the log-likelihood. The complete derivation of the likelihood is shown in the Appendix 2.

2.3 Model Estimation

First, the expectation-maximization (EM) algorithm is implemented. In the E-step, the conditional expectation of the likelihood is obtained, and only the posterior probability of belonging to the first component of the mixture for each i -th patch has to be calculated as:

$$m_{1i}(\theta^{(t)}) = \frac{\exp(S_{1i})}{\sum_{r=1}^2 \exp(S_{ri})} \tag{9}$$

where

$$\begin{aligned} S_{ri} = & -\left(n_i^* + \frac{q}{2}\right) \log(2\pi) - \frac{n_i^*}{2} \log|\Sigma| - \frac{1}{2} \log|\mathbf{H}| + \frac{1}{2} \log|\mathbf{B}_i^{-1}| - \frac{1}{2} \log|\Psi_{cc}| \\ & - \frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} m_r \boldsymbol{\mu}_r^T \\ & - \frac{1}{2} m_r \boldsymbol{\mu}_r^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \boldsymbol{\mu}_r + \frac{1}{2} \mathcal{G}_{ir}^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_{ir} + f(\tilde{\mathbf{c}}_{ir}) + \frac{1}{2} \log| - [f^{(2)}(\tilde{\mathbf{c}}_{ir})] \\ & + \log\left(1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ijr}^{(4)} (\mathbf{Z}_{ij} \mathbf{G}_{ir}^{-1} \mathbf{Z}_{ij}^T)^2 - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ijr}^{(6)} (\mathbf{Z}_{ij} \mathbf{G}_{ir}^{-1} \mathbf{Z}_{ij}^T)^3 \right. \\ & \left. + \frac{15}{72} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ijr}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_{ir}^{-1} \mathbf{Z}_{ij}^T\right)^T \mathbf{G}_{ir}^{-1} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ijr}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_{ir}^{-1} \mathbf{Z}_{ij}^T\right)\right) \end{aligned} \tag{10}$$

In the M-step, the conditional expectation is maximized with respect of m_1 , the probability of belonging to the first component of the mixture, which allow us to obtain an updated estimate $m_1^{(t+1)}$ as:

$$m_1^{(t+1)} = \frac{1}{N} \sum_{i=1}^N m_{1i}(\theta^{(t)}) \tag{11}$$

However, the other parameters in the model cannot be updated analytically. Hence an approximate Fisher scoring numerical maximization procedure is implemented. Score vectors were derived using the score vectors proposed in Olsen and Schafer (2001), but ϕ_{cc} , ϕ_H , and ϕ_Σ represent the vectorized upper triangles of Ψ_{cc} , H , and Σ . Additionally, score vectors were extended to accommodate multivariate data. Expressions for the components of the score vectors for the i -th patch are given in the Appendix 3.

At each iteration, the values of the new estimates for the variance and correlation parameters are validated to assess whether the estimates are outside the parameter space. When needed, a step-halving procedure is applied to return the estimates within the parameter space. Also, a second step-halving procedure was implemented for all parameters in the model to guarantee that the deviance of the new set of estimates is lower than in the previous iteration.

Initial values for β , γ , Ψ_{cc} , and the diagonal in Ψ_{dd} are generated by logistic regression for the occurrence variable and independent linear mixed model for each variable in the intensity vector. Starting values for the off-diagonal elements in Ψ_{dd} are set using the diagonal values and a correlation of 0.1. Initial values for Σ are set as the variances and correlations of the variables in the intensity vector. For simplicity, the starting values for m_1 and μ_I are set as 0.55 and $(-0.05, 0.05)$, respectively.

Approximate empirical Bayes estimates by Pareto smoothed (Vehtari et al., 2015) importance sampling for unnormalized densities are implemented for the random effects. A thousand samples $(c_i^{(1)}, \dots, c_i^{(1000)})$ were taken from a multivariate t distribution with 4 degrees of freedom, centered at the mode \tilde{c}_i with covariance matrix proportional to G_i^{-1} , as proposed by Gelman et al. (1995) and implemented by Olsen and Schafer (2001). Using the drawn sample for c_i , estimated means for c_i and d_i were calculated. All estimation procedures have been implemented in R version 3.4.4.R Core Team (2019).

3 Motivating Example: Radiographic Lung Change Following Radiotherapy of Lung Cancer

Radiation therapy (RT) is the standard treatment for medically inoperable patients in the early stages of Non-Small Cell Lung Cancer (NSCLC) (Jain and Berman, 2018). Stereotactic body radiotherapy (SBRT) is a RT approach that requires high spatial accuracy to deliver high doses to the tumor (Linda et al., (2011). Treatment is delivered over several fractions using beams from different angles to ensure the tumor receives the prescribed dose while limiting exposure to surrounding normal tissue (Matsuo et al., 2012). Nevertheless, healthy tissue near the tumor is exposed to clinically relevant doses of RT, causing changes in the lung surrounding the tumor (Dunne et al., 2018). Following RT, computed tomography (CT) scans are obtained at regular time points to assess tumor control and healthy tissue damage which can be classified according to whether consolidations (dense) or ground-glass opacities (hazy) were observed.

Our preliminary data consists of five CT scans of one NSCLC patient obtained at 3, 6, 12, 18, and 24 months following SBRT with a dose of 57 Gy. Voxels on each CT scan were arranged into larger subvolumes, called patches. Each patch is defined using isodose lines, combining voxels that received a uniform dose (Fig. 3a). Boundaries were drawn at regular dose intervals of 0.2 Gy, and voxels with a received treatment dose within the interval were combined on the same patch. Voxels that did not receive any radiation therapy were combined on the same patch. Figure 3b shows a transverse slice of the isodose-patch approach applied to this

patient. The ocean-green region corresponds to voxels that did not receive any RT dose, and the center of the image, where the smallest region is located (dark blue), is the region with the highest received dose. Transition from voxels to the patches was necessary as the temporal integrity of the voxels was very hard to assure, even after a very rigorous registration process. Patients receiving different levels of radiation go through a substantial amount of tumor shrinkage and tissue deformation over time. Therefore, an aggregated approach, using patch level data, is deemed to be more reliable than voxel-level data.

For this study, one physician manually contoured the area with RILD on the CT scans using commercial contouring software (MIM Maestro v6.6, MIM Software, Cleveland, OH). Even though in the CT scan the extent of radiographic injury is observed in a continuous manner with the RT dose, voxels were classified according to their radiographic findings into three-threshold based ordinal categories of radiographic injury, namely, dense (representing areas of consolidation and fibrosis), hazy (representing ground-glass changes) and normal (Linda et al., 2011; Trovo et al., 2010). After RILD contouring, deformable image registration was performed using a multi-pass b-spline for accounting for major changes in lung architecture after RT. Follow-up CT scans were registered to the baseline treatment planning scan using a commercial registration package (Velocity - Varian Medial Systems).

For this patient, there were 288 patches for analysis. Data for each patch is summarized into a vector $p_{it} = (d_{it}, h_{it}, n_{it})$, where $i = 1, \dots, N$ (total number of patches), $t = 3, 6, 12, 18, 24$ months, d_{it} and h_{it} represent the proportion of dense and hazy voxels in the patch i at time t months, and $n_{it} = 1 - d_{it} - h_{it}$. Further, $0 \leq d_{it} \leq 1$ and $0 \leq h_{it} \leq 1$. This vector is in the form of compositional data measured over time, where the unit of observation is a composition within each patch. Even though some spatial autocorrelation is expected in close patches, the proposed model ignored such a relationship to avoid yet more complexity in the model specification.

As shown in Fig. 1, a notable feature of these data is that, at each time point, less than 50% of the patches presented some tissue damage, meaning that d_{it} and h_{it} equal zero for most of the patches. Also, as shown in Fig. 2, the distribution of the transformed compositional data, for those patches with some tissue damage, seems bimodal.

In Section 5, a two-part mixed-effects heterogeneous mixture model for analyzing the aforementioned data is presented. The model was built as an extension of the two-part mixed model proposed in Olsen and Schafer (2001) and adjusted to fit compositional data, bimodality of the temporal change pattern and the mixture distribution of the random effects.

4 Simulations

Simulations, to evaluate the performance of our proposed model, were conducted using the same framework as the real data. Simulated values of the outcome of interest, namely the composition (%Dense, %Hazy, %Normal) was generated in two steps. (Y_{ij}) , for the i -th ($i = 1, \dots, N$) patch at time t_{ij} ($j = 1, \dots, n_i$) was transformed

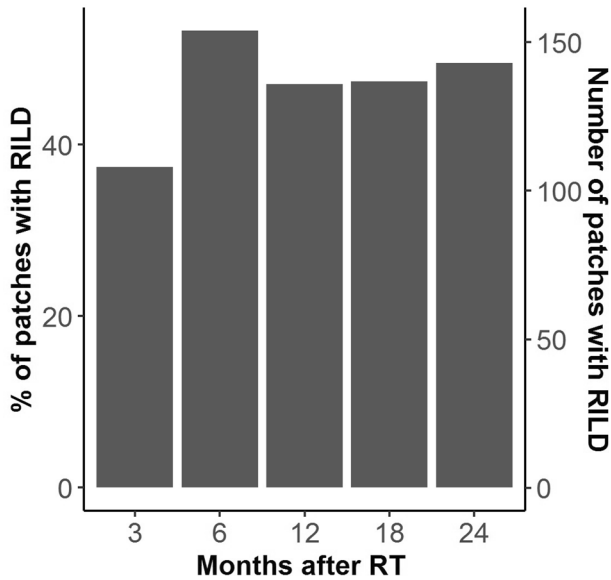


Fig. 1 Bar plot of the observed tissue damage after radiation therapy by follow-up time points

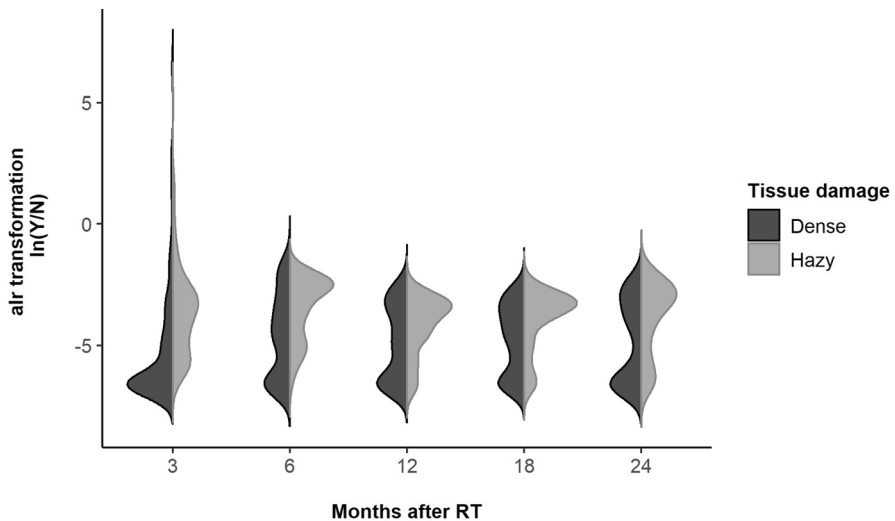


Fig. 2 Violin plots of the kernel density for the proportion of dense and hazy tissue damage after the additive log-ratio transformation (*alt*) by follow-up time points

from the simulated values of the bivariate intensity vector V_{ij} , which is the *alt* transformation of the composition, when $Y_{ij} \neq 0$, and the occurrence variable U_{ij} was generated using a Bernoulli distribution.

The simulation was carried out with three different number of observations ($N = 300, 500, 1000$). Also, three different numbers of follow-up measurements

were considered ($n_i = 3, 4, 5$). The received dose was simulated following a logarithmic distribution. For both parts of the model (occurrence and intensity), intercept, dose-effect, and time after radiotherapy in years (linear and quadratic) were considered as fixed effects to build the data. Only a random intercept was used in both model parts (c_i and d_i for occurrence and intensity models respectively).

The true values for the parameters were chosen to mimic the parameters estimated based on the real data set analyzed in Section 5. For simplicity, true values for the mixture distribution of the random effects for the intensity vector, m_1 and μ_1 , were set as 0.6 and $(-0.1, 0.1)$, respectively. The simulations were repeated 1000 times within each combination of sample size and follow-up time points.

The performance of the algorithm in the estimation of the regression coefficients was assessed by the standardized bias (SB) (difference between the average estimate and the true value as a percentage of the SD estimate), and the coverage rate of the nominal 95% confidence interval (estimate $\pm 1.96SE$). Additionally, the square of the correlation between the observed and predicted data, as well as the mean absolute scaled error (MASE), were calculated for each simulation. As a comparison method, the square of the correlation and MASE were calculated for the “naive” model, using the initial values obtained by a mixed-effects logistic regression for the occurrence variable and independent linear mixed model for each variable in the intensity vector without correlation between the models.

Additionally, to assess the model’s false positive rate of the fixed effects, the same simulation scenario described above was performed but with the coefficients of dose, time, and time squared set equal to zero. Parameter estimates and respective standard errors were calculated using the proposed model, and 95% confidence intervals were obtained. Type I error was calculated as the proportion of intervals that did not contain zero.

For all the simulation scenarios, the procedure failed to converge in less than 5 samples and for those that the procedure converged, the average number of iterations for convergence was between 11 to 14 steps. The average time estimating fixed effects was less than 10 min with 300 patches ($N = 300$), between 10 and 20 min with $N = 500$, and between 35 min and 50 min with $N = 1000$. The estimation of the random effects took longer, around 45 min with $N = 300$, 1 to 1.5 h with $N = 500$, and around 2.5 h with $N = 1000$.

4.1 Estimation of Fixed Effects

For each scenario, Table 1 list the average and the standard deviation (SD) of the point estimates, the average standard error (SE), the standardized bias (SB), the coverage rate, and the type I error. In the occurrence model, the coefficients for the intercept and dose are underestimated, particularly with $N = 300$. In the intensity model, for both $alr(Dense)$ and $alr(Hazy)$, the estimation of the parameters went quite good, even in small sample sizes. The number of time points seems to have a substantial effect on the performance of the estimates in occurrence model, with estimation bias decreasing as the number of time points increases.

Table 1 Simulation results for estimated fixed effects

Population	Occurrence						alr(Dense)						alr(Hazy)					
	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2		
	-10	1.5	-2	-3	-7	1	0.5	-1	-3	2	0.5	-1	-3	2	0.5	-1		
N=1000, t=5																		
Average estimate	-10.38	1.46	-1.93	-3.19	-6.98	1.00	0.49	-0.99	-2.96	1.99	0.48	-0.99	-2.96	1.99	0.48	-0.99	-0.99	
SD estimate	1.32	0.27	1.73	1.34	0.23	0.02	0.52	0.26	0.19	0.02	0.33	0.17	0.19	0.02	0.33	0.17	0.17	
Average SE	1.01	0.07	1.93	0.96	0.22	0.01	0.50	0.30	0.19	0.01	0.33	0.19	0.19	0.01	0.33	0.19	0.19	
Standardized bias (%)	-29.12	-15.86	4.28	-14.50	7.56	-16.86	-2.50	2.24	21.26	-27.98	-6.45	6.20	21.26	-27.98	-6.45	6.20	6.20	
Coverage (%)	87.26	54.26	97.79	95.69	89.77	61.99	88.97	87.96	91.07	63.89	88.26	87.96	91.07	63.89	88.26	87.96	87.96	
Type I error		0.03	0.62	0.22		0.04	0.08	0.11		0.05	0.11	0.12		0.05	0.11	0.12	0.12	
N=1000, t=4																		
Average estimate	-10.29	1.44	-2.09	-3.04	-7.00	0.99	0.50	-1.00	-2.97	1.99	0.50	-1.00	-2.97	1.99	0.50	-1.00	-1.00	
SD estimate	1.17	0.24	1.79	1.14	0.26	0.05	0.69	0.42	0.20	0.02	0.46	0.28	0.20	0.02	0.46	0.28	0.28	
Average SE	1.34	0.13	2.65	1.57	0.21	0.01	0.57	0.35	0.18	0.01	0.39	0.24	0.18	0.01	0.39	0.24	0.24	
Standardized bias (%)	-24.29	-25.28	-5.05	-3.59	0.66	-11.38	-0.02	1.14	17.41	-28.00	-0.71	1.27	17.41	-28.00	-0.71	1.27	1.27	
Coverage (%)	93.76	72.13	98.69	98.09	85.71	63.08	88.13	87.73	89.03	63.58	86.72	86.02	89.03	63.58	86.72	86.02	86.02	
Type I error		0.02	0.42	0.03		0.05	0.08	0.09		0.03	0.11	0.13		0.03	0.11	0.13	0.13	
N=1000, t=3																		
Average estimate	-10.36	1.44	-1.96	-3.17	-6.97	1.00	0.43	-0.94	-2.95	1.99	0.47	-0.98	-2.95	1.99	0.47	-0.98	-0.98	
SD estimate	1.49	0.26	3.40	2.79	0.37	0.02	1.37	1.08	0.29	0.02	0.96	0.76	0.29	0.02	0.96	0.76	0.76	
Average SE	2.32	0.17	6.10	4.65	0.30	0.01	1.10	0.87	0.23	0.01	0.73	0.58	0.23	0.01	0.73	0.58	0.58	
Standardized bias (%)	-24.01	-23.08	1.10	-6.01	8.00	-13.21	-5.14	5.12	18.28	-21.25	-2.87	3.21	18.28	-21.25	-2.87	3.21	3.21	
Coverage (%)	99.70	80.06	99.70	99.90	85.77	60.42	87.07	86.37	84.67	60.92	84.27	83.67	84.67	60.92	84.27	83.67	83.67	
Type I error		0.02	0.02	0.01		0.05	0.11	0.11		0.04	0.14	0.14		0.04	0.14	0.14	0.14	
N=500, t=5																		
Average estimate	-10.52	1.50	-2.12	-3.15	-6.99	1.00	0.49	-1.00	-2.97	1.99	0.52	-1.01	-2.97	1.99	0.52	-1.01	-1.01	

Table 1 continued

Population	Occurrence						air(Dense)						air(Hazy)					
	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2		
	-10	1.5	-2	-3	-7	1	0.5	-1	-3	2	0.5	-1	-3	2	0.5	-1		
SD estimate	1.69	0.31	1.98	1.10	0.23	0.03	0.56	0.27	0.21	0.03	0.42	0.21	0.21	0.03	0.42	0.21		
Average SE	1.27	0.08	2.45	1.27	0.25	0.01	0.57	0.31	0.23	0.01	0.39	0.21	0.23	0.01	0.39	0.21		
Standardized bias (%)	-31.08	-1.12	-6.19	-13.63	3.07	-5.79	-1.34	0.96	12.73	-21.99	3.58	-3.26	12.73	-21.99	3.58	-3.26		
Coverage (%)	86.45	51.71	98.19	95.28	91.47	68.17	90.56	92.17	90.56	68.98	86.04	87.65	90.56	68.98	86.04	87.65		
Type I error		0.02	0.87	0.13		0.04	0.06	0.08		0.05	0.07	0.10		0.05	0.07	0.10		
N=500, t=4																		
Average estimate	-10.58	1.50	-2.01	-3.24	-6.98	1.00	0.48	-0.99	-2.96	2.00	0.48	-0.98	-2.96	2.00	0.48	-0.98		
SD estimate	1.85	0.30	2.42	1.65	0.29	0.02	0.82	0.49	0.24	0.02	0.58	0.34	0.24	0.02	0.58	0.34		
Average SE	2.02	0.18	4.13	2.62	0.28	0.01	0.80	0.50	0.24	0.01	0.53	0.34	0.24	0.01	0.53	0.34		
Standardized bias (%)	-31.51	1.63	-0.34	-14.86	6.89	-13.57	-2.55	2.17	16.12	-17.04	-3.11	4.48	16.12	-17.04	-3.11	4.48		
Coverage (%)	94.98	74.22	99.30	98.29	88.67	65.60	90.67	91.37	91.37	69.01	88.37	88.16	91.37	69.01	88.37	88.16		
Type I error		0.03	0.50	0.03		0.05	0.06	0.08		0.04	0.11	0.11		0.04	0.11	0.11		
N=500, t=3																		
Average estimate	-10.70	1.51	-2.13	-3.18	-7.00	1.00	0.48	-0.98	-2.99	2.00	0.53	-1.02	-2.99	2.00	0.53	-1.02		
SD estimate	2.82	0.46	4.78	3.69	0.42	0.02	1.58	1.24	0.35	0.02	1.23	0.97	0.35	0.02	1.23	0.97		
Average SE	3.35	0.26	8.47	6.43	0.39	0.01	1.47	1.17	0.31	0.01	1.01	0.79	0.31	0.01	1.01	0.79		
Standardized bias (%)	-24.72	2.67	-2.63	-4.96	-0.06	-7.74	-0.97	1.61	3.79	-20.72	2.39	-2.44	3.79	-20.72	2.39	-2.44		
Coverage (%)	99.60	86.83	99.30	99.30	90.55	65.03	91.06	91.16	89.75	68.44	87.04	87.04	89.75	68.44	87.04	87.04		
Type I error		0.01	0.03	0.01		0.03	0.08	0.09		0.04	0.12	0.11		0.04	0.12	0.11		
N=300, t=5																		
Average estimate	-11.23	1.14	-2.30	-3.70	-6.99	1.00	0.52	-1.01	-2.99	2.00	0.53	-1.01	-2.99	2.00	0.53	-1.01		
SD estimate	7.97	9.19	6.74	8.47	0.28	0.03	0.62	0.30	0.24	0.03	0.43	0.20	0.24	0.03	0.43	0.20		
Average SE	1.61	0.09	3.06	1.49	0.30	0.01	0.69	0.37	0.27	0.02	0.47	0.25	0.27	0.02	0.47	0.25		

Table 1 continued

Population	Occurrence						alr(Dense)						alr(Hazy)					
	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2	Int	Dose	Time	Time2		
	-10	1.5	-2	-3	-7	1	0.5	-1	-3	2	0.5	-1	-3	2	0.5	-1		
Standardized bias (%)	-15.47	-3.90	-4.51	-8.28	3.22	-1.78	3.21	-3.31	4.65	-9.14	6.66	-6.91	4.65	-9.14	6.66	-6.91		
Coverage (%)	88.14	49.05	97.99	95.98	91.56	64.82	92.76	92.96	93.87	70.05	92.56	92.56	93.87	70.05	92.56	92.56		
Type I error	0.02	0.02	0.84	0.09	0.05	0.05	0.04	0.06	0.04	0.04	0.06	0.09	0.04	0.04	0.06	0.09		
N=300, t=4																		
Average estimate	-11.41	0.97	-2.74	-3.69	-6.99	1.00	0.50	-1.00	-2.96	2.00	0.51	-1.01	-2.96	2.00	0.51	-1.01		
SD estimate	14.11	12.18	12.04	10.56	0.31	0.02	0.88	0.53	0.26	0.02	0.63	0.38	0.26	0.02	0.63	0.38		
Average SE	2.26	0.20	4.15	2.47	0.33	0.01	0.92	0.57	0.27	0.01	0.62	0.39	0.27	0.01	0.62	0.39		
Standardized bias (%)	-9.97	-4.32	-6.14	-6.50	4.36	-5.79	-0.31	0.21	15.94	-15.68	2.10	-1.62	15.94	-15.68	2.10	-1.62		
Coverage (%)	95.18	74.17	98.99	98.69	93.57	68.54	93.87	93.87	90.85	71.66	89.45	90.65	90.85	71.66	89.45	90.65		
Type I error	0.03	0.03	0.27	0.03	0.05	0.05	0.06	0.07	0.05	0.05	0.06	0.09	0.05	0.05	0.06	0.09		
N=300, t=3																		
Average estimate	-13.22	0.34	-3.38	-4.12	-7.01	1.00	0.60	-1.08	-2.97	2.00	0.46	-0.96	-2.97	2.00	0.46	-0.96		
SD estimate	25.30	19.33	17.48	11.20	0.51	0.03	1.88	1.48	0.38	0.03	1.38	1.08	0.38	0.03	1.38	1.08		
Average SE	4.81	0.35	12.08	9.11	0.50	0.02	1.89	1.50	0.38	0.02	1.26	1.00	0.38	0.02	1.26	1.00		
Standardized bias (%)	-12.71	-6.02	-7.88	-9.97	-2.43	-11.28	5.19	-5.22	8.91	-17.41	-2.99	3.70	8.91	-17.41	-2.99	3.70		
Coverage (%)	98.09	90.24	98.09	97.79	92.25	67.10	92.45	92.45	91.95	69.11	89.84	90.74	91.95	69.11	89.84	90.74		
Type I error	0.02	0.02	0.01	0.02	0.04	0.04	0.06	0.08	0.04	0.04	0.08	0.09	0.04	0.04	0.08	0.09		

Summary of the standardized bias (SB) demonstrates that it is consistently lower than 40% for all parameters, therefore bias impact on the efficiency, coverage and error rate with the proposed model is minimal. However, in some scenarios, the SB is larger than 20%, particularly for the estimation of the intercepts and the dose effect. In most of the cases, the average SE is close to the SD of the estimate; however, with $N = 300$, the average SE and the SD differ substantially. This might indicate that the estimation of the standard errors generated by the proposed method may not adequately represent the variation of the estimator in small sample sizes. Coverage percentages were close to 90% or larger for all parameters but the dose coefficient. Contrary to our common experience, relatively higher coverage rates were observed with the smallest sample size ($N = 300$). This is due to large standard errors resulting from small sample size, not better estimation performance.

The type I error rate for the coefficient associated with dose was equal or less than 0.05 for both model parts for all sample sizes (N) and the number of follow-up time points (n_i). The coefficient for time showed the highest type I error among the model coefficients, particularly for the first part of the model (occurrence model). This issue is a result of the underestimation of that coefficient, which gets worse as the sample size (N) increases. For the coefficient associated with Time², the type I error rate is larger than 0.05 for most of the scenarios. Additionally, an increase in the error is observed with larger sample sizes and number of time points; this is a result of higher precision of the estimates with larger number of time points that led to smaller confidence interval.

4.2 Estimation of Nuisance Parameters

In addition to the estimates of the fixed effects, the proposed model also performs the estimation of the variance components (Σ and Ψ) and the components associated with the mixture model (m and μ). The results obtained in the simulation for these parameters are observed in the Table 2. All the variance components but Ψ_{12} are underestimated and strongly biased. This bias is basically caused by the inclusion of the step-halving procedure that forces a lower deviance in each iteration, not allowing a large and faster change in some parameters. However, the step-halving procedure is necessary and widely used to deal with common convergence problems when using non-standard link functions, such as the log link binomial model used in the occurrence model.

The model performance on the estimation of the parameters of the mixture model was not satisfactory. The average percentage of membership to the most frequent part of the mixture (m) has been shrunk towards 0.50, and the mean towards zero, indicating a no mixture model. This is attributable to our assumed true model parameters that leads to little difference between the two groups, however, it imitates the real data. Even though the nuisance parameters are usually not the main interest in research, they are used on the estimation of the random effects, and therefore will affect the model's predictions, and are also used on the hypothesis testing involving these parameters to find a more parsimonious model.

Table 2 Simulation results for estimated nuisance parameters

Population	Σ_{11} 1	Σ_{22} 0.5	Cor_{12} 0.2	Ψ_{11} 1	Ψ_{12} 0.07	Ψ_{13} 0.2	Ψ_{22} 0.5	Ψ_{33} 1	Ψ_{23} 0.07	m 0.6	μ_1 -0.1	μ_2 0.1
N=1000, t=5												
Average estimate	0.74	0.35	0.59	0.75	0.07	0.11	0.29	0.55	-0.07	0.51	0.00	0.00
SD estimate	0.19	0.14	0.30	0.72	0.05	0.06	0.12	0.13	0.09	0.01	0.01	0.02
N=1000, t=4												
Average estimate	0.74	0.35	0.62	0.71	0.07	0.11	0.27	0.55	-0.06	0.51	0.00	0.00
SD estimate	0.19	0.13	0.30	0.64	0.04	0.06	0.11	0.12	0.09	0.01	0.01	0.01
N=1000, t=3												
Average estimate	0.74	0.33	0.69	0.75	0.07	0.11	0.27	0.53	-0.05	0.51	0.00	0.00
SD estimate	0.19	0.12	0.27	0.81	0.04	0.06	0.11	0.11	0.08	0.01	0.01	0.01
N=500, t=5												
Average estimate	0.81	0.38	0.57	0.86	0.07	0.11	0.27	0.54	-0.05	0.51	-0.01	0.00
SD estimate	0.19	0.14	0.28	1.62	0.05	0.07	0.12	0.16	0.09	0.01	0.01	0.01
N=500, t=4												
Average estimate	0.80	0.37	0.58	0.83	0.07	0.11	0.27	0.53	-0.04	0.51	-0.01	0.00
SD estimate	0.20	0.14	0.28	1.03	0.06	0.08	0.13	0.14	0.09	0.01	0.01	0.01
N=500, t=3												
Average estimate	0.81	0.37	0.62	0.92	0.07	0.12	0.27	0.53	-0.04	0.51	-0.01	0.00
SD estimate	0.20	0.14	0.26	1.47	0.10	0.10	0.14	0.15	0.09	0.01	0.01	0.01
N=300, t=5												
Average estimate	0.86	0.40	0.61	1.27	0.08	0.11	0.27	0.53	-0.03	0.51	-0.01	0.01
SD estimate	0.20	0.14	0.24	7.02	0.14	0.19	0.13	0.15	0.09	0.01	0.01	0.01
N=300, t=4												
Average estimate	0.86	0.39	0.61	1.49	0.08	0.12	0.27	0.53	-0.03	0.51	-0.01	0.01
SD estimate	0.20	0.13	0.24	9.93	0.27	0.21	0.15	0.15	0.08	0.01	0.01	0.01
N=300, t=3												
Average estimate	0.88	0.40	0.64	2.91	0.08	0.14	0.27	0.53	-0.03	0.51	-0.01	0.01
SD estimate	0.21	0.14	0.24	21.3	0.18	0.32	0.16	0.17	0.08	0.01	0.01	0.01

The average squared correlation between observed and predicted values was at least 0.70 for the proposed two-part model, and it was less than 0.40 in the naive approach. Similarly, the two-part model also showed better performance than the naive model when comparing the MASE, with an average MASE lower than 0.71

and 0.58 for dense and hazy, respectively, in the two-part model, and larger than 0.90 in the naive model.

5 Application to Real Data

The two-part mixed-effects mixture model for zero-inflated longitudinal compositional data was applied to patient’s data described in Section 3. As mentioned, the RILD intensity on each patch was consolidated in the compositional vector $p_{it} = (\%Dense, \%Hazy, \%Normal)$. Then, each patch was classified as without RILD presence ($\%Normal=100$), or with RILD presence ($\%Normal \neq 100$).

For applying the proposed model, the outcome variable ($Y_{ij}; i = 1, \dots$, number of patches; $j = 1, \dots, 5$) was the ratio of the vector p_{it} , using the last component (Normal) as denominator ($Y_{ij}=(\%Dense/\%Normal, \%Hazy/\%Normal)$). When the whole patch presented RILD ($\%Normal=0$), or when only one of the damage components, dense or hazy, equals zero, the zero replacement method proposed by Fry et al. (2000) with $\delta = 0.003$ was applied before calculating the *alr* transformation. Therefore, $Y_{ij} = \mathbf{0}$ represents no observed tissue damage after radiation in the patch i at the time point j ($\%Dense=0, \%Hazy=0, \%Normal=100$).

As commented on Section 3 (Fig. 1), there is a large proportion of zeros on these data set. Hence, the occurrence variable is defined as

$$U_{ij} = \begin{cases} 1 & \text{If RILD is observed in the patch } i \text{ at the time point } j \\ 0 & \text{otherwise} \end{cases}$$

and the intensity variable $V_{ij} = \log(Y_{ij})$, when $Y_{ij} \neq \mathbf{0}$. Then, V_{ij} is the *alr* transformation of the compositional vector ($\%Dense, \%Hazy, \%Normal$). Median patch dose, time in years, and time squared were used as explanatory variables for modeling the fixed effects for both outcomes (occurrence and intensity). To account for the interdependence over time, a random intercept was used in both models. No other factors were considered as random effects. Iterative estimation was performed until convergence was attained (tolerance = 1e-8). Table 3 shows the estimates of the fixed effects in the model.

5.1 Results

Images of this patient rendered 288 isodose patches, of which 58.48% present RILD at some point after RT. The model converged after 6 iterations, which took 8.1 min to a maximum relative parameter change of 1.2E-13. Also, the estimation of the random effects took 30 min. The odds of RILD, for a particular patch, increased 26% for each Gy increase in the dose (Table 3). Likewise, the odds of RILD increase until 17 months after RT, then the odds decrease over time. In the results for the intensity vector, the estimated baseline composition was (Dense= 0, Hazy= 0, Normal= 1). An increase of 1 Gy in the dose shifts the baseline composition by (0.35, 0.35, 0.30). Consequently, the dose is positively correlated with both dense and hazy. Furthermore, over time the changes in the compositions are characterized by an increment in the proportion of dense. The estimated value of the parameter capturing

Table 3 Parameter estimates and standard errors of the fixed effects for modeling RILD

Parameter	Estimate	S.E.	$exp(Estimate)$	95%CI
Occurrence model				
Intercept	-9.89	0.76		
Median dose	0.23	0.02	1.26	(1.21, 1.31)
Years	2.35	0.90	10.48	(1.78, 61.58)
Years2	-0.84	0.44	0.43	(0.18, 1.01)
Intensity model				
Dense/Normal				
Intercept	-13.02	1.8E-06		
Median dose	0.16	3.2E-08	1.18	(1.18, 1.18)
Years	2.72	0.39	15.20	(7.11, 32.52)
Years2	-0.99	0.19	0.37	(0.26, 0.54)
Hazy/Normal				
Intercept	-9.82	0.25		
Median dose	0.14	4.3E-03	1.15	(1.14, 1.16)
Years	0.50	0.28	1.65	(0.95, 2.85)
Years2	-0.16	0.14	0.85	(0.65, 1.11)

the correlation of the occurrence and intensity model, namely $\psi_{dc}=(0.44, 0.38)$. This indicates positive dependence between the two components of our model, thus enforcing a joint model for the two processes.

5.2 Assessing Model Adequacy

Assumptions and goodness of fit of the model were assessed and results are presented on the Appendix 4. Apparent deviations of the normality assumption of the random effects for the occurrence model (c_i) are observed (Fig. 4a), where the tails correspond to patches with RILD. The bivariate vector of random effects in the intensity model is distributed by a mixture of normal distributions. A goodness-of-fit test of a mixture of normal distributions using the Cramér-von Mises Garcia Portugues (2020) can be implemented for this objective. However, these results are not presented in the current work. However, as shown in Verbeke and Lesaffre (1997), and in Butler and Louis (1992), the inference of the fixed effects might be robust to the non-normality of random effects, but it can affect the predictions at the patch level.

Figure 4b shows the q-q plots for the errors of the intensity model. There is also evidence of a lack of normality in the residuals, and mainly patches that move further from the straight line correspond to patches without RILD, which is expected given that the intensity model is conditional on the presence of RILD. However, the number of patches is larger than 250. Therefore, Central Limit Theorem should assure some robustness, provided the model is asymptotically unbiased.

Figure 5 shows the scatter plot of the standardized residuals for dense and hazy versus dose. In these plots, patches with no observed RILD at any occasion cluster

along lines, and since the estimates of the random effects (d_i) are set at the population-level predictions, large standard errors for the random effects in these patches are expected. Residuals in doses less than 20 Gy have zero variability because those patches did not present RILD at any time point. After 20 Gy, the residuals seem to have larger variability at larger doses of RT.

For assessing the assumption that the two models, occurrence and intensity, are correlated, Fig. 6 shows the estimated log odds for each patch and the expected $alr(Dense)$ and $alr(Hazy)$. As mentioned, the points cluster along lines representing the patches that did not have RILD at any time point after RT. Patches with low propensities to have RILD also tend to have lower RILD intensity. This provides evidence of the correlation between the models and the importance of not ignoring this relationship in the model specification.

Figure 7 shows the observed number of times that each patch presented RILD, and the sum of the predicted probabilities at each time point ($\sum_{j=1}^{n_i} \hat{U}_{ij}$). As observed, the prediction in patches of normal tissue during the whole follow-up (Observed = 0) is very good. However, some discrepancies in the fit of the occurrence pattern is observed for patches with RILD at some point. The scaled residuals $(U_i - \hat{U}_i) / \sqrt{\hat{U}_i(n_i - \hat{U}_i) / n_i}$ were calculated, and outliers were identified as scaled residuals with magnitude larger than 2.5. In this data set, only 1.0 % of the patches were classified as outliers.

Figure 8 shows the scatter plots of predicted versus observed transformed dense and hazy proportions. The predicted values from the intensity model are labeled as “subpredicted” and are in the first column of this set of plots. The second column shows the predictions from the first model multiplied by the estimated probability of RILD for each patch. This shows the effect of the underestimated probabilities of RILD from the occurrence model on the predicted values from the intensity model adversely resulting in inferior predictions.

6 Discussion

Longitudinal compositional data, with a peak at zero, occur in many applications in health sciences. Different sources of correlation in the data are observed between subjects over time and between parts of the compositional vector. Additionally, the presence of singular zero represents a particular challenge in the analysis of such data. A new approach for the analysis of zero-inflated longitudinal compositional data, which additionally shows bimodal distributions of the intensity component, is proposed in this work. This model is fully parametric and uses asymptotic approximations to estimate the parameters. The likelihood was approximated using a sixth-order multivariate Laplace method, and even though a higher-order could have been used, empirical evidence shows that it would not improve the accuracy (Raudenbush et al., 2000). The EM algorithm was used to estimate the probability of the mixture model, and the approximate Fisher scoring procedure was implemented to estimate other parameters in the model. Finally, approximate empirical Bayes

estimates by Pareto smoothed importance sampling were implemented for the estimation of the random effects.

The simulation study showed mixed performance of the proposed method, with reasonable estimates of the model's fixed effects, but with highly biased estimates of other parameters. The average time that the procedure takes for the estimation of the fixed effects seems to be in the feasible range for statistical analysis.

Although the method seems to have good convergence, in future studies, it is important to assess whether the implemented step-halving procedure is causing false convergence. The implemented procedure is similar to the implemented "loop3" in the R function *glm*, (Marschner, 2011) which invokes the step-halving until convergence, until the difference in the relative deviance goes from one side of the convergence region to the other, or until this process has been repeated a fixed number of times. In this last situation, the new estimates can be very close to the previous ones and fall inside the convergence region, which could be causing a false convergence (Marschner, 2011). In this study, this step-halving procedure has repeated a maximum of 50 times. The effects in the model performance on those cases when convergence is attained after reaching the maximum number of repetitions need to be assessed.

The proposed method seems to be making an adequate estimate of the variability of the estimates through the SE. However, for $N = 300$ performance of the estimated SEs was inadequate. Since it is very common to find studies with sample sizes smaller than 300, it is advisable to use the method with caution in such cases. In our application, increasing sample size requires more imaging follow up. This is very difficult to achieve.

Estimation of nuisance parameters was somewhat unsatisfactory. As mentioned above, "loop 3" is the most likely cause of poor model performance in estimating variance components. By performing the step-halving procedure in each iteration, the change of each parameter is reduced more and more with each time this loop runs. Therefore, the changes in the estimates between iterations can be tiny for some parameters. Improvements to the step-halving procedure, perhaps limiting it only to fixed effects, could be implemented to mend this lack of performance.

The real data application of this model using the lung scans of a patient with NSCLC shows positive association of RILD with RT dose. Also the results suggest a positive correlation between the random effects of both models (occurrence and intensity). The methodology proposed in this study considers this correlation in the estimation process. Ignoring this relationship, by fitting the logit and the linear models separately, would introduce substantial bias in the estimates of coefficients, particularly in the coefficients of the intensity model.

It is important to consider that the estimated coefficients in the occurrence model represent the effect of changes in the variables on a particular patch, not a population-average effect. That is, the model responds to the probability of RILD for each patch, not to the percentage of patches with RILD. Population-average estimates tend to maintain the same direction as the mixed model estimates but with smaller magnitude, as well as smaller standard errors. Some approaches to estimate the population-averaged effects based on mixed model estimates have been discussed in

the literature (Neuhaus et al., 1991), but were not implemented here since the main interest was the effect of the variables in each patch and not the overall effect.

Lack of agreement with the normality assumption was observed, for both the random effects and the error term in the intensity model. Exclusion or misspecification of the variables in the model might be one of the explanations of this finding. In this particular example, the number of patches is larger than 250, Central Limit Theorem should allow asymptotic approximation to a reasonable level. Therefore, the inference of the fixed effects might be robust to the non-normality of random effects (Butler and Louis, 1992; Verbeke and Lesaffre, 1997). Nevertheless, violation of the homoscedasticity assumption of the residuals might have a negative impact on the variance estimates of the parameters in the model.

Notably, the patches with no RILD at all time points represent an issue for both parts of the model. First, in the occurrence model, these patches might have an actual logit of probabilities of RILD of $-\infty$ (Olsen and Schafer, 2001). Also, even though the intensity model can provide predictions on these patches, these predictions have large standard errors (Olsen and Schafer, 2001). Therefore, these patches have a substantial impact on the model assumptions and model performance. Furthermore, most of the procedures for diagnosing lack of adequacy in this model are based on the independent results or residuals of each of the parts (occurrence and intensity); however, future research should involve methods for model adequacy assessment using the residuals from the two-part model.

Additionally, the model presented in this section is a simplified version of the radiographic lung change following radiotherapy of lung cancer, being able to predict an average trajectory overtime for all patches, which moves higher as the received dose increases, and starts at different points for each patch due to the random intercept. However, this model does not allow us to identify different trajectories over time. Therefore, future applications in this data should include more complex models that better adjust the behavior of different patches over time, e.g. adding interactions between time and received dose.

In conclusion, the two-part mixed-effects mixture model for zero-inflated longitudinal compositional data has shown reasonable performance in estimating fixed effects. However, the model would benefit from improvements in the estimation of the nuisance parameters. Additionally, future research should study the impact of the inclusion of other random effects that represent clustering, or random slopes, as well as the effect of missing data.

Appendix 1: Isodose Patch Definition

See Fig. 3

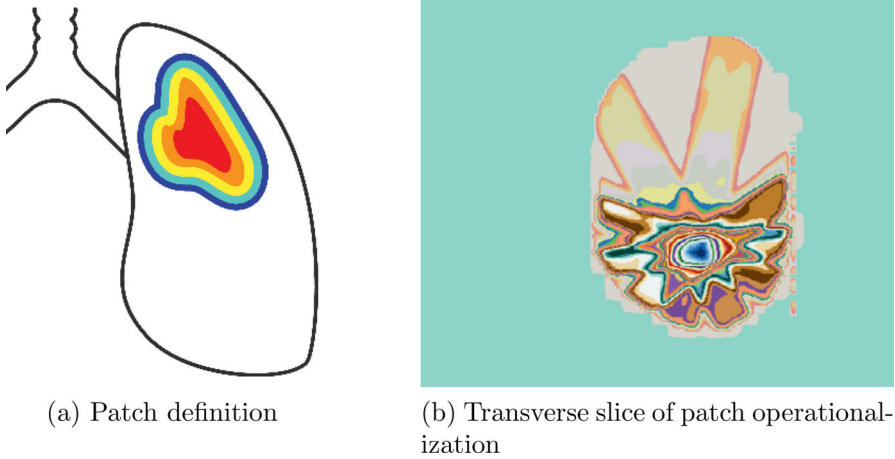


Fig. 3 Isodose patch, as defined using the isodose lines with dose interval 0.2 Gy

Appendix 2: The Likelihood

$$L \propto \prod_{i=1}^N \int_{c_i} \int_{\tilde{d}_i} f(y_i | \beta, \gamma, c_i, \tilde{d}_i, \Sigma) f(c_i, \tilde{d}_i | m_1, \mu_1, \Psi) dc_i \tilde{d}_i \tag{12}$$

assuming Y_i are *i.i.d.*:

$$= \prod_{i=1}^N \int_{c_i} \int_{\tilde{d}_i} \prod_{j=1}^{n_i} f(y_{ij} | \beta, \gamma, c_i, \tilde{d}_i, \Sigma) f(c_i, \tilde{d}_i | m_1, \mu_1, \Psi) dc_i \tilde{d}_i$$

with the two-parts model and using conditional distribution for random effects:

$$= \prod_{i=1}^N \int_{c_i} \int_{\tilde{d}_i} \prod_{j=1}^{n_i} (1 - \pi_{ij})^{(1-U_{ij})} \pi_{ij}^{U_{ij}} [f(V_{ij} | \gamma, \tilde{d}_i, \Sigma)]^{U_{ij}} \times f(c_i | \Psi_{cc}) f(\tilde{d}_i | m_1, \mu_1, c_i, \Psi_{dd}) dc_i \tilde{d}_i$$

re-expressing internal product:

$$= \prod_{i=1}^N \int_{c_i} \int_{\tilde{d}_i} \prod_{j=1}^{n_i} (1 - \pi_{ij})^{(1-U_{ij})} \pi_{ij}^{U_{ij}} \prod_{j=1}^{n_i} [f(V_{ij} | \gamma, \tilde{d}_i, \Sigma)]^{U_{ij}} f(c_i | \Psi_{cc}) f(\tilde{d}_i | m_1, \mu_1, c_i, \Psi_{dd}) dc_i \tilde{d}_i$$

using that $\exp(\log(x)) = x$:

$$\begin{aligned}
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \exp \left(\log \left(\prod_{j=1}^{n_i} (1 - \pi_{ij})^{(1-U_{ij})} \pi_{ij}^{U_{ij}} \right) \right) \times \exp \left(\log \left(\prod_{j=1}^{n_i} [f(\mathbf{V}_{ij} | \boldsymbol{\gamma}, \tilde{\mathbf{d}}_i, \boldsymbol{\Sigma})]^{U_{ij}} \right) \right) \\
 &\quad \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \\
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \exp \left(\sum_{j=1}^{n_i} \left((1 - U_{ij}) \log(1 - \pi_{ij}) + U_{ij} \log(\pi_{ij}) \right) \right) \\
 &\quad \times \exp \left(\sum_{j=1}^{n_i} U_{ij} \left[-\log(2\pi) - \frac{1}{2} \log(|\boldsymbol{\Sigma}|) \right. \right. \\
 &\quad \left. \left. - \frac{1}{2} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i)) \boldsymbol{\Sigma}^{-1} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i))^T \right] \right) \\
 &\quad \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i
 \end{aligned}$$

Let n_i^* the number of non-zero vectors in \mathbf{y}_i

$$\begin{aligned}
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \times \exp \left(\sum_{j=1}^{n_i} \left(U_{ij} \log \left(\frac{\pi_{ij}}{(1 - \pi_{ij})} \right) + \log(1 - \pi_{ij}) \right) \right) \\
 &\quad \times \exp \left(-n_i^* \log(2\pi) - \frac{n_i^*}{2} \log(|\boldsymbol{\Sigma}|) \right. \\
 &\quad \left. - \frac{1}{2} \sum_{j=1}^{n_i} U_{ij} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i)) \boldsymbol{\Sigma}^{-1} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i))^T \right) \\
 &\quad \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \\
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \exp \left(\sum_{j=1}^{n_i} \left(U_{ij} \eta_{ij} + \log(1 - \pi_{ij}) \right) \right) \\
 &\quad \times \exp \left(-n_i^* \log(2\pi) - \frac{n_i^*}{2} \log(|\boldsymbol{\Sigma}|) \right. \\
 &\quad \left. - \frac{1}{2} \sum_{j=1}^{n_i} U_{ij} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i)) \boldsymbol{\Sigma}^{-1} (\mathbf{V}_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i))^T \right) \\
 &\quad \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \\
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \int_{\tilde{\mathbf{d}}_i} \exp(l_{U_{ij}}) \times \exp(l_{V_{ij}}) \times f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\mathbf{c}_i d\tilde{\mathbf{d}}_i \\
 &= \prod_{i=1}^N \int_{\mathbf{c}_i} \exp(l_{U_{ij}}) f(\mathbf{c}_i | \boldsymbol{\Psi}_{cc}) \left(\int_{\tilde{\mathbf{d}}_i} \exp(l_{V_{ij}}) f(\tilde{\mathbf{d}}_i | m_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dd}) d\tilde{\mathbf{d}}_i \right) d\mathbf{c}_i
 \end{aligned} \tag{13}$$

Then, working with the inner integral in (13)

$$\begin{aligned}
 & \int_{d_i} \exp(l_{V_{ij}}) f(\tilde{d}_i | \boldsymbol{\mu}_1, \boldsymbol{\mu}_1, \mathbf{c}_i, \boldsymbol{\Psi}_{dc}) d\tilde{d}_i \\
 &= \int_{d_i} \exp(l_{V_{ij}}) [m_1 N(\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})]^{\Delta_i} [(1 - m_1) N(\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})]^{1 - \Delta_i} d\tilde{d}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} \int_{d_i} \exp(l_{V_{ij}}) N(\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})^{\Delta_i} N(\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i, \mathbf{H})^{1 - \Delta_i} d\tilde{d}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-q^*} |H|^{-\frac{1}{2}} \int_{d_i} \exp\left(-n_i^* \log(2\pi) - \frac{n_i^*}{2} \log(|\Sigma|)\right. \\
 &\quad \left. - \frac{1}{2} \sum_{j=1}^{n_i} U_{ij} (V_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i)) \Sigma^{-1} (V_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i))^T\right) \\
 &\quad \times \exp\left(-\Delta_i \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) \\
 &\quad \times \exp\left(-(1 - \Delta_i) \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) d\tilde{d}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-q^*} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |H|^{-\frac{1}{2}} \\
 &\quad \times \int_{d_i} \exp\left(\frac{1}{2} \sum_{j=1}^{n_i} U_{ij} (V_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i)) \Sigma^{-1} (V_{ij} - (\mathbf{X}_{ij}^* \boldsymbol{\gamma} + \mathbf{Z}_{ij}^* \mathbf{d}_i))^T\right) \\
 &\quad \times \exp\left(-\Delta_i \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) \\
 &\quad \times \exp\left(-(1 - \Delta_i) \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) d\tilde{d}_i
 \end{aligned}$$

Let i^* identify the group of observations $y_{ij} \neq \mathbf{0}$ in \mathbf{y}_i

$$\begin{aligned}
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-q^* - n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |H|^{-\frac{1}{2}} \\
 &\quad \times \int_{d_i} \exp\left(-\frac{1}{2} (\tilde{V}_{i^*} - (\tilde{\mathbf{X}}_{i^*}^* \tilde{\boldsymbol{\gamma}} + \tilde{\mathbf{Z}}_{i^*}^* \tilde{d}_i))^T \tilde{\Sigma}^{-1} (\tilde{V}_{i^*} - (\tilde{\mathbf{X}}_{i^*}^* \tilde{\boldsymbol{\gamma}} + \tilde{\mathbf{Z}}_{i^*}^* \tilde{d}_i))\right) \\
 &\quad \times \exp\left(-\Delta_i \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) \\
 &\quad \times \exp\left(-(1 - \Delta_i) \left(-\frac{1}{2} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))^T \mathbf{H}^{-1} (\tilde{d}_i - (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i))\right)\right) d\tilde{d}_i
 \end{aligned}$$

where $\tilde{V}_{i^*} = \text{vec}(V_{i^*})$, $\tilde{\boldsymbol{\gamma}} = \text{vec}(\boldsymbol{\gamma})$, $\tilde{\mathbf{X}}_{i^*}^* = I_{2 \times 2} \otimes \mathbf{X}_{i^*}^*$, $\tilde{\mathbf{Z}}_{i^*}^* = I_{2 \times 2} \otimes \mathbf{Z}_{i^*}^*$, and $\tilde{\Sigma}^{-1} = \Sigma^{-1} \otimes I_{n_{i^*} \times n_{i^*}}$

Let $\mathbf{A}_{i^*} = \tilde{V}_{i^*} - \tilde{\mathbf{X}}_{i^*}^* \tilde{\boldsymbol{\gamma}}$; $\mathbf{f}_{i1} = \boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i$, and $\mathbf{f}_{i2} = \boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i$. Then,

$$\begin{aligned}
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-q^* - n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \int_{\tilde{\mathbf{d}}_i} \exp \left(-\frac{1}{2} (\mathbf{A}_{i^*} - \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i)^T \tilde{\Sigma}^{-1} (\mathbf{A}_{i^*} - \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i) \right) \\
 &\quad \times \exp \left(-\Delta_i \left(-\frac{1}{2} (\tilde{\mathbf{d}}_i - \mathbf{f}_{i1})^T \mathbf{H}^{-1} (\tilde{\mathbf{d}}_i - \mathbf{f}_{i1}) \right) \right) \\
 &\quad \times \exp \left(-(1 - \Delta_i) \left(-\frac{1}{2} (\tilde{\mathbf{d}}_i - \mathbf{f}_{i2})^T \mathbf{H}^{-1} (\tilde{\mathbf{d}}_i - \mathbf{f}_{i2}) \right) \right) d\tilde{\mathbf{d}}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-q^* - n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \int_{\tilde{\mathbf{d}}_i} \exp \left(-\frac{1}{2} (\mathbf{A}_{i^*}^T - \tilde{\mathbf{d}}_i^T \tilde{\mathbf{Z}}_{i^*}^{*T}) (\tilde{\Sigma}^{-1} \mathbf{A}_{i^*} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i) \right) \\
 &\quad \times \exp \left(-\Delta_i \left(-\frac{1}{2} (\tilde{\mathbf{d}}_i^T - \mathbf{f}_{i1}^T) (\mathbf{H}^{-1} \tilde{\mathbf{d}}_i - \mathbf{H}^{-1} \mathbf{f}_{i1}) \right) \right) \\
 &\quad \times \exp \left(-(1 - \Delta_i) \left(-\frac{1}{2} (\tilde{\mathbf{d}}_i^T - \mathbf{f}_{i2}^T) (\mathbf{H}^{-1} \tilde{\mathbf{d}}_i - \mathbf{H}^{-1} \mathbf{f}_{i2}) \right) \right) d\tilde{\mathbf{d}}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \\
 &\quad \times \int_{\tilde{\mathbf{d}}_i} \exp \left(-\frac{1}{2} (\mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \mathbf{A}_{i^*} - \mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i - \tilde{\mathbf{d}}_i^T \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \mathbf{A}_{i^*} + \tilde{\mathbf{d}}_i^T \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i) \right. \\
 &\quad \left. -\frac{1}{2} \Delta_i (\tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i - \tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \mathbf{f}_{i1} - \mathbf{f}_{i1}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i + \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1}) \right. \\
 &\quad \left. -\frac{1}{2} (1 - \Delta_i) (\tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i - \tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \mathbf{f}_{i2} - \mathbf{f}_{i2}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i + \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2}) \right) d\tilde{\mathbf{d}}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \\
 &\quad \times \int_{\tilde{\mathbf{d}}_i} \exp \left(-\frac{1}{2} (\mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \mathbf{A}_{i^*} - 2\mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i + \tilde{\mathbf{d}}_i^T \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i) \right. \\
 &\quad \left. -\frac{1}{2} \Delta_i (\tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i - 2\mathbf{f}_{i1}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i + \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1}) \right. \\
 &\quad \left. -\frac{1}{2} (1 - \Delta_i) (\tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i - 2\mathbf{f}_{i2}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i + \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2}) \right) d\tilde{\mathbf{d}}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \\
 &\quad \times \exp \left(-\frac{1}{2} \mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \mathbf{A}_{i^*} - \frac{1}{2} \Delta \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1} - \frac{1}{2} (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2} \right) \\
 &\quad \times \int_{\tilde{\mathbf{d}}_i} \exp \left(\mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i + \Delta \mathbf{f}_{i1}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i \right. \\
 &\quad \left. -\frac{1}{2} (1 - \Delta_i) \tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i - \frac{1}{2} \tilde{\mathbf{d}}_i^T \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* \tilde{\mathbf{d}}_i - \frac{1}{2} \Delta_i \tilde{\mathbf{d}}_i^T \mathbf{H}^{-1} \tilde{\mathbf{d}}_i \right) d\tilde{\mathbf{d}}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \\
 &\quad \times \exp \left(-\frac{1}{2} \mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \mathbf{A}_{i^*} - \frac{1}{2} \Delta \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1} - \frac{1}{2} (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2} \right) \\
 &\quad \times \int_{\tilde{\mathbf{d}}_i} \exp \left((\mathbf{A}_{i^*}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* + \Delta \mathbf{f}_{i1}^T \mathbf{H}^{-1} + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1}) \tilde{\mathbf{d}}_i - \frac{1}{2} \tilde{\mathbf{d}}_i^T (\mathbf{H}^{-1} + \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^*) \tilde{\mathbf{d}}_i \right) d\tilde{\mathbf{d}}_i
 \end{aligned}$$

Let $\mathbf{B}_i = \tilde{\mathbf{Z}}_{i^*}^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_{i^*}^* + \mathbf{H}^{-1}$

$$\begin{aligned}
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} \\
 &\quad \times \exp\left(-\frac{1}{2} \mathbf{A}_i^T \tilde{\Sigma}^{-1} \mathbf{A}_i - \frac{1}{2} \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1} - \frac{1}{2} (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2}\right) \\
 &\quad \times \int_{\tilde{d}_i} \exp\left(\left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* + \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1}\right) \tilde{d}_i - \frac{1}{2} \tilde{d}_i^T \mathbf{B}_i \tilde{d}_i\right) d\tilde{d}_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + q^*)} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} (2\pi)^{q^*} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} \\
 &\quad \times \exp\left(-\frac{1}{2} \mathbf{A}_i^T \tilde{\Sigma}^{-1} \mathbf{A}_i - \frac{1}{2} \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{f}_{i1} - \frac{1}{2} (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{f}_{i2}\right) \\
 &\quad \times \exp\left(\frac{1}{2} \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* + \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1}\right) \mathbf{B}_i^{-1}\right. \\
 &\quad \left. \times \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* + \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1}\right)^T\right) \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} \\
 &\quad \times \exp\left(-\frac{1}{2} \mathbf{A}_i^T \tilde{\Sigma}^{-1} \mathbf{A}_i - \frac{1}{2} \Delta_i (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)^T \mathbf{H}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)\right. \\
 &\quad \left. - \frac{1}{2} (1 - \Delta_i) (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)^T \mathbf{H}^{-1} (\boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)\right) \\
 &\quad \times \exp\left(\frac{1}{2} \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* + \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1}\right)\right. \\
 &\quad \left. \times \left(\mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \mathbf{A}_i + \Delta_i \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i1} + (1 - \Delta_i) \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i2}\right)\right) \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} \\
 &\quad \times \exp\left(-\frac{1}{2} \mathbf{A}_i^T \tilde{\Sigma}^{-1} \mathbf{A}_i - \frac{1}{2} \Delta_i (\boldsymbol{\mu}_1^T + \mathbf{c}_i^T \boldsymbol{\Psi}_{cc}^{-1} \boldsymbol{\Psi}_{dc}^T) (\mathbf{H}^{-1} \boldsymbol{\mu}_1 + \mathbf{H}^{-1} \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)\right. \\
 &\quad \left. - \frac{1}{2} (1 - \Delta_i) (\boldsymbol{\mu}_2^T + \mathbf{c}_i^T \boldsymbol{\Psi}_{cc}^{-1} \boldsymbol{\Psi}_{dc}^T) (\mathbf{H}^{-1} \boldsymbol{\mu}_2 + \mathbf{H}^{-1} \boldsymbol{\Psi}_{dc} \boldsymbol{\Psi}_{cc}^{-1} \mathbf{c}_i)\right) \\
 &\quad \times \exp\left(\frac{1}{2} \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \mathbf{A}_i + \Delta_i \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i1}\right.\right. \\
 &\quad \left. + (1 - \Delta_i) \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i2} + \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \mathbf{A}_i\right. \\
 &\quad \left. + \Delta_i^2 \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i1} + (1 - \Delta_i) \Delta_i \mathbf{f}_{i1}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i2}\right. \\
 &\quad \left. + (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \mathbf{A}_i + \Delta_i (1 - \Delta_i) \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i1}\right. \\
 &\quad \left. + (1 - \Delta_i)^2 \mathbf{f}_{i2}^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{f}_{i2}\right)
 \end{aligned}$$

Let $E_i^T = A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc}^{-1}$, $M_i^T = \Delta_i \mu_1^T + (1 - \Delta_i) \mu_2^T$,
 and $D_i = \Psi_{cc}^{-1} \Psi_{dc}^T (H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^*) \Psi_{dc} \Psi_{cc}^{-1}$. Then,

$$\begin{aligned}
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |H|^{-\frac{1}{2}} |B_i^{-1}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} A_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1}) A_i \right) \\
 &\quad \times \exp \left(A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1} M_i - \frac{1}{2} \Delta_i \mu_1^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_1 \right. \\
 &\quad \left. - \frac{1}{2} (1 - \Delta_i) \mu_2^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_2 + E_i^T c_i \right. \\
 &\quad \left. - M_i^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \Psi_{dc} \Psi_{cc}^{-1} c_i - \frac{1}{2} c^T_i D_i c_i \right)
 \end{aligned}$$

Let $G_i^T = E_i^T - M_i^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \Psi_{dc} \Psi_{cc}^{-1}$

$$\begin{aligned}
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |H|^{-\frac{1}{2}} |B_i^{-1}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} A_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1}) A_i \right) \\
 &\quad \times \exp \left(A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1} M_i - \frac{1}{2} \Delta_i \mu_1^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_1 \right. \\
 &\quad \left. - \frac{1}{2} (1 - \Delta_i) \mu_2^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_2 + G_i^T c_i - \frac{1}{2} c^T_i D_i c_i \right)
 \end{aligned}$$

(14)

Now, going back to the integral in (13):

$$\begin{aligned}
 & \int_{c_i} \exp(l_{U_{ij}}) f(c_i | \Psi_{cc}) \left(\int_{\tilde{d}_i} \exp(l_{V_{ij}}) f(\tilde{d}_i | m_1, \mu_1, c_i, \Psi_{dd}) d\tilde{d}_i \right) dc_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-n_i^*} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i \right) \\
 & \quad \times \exp \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 \right. \\
 & \quad \left. - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \right) \\
 & \quad \times \int_{c_i} \exp(l_{U_{ij}}) f(c_i | \Psi_{cc}) \exp \left(\mathcal{G}_i^T c_i - \frac{1}{2} c_i^T \mathbf{D}_i c_i \right) dc_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \quad \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i \right) \\
 & \quad \times \exp \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 \right. \\
 & \quad \left. - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \right) \\
 & \quad \times \int_{c_i} \exp \left(l_{U_{ij}} - \frac{1}{2} c_i^T \Psi_{cc}^{-1} c_i + \mathcal{G}_i^T c_i - \frac{1}{2} c_i^T \mathbf{D}_i c_i \right) dc_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \quad \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i \right) \\
 & \quad \times \exp \left(\mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 \right. \\
 & \quad \left. - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \right) \\
 & \quad \times \int_{c_i} \exp \left(l_{U_{ij}} - \frac{1}{2} c_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i) c_i + \mathcal{G}_i^T c_i \right) dc_i \\
 &= m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \quad \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i \right. \\
 & \quad \left. - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \right. \\
 & \quad \left. + \frac{1}{2} \mathcal{G}_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i \right) \\
 & \quad \times \int_{c_i} \exp \left(l_{U_{ij}} - \frac{1}{2} (c_i - (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i)^T (\Psi_{cc}^{-1} + \mathbf{D}_i) (c_i - (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i) \right) dc_i
 \end{aligned}$$

(15)

The integral in (15) is intractable, but it is similar to the likelihood in a mixed effects

logistic regression where the random effects follow a normal distribution with different location and scale parameters. As discussed for different authors, several strategies have been implemented to evaluate this integral. Using Olsen and Schafer (2001) approach with a sixth-order Laplace approximation, we have that:

$$\begin{aligned}
 & \int_{c_i} \exp(f(c_i)) dc_i \\
 &= \int_{c_i} \exp\left(l_{U_{ij}} - \frac{1}{2} (c_i - (\Psi_{cc}^{-1} + D_i)^{-1} G_i)^T (\Psi_{cc}^{-1} + D_i) (c_i - (\Psi_{cc}^{-1} + D_i)^{-1} G_i)\right) dc_i \\
 &\approx \exp(f(\tilde{c}_i)) | -f^{(2)}(\tilde{c}_i) |^{-1/2} \left(1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} (Z_{ij} G_i^{-1} Z_{ij}^T)^2 - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} (Z_{ij} G_i^{-1} Z_{ij}^T)^3 \right. \\
 &\quad \left. + \frac{15}{72} \left(\sum_{j=1}^{n_i} Z_{ij}^T \tilde{m}_{ij}^{(3)} Z_{ij} G_i^{-1} Z_{ij}^T \right)^T G_i^{-1} \left(\sum_{j=1}^{n_i} Z_{ij}^T \tilde{m}_{ij}^{(3)} Z_{ij} G_i^{-1} Z_{ij}^T \right) \right)
 \end{aligned} \tag{16}$$

where:

- $\tilde{c}_i = (\Psi_{cc}^{-1} + D_i + Z_i^T \tilde{W}_i Z_i)^{-1} (Z_i^T \tilde{W}_i (U_i^* - X_i \beta) + G_i)$
- $U_i^* = \tilde{W}_i^{-1} (U_i - \tilde{\pi}_i) + \tilde{\eta}_i$
- W_i is a diagonal matrix with elements $\pi_{ij}(1 - \pi_{ij})$
- $\tilde{W}_i, \tilde{W}_i^{-1}, \tilde{\pi}_i,$ and $\tilde{\eta}_i$ are evaluated at $c_i = \tilde{c}_i$
- $-f^{(2)}(\tilde{c}_i) = Z_i^T \tilde{W}_i Z_i + \Psi_{cc}^{-1} + D_i = G_i$
- $\tilde{m}_{ij}^{(k)}$ is the $(k - 1)$ th derivative of π_{ij} with respect to η_{ij} evaluated at \tilde{c}_i

Now, going back to (15),

$$\begin{aligned}
 & m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i \right. \\
 & - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \\
 & \left. + \frac{1}{2} \mathcal{G}_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i \right) \\
 & \times \int_{c_i} \exp \left(l_{U_{ij}} - \frac{1}{2} (c_i - (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i)^T (\Psi_{cc}^{-1} + \mathbf{D}_i) (c_i - (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i) \right) dc_i \\
 & \approx m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i \right. \\
 & - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \\
 & \left. + \frac{1}{2} \mathcal{G}_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i \right) \\
 & \times \exp(f(\tilde{c}_i)) - [f^{(2)}(\tilde{c}_i)]^{-1} \left| 1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^2 - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^3 \right. \\
 & \left. + \frac{15}{72} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right)^T \mathbf{G}_i^{-1} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right) \right)
 \end{aligned} \tag{17}$$

Then,

$$\begin{aligned}
 L & \approx \prod_{i=1}^N m_1^{\Delta_i} (1 - m_1)^{1 - \Delta_i} (2\pi)^{-(n_i^* + \frac{q}{2})} |\Sigma|^{-\frac{n_i^*}{2}} |\mathbf{H}|^{-\frac{1}{2}} |\mathbf{B}_i^{-1}|^{\frac{1}{2}} |\Psi_{cc}|^{-\frac{1}{2}} \\
 & \times \exp \left(-\frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i \right. \\
 & - \frac{1}{2} \Delta_i \mu_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_1 - \frac{1}{2} (1 - \Delta_i) \mu_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mu_2 \\
 & \left. + \frac{1}{2} \mathcal{G}_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i \right) \\
 & \times \exp(f(\tilde{c}_i)) - [f^{(2)}(\tilde{c}_i)]^{-1} \left| 1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^2 - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^3 \right. \\
 & \left. + \frac{15}{72} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right)^T \mathbf{G}_i^{-1} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right) \right)
 \end{aligned} \tag{18}$$

Finally, the approximate log-likelihood is:

$$\begin{aligned}
 l \approx \sum_{i=1}^N & \left[\Delta_i \log(m_1) + (1 - \Delta_i) \log(1 - m_1) - \left(n_i^* + \frac{q}{2}\right) \log(2\pi) - \frac{n_i^*}{2} \log|\Sigma| - \frac{1}{2} \log|\mathbf{H}| + \frac{1}{2} \log|\mathbf{B}_i^{-1}| \right. \\
 & - \frac{1}{2} \log|\Psi_{cc}| - \frac{1}{2} \mathbf{A}_i^T (\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1}) \mathbf{A}_i + \mathbf{A}_i^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \mathbf{M}_i \\
 & - \frac{1}{2} \Delta_i \boldsymbol{\mu}_1^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \boldsymbol{\mu}_1 - \frac{1}{2} (1 - \Delta_i) \boldsymbol{\mu}_2^T \mathbf{H}^{-1} \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \boldsymbol{\mu}_2 \\
 & + \frac{1}{2} \mathcal{G}_i^T (\Psi_{cc}^{-1} + \mathbf{D}_i)^{-1} \mathcal{G}_i + f(\tilde{c}_i) + \frac{1}{2} \log| - [f^{(2)}(\tilde{c}_i)] + \log \left(1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^2 \right. \\
 & \left. - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^3 + \frac{15}{72} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right)^T \mathbf{G}_i^{-1} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right) \right] \quad (19)
 \end{aligned}$$

where

$$\begin{aligned}
 \bullet \quad P_i &= 1 - \frac{1}{8} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^2 - \frac{1}{48} \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T)^3 \\
 & + \frac{15}{72} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right)^T \mathbf{G}_i^{-1} \left(\sum_{j=1}^{n_i} \mathbf{Z}_{ij}^T \tilde{m}_{ij}^{(3)} \mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T \right)
 \end{aligned}$$

Appendix 3: The Score Vectors

Score vectors were derived using the proposed score vectors by Olsen and Schafer (2001) and extended to accommodate multivariate data.

$$\begin{aligned}
 S_{\beta_i} &= X_i^T \tilde{W}_i (U_i^* - X_i \beta - Z_i \tilde{c}_i) + \sum_{j=1}^{n_i} \left(-\frac{1}{2} \tilde{m}_{ij}^{(3)} (\mathbf{Z}_{ij} \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T) \right) (X_{ij}^T - X_i^T \tilde{W}_i Z_i \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T) \\
 & + \frac{1}{P_i} \sum_{j=1}^{n_i} R_{ij}^* (X_{ij}^T - X_i^T \tilde{W}_i Z_i \mathbf{G}_i^{-1} \mathbf{Z}_{ij}^T) \quad (20)
 \end{aligned}$$

$$\begin{aligned}
 S_{\gamma_i} &= \tilde{X}_i^{*T} \tilde{\Sigma}^{-1} \tilde{V}_i - \tilde{X}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{V}_i - \tilde{X}_i^{*T} \tilde{\Sigma}^{-1} \tilde{X}_i^* \tilde{\gamma} \\
 & + \tilde{X}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \tilde{\mathbf{Z}}_i^{*T} \tilde{\Sigma}^{-1} \tilde{X}_i^* \tilde{\gamma} \\
 & - \tilde{X}_i^{*T} \tilde{\Sigma}^{-1} \tilde{\mathbf{Z}}_i^* \mathbf{B}_i^{-1} \mathbf{H}^{-1} \Psi_{dc} \Psi_{cc}^{-1} \tilde{c}_i + t_i^T E_{ij} \quad (21)
 \end{aligned}$$

$$S_{\mu_{1i}} = -\frac{1}{2} \left(m_1 + \frac{m_1^2}{(1 - m_1)} \right) (H^{-1} B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* + \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1}) \mu_{1i} \quad (22)$$

$$\begin{aligned} S_{\phi_{\Sigma}} = & \left(\frac{\partial \text{vec}(\Sigma^{-1})}{\partial \phi_{\Sigma}^T} \right)^T \left(\frac{n_i^*}{2} \text{vec}(\Sigma) \right) \\ & + \left(\frac{\partial \text{vec}(\tilde{\Sigma}^{-1})}{\partial \phi_{\Sigma}^T} \right)^T \left(\frac{1}{2} \text{vec}(-\tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} - A_i A_i^T + 2A_i A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \right. \\ & - \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} A_i A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T}) - \frac{m_1}{2} \text{vec}(\tilde{Z}_i^* B_i^{-1} H^{-1} \mu_1 \mu_1^T \tilde{Z}_i^{*T} \\ & - \tilde{Z}_i^* B_i^{-1} H^{-1} \mu_1 \mu_1^T \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T}) - \frac{1 - m_1}{2} \text{vec}(\tilde{Z}_i^* B_i^{-1} H^{-1} \mu_2 \mu_2^T \tilde{Z}_i^{*T} \\ & - \tilde{Z}_i^* B_i^{-1} H^{-1} \mu_2 \mu_2^T \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T}) + P_i^{*T} E_{ij} - \frac{1}{2} [(\tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \otimes \tilde{Z}_i^* \Psi_{dc} \Psi_{cc_i}^{-1}) \\ & - (\tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \otimes \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* \Psi_{dc} \Psi_{cc_i}^{-1})] \text{vec}(G_i^{-1}) \\ & + \frac{1}{2} \text{vec}[\tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \tilde{c}_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc_i}^{-1})^T \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \\ & - \tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \tilde{c}_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc_i}^{-1})^T \tilde{Z}_i^{*T} + 2A_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc_i}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T} \\ & - 2\tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} A_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc_i}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T}] \\ & + \frac{1}{P_i} \left[\sum_{j=1}^{n_i} \text{vec}(\tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} G_i^{-1} Z_{ij}^T Z_{ij} G_i^{-1} (\Psi_{dc} \Psi_{cc_i}^{-1})^T (\tilde{Z}_i^{*T} - \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T})) F_{ij}^* \right. \\ & \left. - \frac{15}{72} \text{vec}(\tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} G_i^{-1} K_i K_i^T G_i^{-1} (\Psi_{dc} \Psi_{cc_i}^{-1})^T (\tilde{Z}_i^{*T} - \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T})) \right] \Big) \end{aligned} \quad (23)$$

$$\begin{aligned} S_{\phi_{cc_i}} = & \left(\frac{\partial \text{vec}(\Psi_{cc}^{-1})}{\partial \psi_{cc}^T} \right)^T \left(\frac{1}{2} \text{vec}(\Psi_{cc} - G_i^{-1} - \tilde{c}_i \tilde{c}_i^T) + D_i^{*T} E_{ij} \right. \\ & + \frac{1}{P_i} \left(\sum_{j=1}^{n_i} \text{vec}(G_i^{-1} Z_{ij}^T Z_{ij} G_i^{-1}) F_{ij}^* \right. \\ & \left. \left. - \frac{15}{72} \text{vec}(G_i^{-1} K_i K_i^T G_i^{-1}) \right) \right) \end{aligned} \quad (24)$$

$$\begin{aligned} S_{\Psi_{dc} \Psi_{cc_i}^{-1}} = & -(I_p \otimes H_i^*) (\text{vec}(G_i^{-1}) + \text{vec}(\tilde{c}_i \tilde{c}_i^T)) + \text{vec}(H^{-1} B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} A_i \tilde{c}_i^T) + S_i^{*T} E_{ij} \\ & + \frac{2}{P_i} \left(\sum_{j=1}^{n_i} \text{vec}(H_i^* G_i^{-1} Z_{ij}^T Z_{ij} G_i^{-1}) F_{ij}^* - \frac{15}{36} \text{vec}(H_i^* G_i^{-1} K_i K_i^T G_i^{-1}) \right) \end{aligned} \quad (25)$$

$$\begin{aligned}
 S_{\phi_{H_i}} = & \left(\frac{\partial \text{vec}(H^{-1})}{\partial \phi_H^T} \right)^T \left(-\frac{1}{2} ((B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \Psi_{dc} \Psi_{cc_i}^{-1} \otimes \Psi_{dc} \Psi_{cc_i}^{-1}) \right. \\
 & - (B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \otimes B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \Psi_{dc} \Psi_{cc_i}^{-1})) \text{vec}(G_i^{-1}) \\
 & + \frac{1}{2} \text{vec}[H - B_i^{-1} - B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} A_i A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} + m_1 B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_1 \mu_1^T H^{-1} B_i^{-1} \\
 & - m_1 B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_1 \mu_1^T + (1 - m_1) B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_2 \mu_2^T H^{-1} B_i^{-1} \\
 & - (1 - m_1) B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* \mu_2 \mu_2^T - \Psi_{dc} \Psi_{cc_i}^{-1} \tilde{c}_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \\
 & + 2 B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} A_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc}^{-1})^T + B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc_i}^{-1} \tilde{c}_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \\
 & - 2 B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} A_i \tilde{c}_i^T (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1}] + M_i^{*T} E_{ij} \\
 & + \frac{1}{P_i} \left[\sum_{j=1}^{n_i} \text{vec}((I_{2q^*} - B_i^{-1} H^{-1}) \Psi_{dc} \Psi_{cc}^{-1} G_i^{-1} Z_{ij}^T Z_{ij} G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1}) F_{ij}^* \right. \\
 & \left. - \frac{15}{72} \text{vec}((I_{2q^*} - B_i^{-1} H^{-1}) \Psi_{dc} \Psi_{cc}^{-1} G_i^{-1} K_i K_i^T G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1}) \right] \Big)
 \end{aligned} \tag{26}$$

where

$$\begin{aligned}
 P_i^* = & (\partial \tilde{c}_i) / (\partial \text{vec}^T(\tilde{\Sigma}^{-1})) = (Y_i^{*T} G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \\
 & \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T}) \\
 & - (Y_i^{*T} G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^{*T} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T}) \\
 & - (A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \tilde{Z}_i^{*T} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T}) \\
 & + (A_i^T \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^{*T})
 \end{aligned} \tag{27}$$

$$\begin{aligned}
 M_i^* = & -(Y_i^{*T} G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T) \\
 & + (A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T) \\
 & - (A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1}) \\
 & + (Y_i^{*T} G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} \otimes G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1})
 \end{aligned} \tag{28}$$

$$D_i^* = -(Y_i^{*T} G_i^{-1} \otimes G_i^{-1}) \tag{29}$$

$$t_i = -G_i^{-1} (\Psi_{dc} \Psi_{cc}^{-1})^T H^{-1} B_i^{-1} \tilde{Z}_i^* \tilde{\Sigma}^{-1} \tilde{X}_i^* \tag{30}$$

$$S_i^* = ((A_i^T \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1} \otimes G_i^{-1}) - (G_i^{-1} \otimes Y_i^{*T} G_i^{-1} H_i^{*T}) - (Y_i^{*T} G_i^{-1} \otimes G_i^{-1} H_i^{*T})) \tag{31}$$

$$E_{ij} = E_i - (\Psi_{cc}^{-1} + D_i) \tilde{c}_i + \sum_{j=1}^{n_i} Z_{ij}^T \left(-\frac{1}{2} \tilde{m}_{ij}^{(3)} (Z_{ij} G_i^{-1} Z_{ij}^T) + (U_{ij} - \tilde{\pi}_{ij}) + \frac{1}{P_i} R_{ij}^* \right) \tag{32}$$

$$H_i^* = \tilde{Z}_i^{*T} \tilde{\Sigma}^{-1} \tilde{Z}_i^* B_i^{-1} H^{-1} \Psi_{dc} \Psi_{cc}^{-1} \tag{33}$$

$$Y_i^* = Z_i^T \tilde{W}_i (U_i^* - X_i \beta) + E_i \tag{34}$$

$$F_{ij}^* = \frac{1}{4} \tilde{m}_{ij}^{(4)} (Z_{ij} G_i^{-1} Z_{ij}^T) + \frac{1}{16} \tilde{m}_{ij}^{(6)} (Z_{ij} G_i^{-1} Z_{ij}^T)^2 - \frac{15}{36} \tilde{m}_{ij}^{(3)} (Z_{ij} G_i^{-1} K_i) \tag{35}$$

$$R_{ij}^* = -\frac{1}{8} \tilde{m}_{ij}^{(5)} (Z_{ij} G_i^{-1} Z_{ij}^T)^2 + \frac{1}{4} \tilde{m}_{ij}^{(3)} Z_{ij} F_i Z_{ij}^T - \frac{1}{48} \tilde{m}_{ij}^{(7)} (Z_{ij} G_i^{-1} Z_{ij}^T)^3 + \frac{1}{16} \tilde{m}_{ij}^{(3)} Z_{ij} f_i Z_{ij}^T - \frac{15}{72} \tilde{m}_{ij}^{(3)} (Z_{ij} G_i^{-1} K_i)^2 + \frac{15}{36} \tilde{m}_{ij}^{(4)} (K_i^T G_i^{-1} Z_{ij}^T) (Z_{ij} G_i^{-1} Z_{ij}^T) - \frac{15}{36} \tilde{m}_{ij}^{(3)} Z_{ij} h_i Z_{ij}^T \tag{36}$$

$$F_i = \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(4)} G_i^{-1} Z_{ij}^T (Z_{ij} G_i^{-1} Z_{ij}^T) Z_{ij} G_i^{-1} \tag{37}$$

$$f_i = \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(6)} G_i^{-1} Z_{ij}^T (Z_{ij} G_i^{-1} Z_{ij}^T)^2 Z_{ij} G_i^{-1} \tag{38}$$

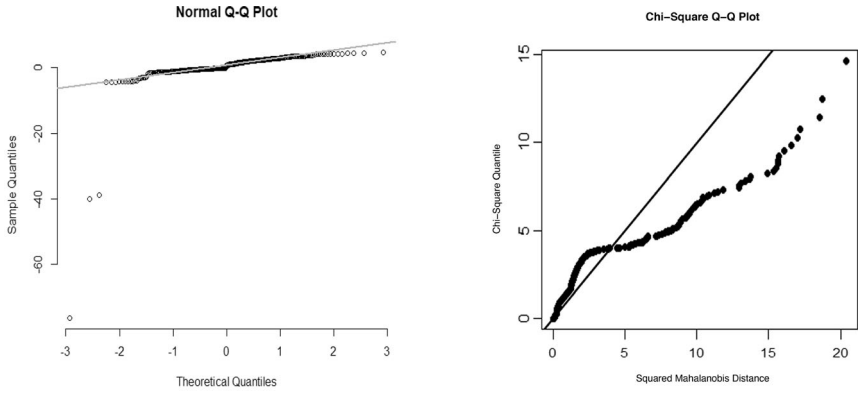
$$h_i = \sum_{j=1}^{n_i} \tilde{m}_{ij}^{(3)} G_i^{-1} Z_{ij}^T Z_{ij} G_i^{-1} (Z_{ij} G_i^{-1} K_i) \tag{39}$$

and

$$K_i = \sum_{j=1}^{n_i} Z_{ij}^T \tilde{m}_{ij}^{(3)} Z_{ij} G_i^{-1} Z_{ij}^T \tag{40}$$

Appendix 4: Assumptions and Goodness of Fit of the Model

See Figs. 4, 5, 6, 7 and 8.



(a) Normal Q-Q plots for random effects in the occurrence model (c_i)

(b) Chi-Square Q-Q plots for bivariate residuals from the intensity model

Fig. 4 Q-Q plots for random effects in the occurrence and bivariate residuals from the intensity model

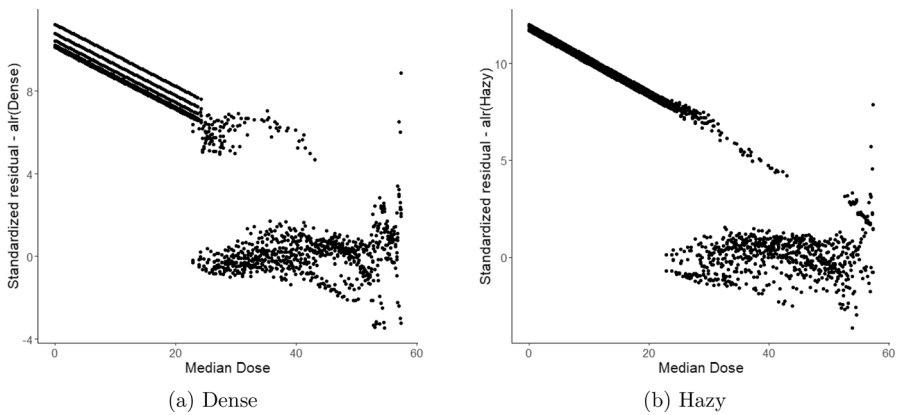


Fig. 5 Standardized residuals of the intensity model versus median dose per patch

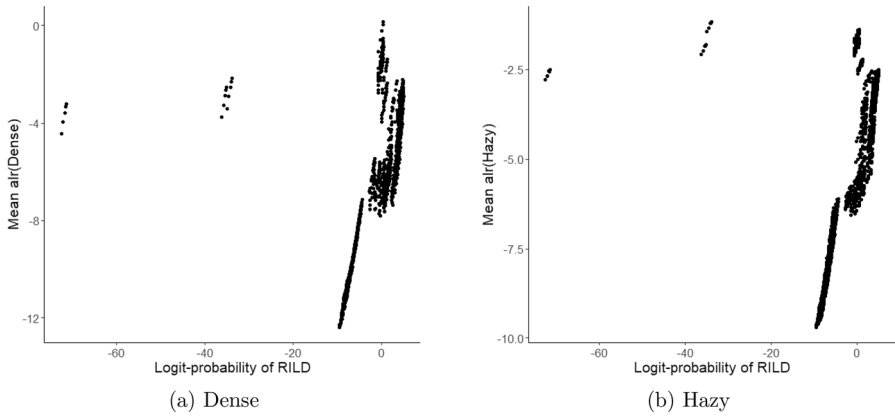


Fig. 6 Estimated mean $alr(Dense)$ and $alr(Hazy)$ versus logit-probability of RILD

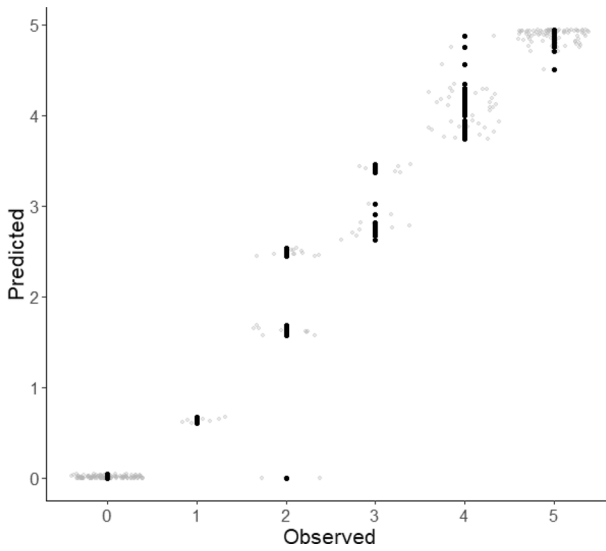


Fig. 7 Observed number of times with RILD versus expected probability of RILD

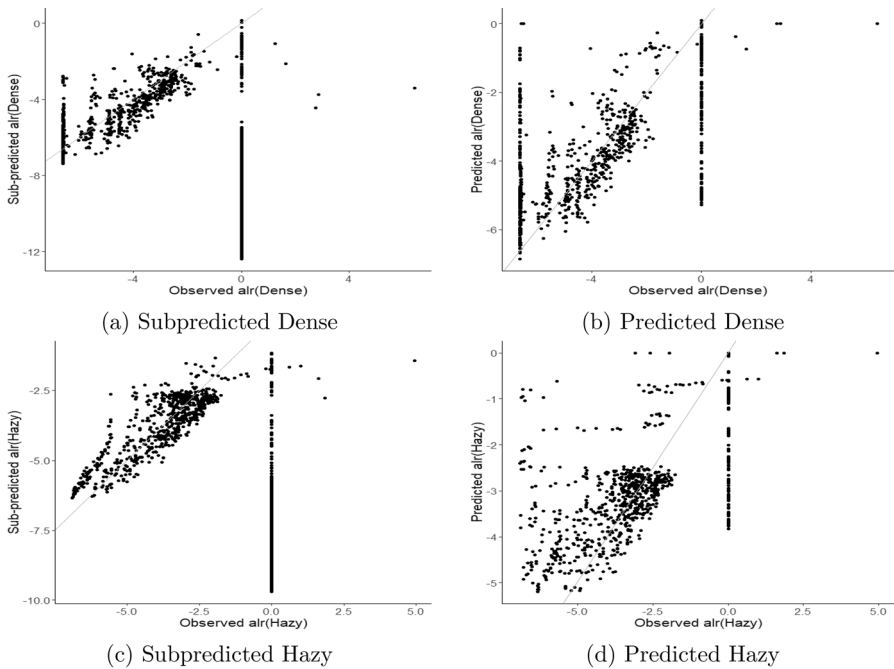


Figure D.5: Observed vs. predicted $alr(Dense)$ and $alr(Hazy)$

Fig. 8 Observed versus predicted $alr(Dense)$ and $alr(Hazy)$. Subpredicted values are the predictions of the intensity model, the predicted values are the prediction of the two-part model (integration between occurrence and intensity models)

Acknowledgements Data collection was partially supported by Award No. P30 CA016059NIH from the NCI Cancer Center Support Grant. VAR's effort was provided by the Colciencias-Fulbright Scholarship No. 481 of 2014.

Author Contributions All authors had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conceptualization, N.M., E.W., V.A.R., and R.N.M.; Methodology, N.M., V.A.R.; Software, V.A.R.; Formal analysis, N.M., V.A.R.; Resources, E.W. R.N.M.; Writing—original draft, N.M., V.A.R.; Writing—review and editing, N.M., E.W., V.A.R., and R.N.M.; Visualization, V.A.R.; Supervision, N.M., E.W.; Funding Acquisition, N.M., E.W.

Funding Services in support of this research were provided by the VCU Massey Cancer Center Biostatistics Shared Resource, supported in part with funding from NIH-NCI Cancer Center Support Grant P30 CA016059 and fullbright U.S. Scholar Program (Grant No 481).

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request. However, all analysis code is available at <https://github.com/vivifj03/TumorModel>.

Declarations

Conflict of interest The authors declare no Conflict of interest

Ethical Approval This study was approved by the institutional review board at VCU (HM20013924). Anonymized data stored and analyzed on a password-protected computer, which were only accessible to the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall Ltd, London
- Aitchison J, Egozcue JJ (2005) Compositional data analysis: Where are we and where should we be heading? *Math Geol* 37(7):829–850
- Aitchison J, Kay JW (2003) Possible solution of some essential zero problems in compositional data analysis. In: Thió-Henestrosa S, Martín-Fernández JA (eds) *CoDaWork'03*. Universitat de Girona, Girona, Spain, pp 1–6
- Bacon-Shone J (2003) Modelling structural zeros in compositional data. In: Thió-Henestrosa S, Martín-Fernández JA (eds) *CoDaWork'03*. Universitat de Girona, Girona, Spain, pp 1–4
- Bear J, Billheimer D (2016) A logistic normal mixture model for compositional data allowing essential zeros. *Austrian J Stat* 45:3–23
- Butler SM, Louis TA (1992) Random effects models with non-parametric priors. *Stat Med* 11(14–15):1981–2000 (<https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780111416>)
- Chen EZ, Li H (2016) A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics* 32(17):2611–2617
- Dunne EM, Fraser IM, Liu M (2018) Stereotactic body radiation therapy for lung, spine and oligometastatic disease: current evidence and future directions. *Ann Transl Med* 6(14):283
- Fry JM, Fry TRL, McLaren KR (2000) Compositional data analysis and zeros in micro data. *Appl Econ* 32(8):953–959
- Garcia Portugues E (2020) Goodness-of-fit tests for distribution models. Available from: <https://bookdown.org/egarpor/NP-UC3M/nptests.html>
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall/CRC
- Jain V, Berman AT (2018) Radiation pneumonitis: old problem, new tricks. *Cancers* 10(7):222
- Linda A, Trovo M, Bradleyc JD (2011) Radiation injury of the lung after stereotactic body radiation therapy (SBRT) for lung cancer: a timeline and pattern of CT changes. *Eur J Radiol* 79(1):147–154
- Marschner IC (2011) glm2: fitting generalized linear models with convergence problems. *R J* 3(2):12–15
- Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003) Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math Geol* 35(3):253–278
- Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J (2012) Model-based replacement of rounded zeros in compositional data: classical and robust approaches. *Comput Stat Data Anal* 56(9):2688–2704
- Matsuo Y, Shibuya K, Nakamura M, Narabayashi M, Sakanaka K, Ueki N et al (2012) Dose-volume metrics associated with radiation pneumonitis after stereotactic body radiation therapy for lung cancer. *Int J Radiat Oncol Biol Phys* 83(4):e545–e549
- Neuhaus JM, Kalbfleisch JD, Hauck WW (1991) A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev/Rev Int Stat* 59(1):25–35 (<http://www.jstor.org/stable/1403572>)
- Olsen MK, Schafer JL (2001) A two-part random-effects model for semicontinuous longitudinal data. *J Am Stat Assoc* 96(454):730–745. <https://doi.org/10.1198/016214501753168389>
- Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and analysis of compositional data, 1st edn. John Wiley & Sons Inc, Chichester, United Kingdom

- R Core Team. R (2019) A language and environment for statistical computing. Vienna, Austria; Available from: <https://www.R-project.org/>
- Raudenbush SW, Yang ML, Yosef M (2000) Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *J Comput Graph Stat* 9 (1):141
- Silverman JD, Durand HK, Bloom RJ, Mukherjee S, David LA (2018) Dynamic linear models guide design and analysis of microbiota studies within artificial human guts. *Microbiome* 6:202
- Templ M, Hron K, Filzmoser P (2017) Exploratory tools for outlier detection in compositional data with structural zeros. *J Appl Stat* 44(4):77–85
- Trovo M, Linda A, El Naqa I, Javidan-Nejad C, Bradley J (2010) Early and late lung radiographic injury following stereotactic body radiation therapy (SBRT). *Lung Cancer* 69(1):77–85
- Vehtari A, Simpson D, Gelman A, Yao Y, Gabry J (2015) Pareto smoothed importance sampling
- Verbeke G, Lesaffre E (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Comput Stat Data Anal* 23(4):541–556 (<http://www.sciencedirect.com/science/article/pii/S0167947396000473>)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.