

Modelos multiestado com fragilidade aninhada para análise de sobrevivência: aplicação a pacientes de recuperação de transplante de medula óssea

Katy Rocio Cruz Molina ¹
Jonathan Kevin Jordan Vasquez ²
Vera L. D. Tomazella ³

Resumo

Modelos multiestado permitem que indivíduos se movimentem por uma série de estados ao longo do tempo, possibilitando a estimativa das probabilidades de transição e intensidades entre estados e o efeito de covariáveis associadas a cada transição. Modelos multiestado são úteis para o estudo da evolução de indivíduos ao longo do tempo, particularmente em dados agrupados, em que os indivíduos estão próximos e podem compartilhar fatores de risco comuns. A fragilidade pode ser usada em modelos multiestado para capturar a heterogeneidade entre grupos em risco para diferentes tipos de transições, assim como a estrutura de dependência entre as transições de indivíduos no mesmo grupo. As fragilidades são variáveis aleatórias não observadas, assumidas independentes entre diferentes tipos de transições, mas compartilhadas por grupos de indivíduos. Neste trabalho, propomos um modelo multiestado com fragilidade para capturar a heterogeneidade entre grupos em risco para diferentes tipos de transições, assim como a estrutura de dependência entre as transições de indivíduos no mesmo grupo. Supomos que as fragilidades são independentes entre diferentes tipos de transições, mas compartilhadas por grupos de indivíduos. Discutimos a estimação paramétrica e semiparamétrica dos parâmetros do modelo e avaliamos o desempenho do modelo em estudos de simulação. Finalmente, aplicamos o modelo a um conjunto de dados de pacientes em recuperação de transplante de medula óssea.

Palavras-chave: Modelos multiestado, modelos de fragilidade, modelos multiestado com fragilidade, fragilidade aninhada, recuperação de transplante de medula óssea.

¹Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP , São Carlos, Brasil - rocio.cm@usp.br

²Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP , São Carlos, Brasil - kevin.jv@usp.br

³Departamento de Estatística UFSCar, São Carlos, Brasil - vera@ufscar.br

1 Introdução

Na vida de um indivíduo ou sistema existem diversas causas de falha presentes ao mesmo tempo e esses eventos são os chamados eventos competitivos. Então, dizemos que essas causas competem entre si para que ela falhe. Porém, apenas uma dessas causas é responsável por determinada falha, evitando assim a ocorrência das demais. Esta abordagem foi desenvolvida por diferentes autores (Klein and Andersen, 2005; Bakoyannis and Touloumi, 2012; Hudgens et al., 2014; Andersen and Keiding, 2002; Putter et al., 2007; Pintilie, 2006; Andersen et al., 2012; Willekens, 2014).

O modelo de risco competitivo é um caso especial de modelos multiestados, que descreve transições entre eventos intermediários. Modelos multiestado são frequentemente usados para descrever dados longitudinais, que são definidos por um processo estocástico de tempo contínuo. Neste trabalho, os indivíduos transitam por um número finito de estados, possibilitando estimar as probabilidades e os riscos de transição entre eles, bem como os efeitos das covariáveis associadas a cada transição. A complexidade dos modelos multiestados depende do número de estados definidos e do número de transições permitidas entre eles. Em contraste, o problema da heterogeneidade devido a factores de risco não observados, tais como variáveis ambientais e genéticas, ou factores que foram inadvertidamente omitidos durante a fase de planeamento, foi investigado pela primeira vez por Vaupel et al. (1979) e propuseram um modelo de efeitos aleatórios para dar conta desta heterogeneidade inerente, que posteriormente abriu caminho para um extenso corpo de literatura agora referido como modelos de fragilidade. Hoje, numerosos esforços de investigação aplicaram modelos de fragilidade para enfrentar este desafio, exemplificados por estudos notáveis como Molina et al. (2021); Calsavara et al. (2020); Vasquez et al. (2022); Cancho et al. (2020); Caroni et al. (2010); Ata and Özel (2013), e muitos outros.

2 Modelo multiestado com fragilidade aninhada

O modelo de fragilidade aninhado é apropriado em situações onde os dados são naturalmente agrupados em diferentes níveis hierárquicos. Portanto, para considerar a heterogeneidade entre grupos em risco para diferentes tipos de transição, devido a fatores de risco comuns não observados, e para considerar a estrutura de dependência entre transições de indivíduos dentro do mesmo grupo, um modelo multiestado com fragilidade aninhada é definido através da transição- função de risco condicional específica do tipo q ($q = 1, \dots, Q$). Inicialmente consideraremos que existem grupos H , que são independentes entre si. Porém, existe dependência entre os indivíduos dentro de cada grupo h ($h = 1, \dots, H$), e dentro de cada grupo h , existe dependência entre as transições de seus indivíduos. Portanto, para alguns indivíduos i ($i = 1, \dots, n_h$) do grupo h e transição específica q , a função de risco condicionada à fragilidade u_{qh} é dada por

$$\lambda_{qhi}(t|u_{qh}) = u_{qh} \lambda_{q0}(t) \exp(\beta_q^\top X_{qhi}), \quad (1)$$

onde, $\lambda_{q0}(t)$ é a função de risco basal, β_q^\top é o vetor de parâmetros de regressão e X_{qhi} é o vetor de covariáveis para a transição q . Consideramos o termo de fragilidade u_{qh} para a transição q dos indivíduos do grupo h como uma realização não observada da variável aleatória, tal que

$$U_{qh} = V_h W_{qh},$$

com

$$\begin{aligned} V_h &\stackrel{iid}{\sim} f_V(v, \theta_v), \quad h = 1, \dots, H, \\ W_{qh} &\stackrel{iid}{\sim} f_W(w, \theta_q), \quad h = 1, \dots, H, \\ V_h &\perp W_{qh}, \forall (h, q), \end{aligned} \quad (2)$$

onde V_h representa o efeito geral do grupo e W_{qh} representa o efeito aleatório no nível de transição no grupo h , considerando a interação entre o tipo de transição e o grupo (Rotolo, 2013).

Na Figura 1, mostramos os casos particulares obtidos a partir do modelo de fragilidade aninhado (1):

1. Modelo multiestado sem fragilidade: ocorre quando $v_h \equiv 1$ e $w_{qh} \equiv 1, \forall h$ e q .
2. Modelo multiestado com fragilidade compartilhada apenas para o efeito de grupo: Se $w_{qh} \equiv 1$, então são obtidas fragilidades compartilhadas, resultando em $u_{qh} = u_{q'h} = v_h$. Assim, temos o modelo com efeito aleatório apenas para o grupo, que é o caso do modelo de fragilidade compartilhada
3. Modelo multiestado com fragilidade compartilhada apenas para o efeito de interação do grupo de transição: Quando $v_h \equiv 1$, as fragilidades são consideradas independentes entre as transições, e o modelo se reduz ao caso em que $\text{Corr}(U_{qh}, U_{q'h}) = 0$ ($U_{qh} \perp U_{q'h}$), semelhante aos modelos propostos por Ma et al. (2010), Yen et al. (2010) e Lique et al. (2012). Esta suposição de fragilidade independente dentro de grupos ignora a estrutura de correlação de diferentes eventos que são influenciados pelos mesmos fatores de risco não observados para indivíduos dentro do mesmo grupo

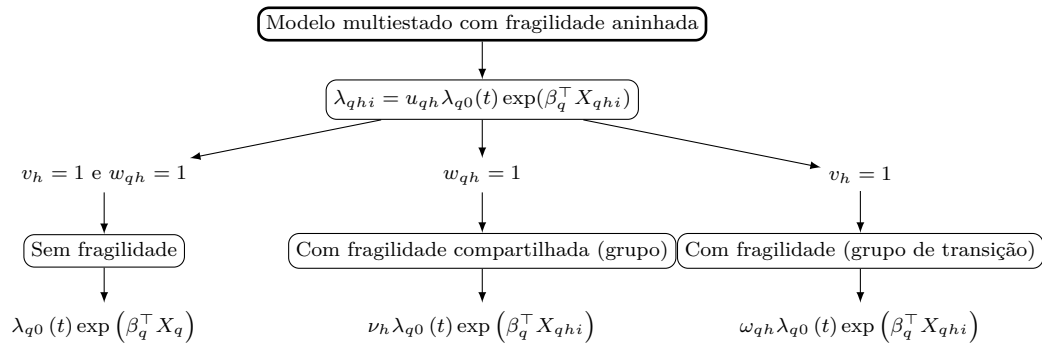


Figure 1: Casos particulares do modelo multiestado com fragilidade aninhada.

3 Inferência

Para este trabalho o procedimento inferencial para os parâmetros do modelo com base nas abordagens paramétrica e semiparamétrica. Rotolo (2013) mostrou que os indivíduos só estarão em risco para um determinado tipo de evento a partir do momento em que entrarem no estado inicial associado a esse evento. De acordo com a estrutura multiestado, para cada

indivíduo i ($i = 1, \dots, n_h$), existem Q tipos de transição possíveis. Consideraremos um total de n indivíduos agrupados em grupos H de tamanhos n_1, \dots, n_H , com $n = \sum_{h=1}^H n_h$. O tempo de transição, t_{qhi} , para o indivíduo i no grupo h pode ser observado ou censurado, com o indicador de evento/censura δ_{qhi} definido como:

$$\delta_{qhi} = \begin{cases} 1, & \text{se o tempo na transição } q \text{ para o indivíduo } i \text{ dentro do grupo } h \text{ for observado,} \\ 0, & \text{se o tempo na transição } q \text{ para o indivíduo } i \text{ dentro do grupo } h \text{ for censurado.} \end{cases}$$

Devido às suposições de independência condicional dadas em (2), a função de verossimilhança condicional para o modelo (1) pode ser obtida como o produto de todos os grupos e as contribuições dos indivíduos apenas nas transições pelas quais eles estão em risco, Ou seja, aqueles que têm tempo de truncamento esquerdo finito. Desta forma, a função de verossimilhança condicional é definida como:

$$\begin{aligned} L_C(\xi, \beta) &= \prod_{h=1}^H L_{c,h}(\xi, \beta) = \prod_{h=1}^H \prod_{i=1}^{n_h} \prod_{q=1}^Q [\lambda_{qhi}(t_{qhi}|v_h w_{qh})]^{\delta_{qhi}} \frac{S(t_{qhi}|v_h w_{qh})}{S(\tau_{qhi}|v_h w_{qh})} \\ &= \prod_{h=1}^H \prod_{i=1}^{n_h} \prod_{q=1}^Q \left[v_h w_{qh} \lambda_{q0}(t_{qhi}) \exp(\beta_q^\top X_{qhi}) \right]^{\delta_{qhi}} \\ &\quad \times \exp \left\{ -v_h w_{qh} [\Lambda_{q0}(t_{qhi}) - \Lambda_{q0}(\tau_{qhi})] \exp(\beta_q^\top X_{qhi}) \right\}, \end{aligned} \quad (3)$$

onde τ_{qhi} são os tempos de truncamento à esquerda, ou seja, aqueles correspondentes ao tempo da transição anterior, $\xi = (\xi_1^\top, \dots, \xi_Q^\top)^\top$ é o vetor de parâmetros da função de risco de linha de base, e $\beta = (\beta_1^\top, \dots, \beta_Q^\top)^\top$ é o vetor de parâmetros de regressão. As variáveis aleatórias não são observáveis e a função de verossimilhança (3) não pode ser maximizada diretamente, portanto consideraremos uma abordagem semiparamétrica.

4 Aplicação

Para ilustrar a aplicabilidade dos modelos previamente estudados, utilizamos um conjunto de dados de transplante de medula óssea. Utilizamos o conjunto de dados do livro de Klein and Moeschberger (2003), que também está disponível no pacote de Wreede et al. (2010); ? do software R (R Core Team, 2019). Este conjunto de dados trata do processo de recuperação do transplante de medula óssea, que é um tratamento padrão para a leucemia aguda. A recuperação após o transplante é um processo complexo, pois o prognóstico de recuperação pode depender de fatores de risco conhecidos no momento do transplante, como a idade do paciente e do doador, além de outros fatores. O prognóstico final também pode mudar à medida que a história pós-transplante do paciente se desenvolve com a ocorrência de eventos durante o processo de recuperação, como o desenvolvimento de doença aguda do enxerto versus hospedeiro (AGvHD), recuperação de plaquetas para níveis normais, retorno de granulócitos para níveis normais ou o desenvolvimento de infecções. O transplante pode ser considerado um fracasso quando há recorrência da leucemia no paciente ou quando o paciente morre enquanto a doença está em remissão, ou seja, a morte está relacionada ao tratamento.

Estudamos quatro modelos semiparamétricos para o conjunto de dados do transplante de

medula óssea. Os modelos propostos para a aplicação foram, Modelo 1 que é o modelo multiestado sem fragilidade (MMWF), Modelo 2 é modelo multiestado com fragilidade compartilhada apenas para efeito de grupo (MMSF1), Modelo 3 é modelo multiestado com fragilidade compartilhada apenas para efeito de interação de grupo de transição (MMSF2) e o Modelo 4 é o modelo multiestado com fragilidade aninhada (MMNF). Ajustamos os modelos semiparametricamente e as estimativas dos parâmetros dos modelos MMWF, MMSF1 e MMSF2 foram obtidas com o auxílio do pacote `mstate`, enquanto o MMNF foi ajustado com o auxílio do pacote `mlfm` (Rotolo and Horny, 2012), sob a restrição que todas as interações do grupo de transição W_{qh} têm a mesma variância ($\theta_W := \theta_1 = \dots = \theta_{10}$), esta interação sendo agora denotada por W_q . Em 1 são dados os resultados das estimativas dos parâmetros de regressão $\beta_q = (\beta_1, \beta_2)$ ($q = 1, \dots, 10$) e os respectivos erros padrão (SE) para estes modelos. Em relação aos resultados dos Table 1, as estimativas dos parâmetros de regressão permanecem positivas ou negativas em todos os modelos, os seus valores são geralmente muito próximos entre si e os EP são próximos de zero para estes modelos.

Referências Bibliográficas

- Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (2012). *Statistical models based on counting processes*. Springer Science & Business Media.
- Andersen, P. K. and N. Keiding (2002). Multi-state models for event history analysis. *Statistical methods in medical research* 11(2), 91–115.
- Ata, N. and G. Özel (2013). Survival functions for the frailty models based on the discrete compound poisson process. *Journal of Statistical Computation and Simulation* 83(11), 2105–2116.
- Bakoyannis, G. and G. Touloumi (2012). Practical methods for competing risks data: a review. *Statistical methods in medical research* 21(3), 257–272.
- Calsavara, V. F., E. A. Milani, E. Bertolli, and V. Tomazella (2020). Long-term frailty modeling using a non-proportional hazards model: Application with a melanoma dataset. *Statistical methods in medical research* 29(8), 2100–2118.
- Cancho, V. G., M. A. Macera, A. K. Suzuki, F. Louzada, and K. E. Zavaleta (2020). A new long-term survival model with dispersion induced by discrete frailty. *Lifetime data analysis* 26(2), 221–244.
- Caroni, C., M. Crowder, and A. Kimber (2010). Proportional hazards models with discrete frailty. *Lifetime data analysis* 16(3), 374–384.
- de Wreede, L., M. Fiocco, and H. Putter (2010). The `mstate` package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* 99(3), 261–74.
- Hudgens, M. G., C. Li, and J. P. Fine (2014). Parametric likelihood inference for interval censored competing risks data. *Biometrics* 70(1), 1–9.
- Klein, J. and M. Moeschberger (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.
- Klein, J. P. and P. K. Andersen (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* 61(1), 223–229.
- Liquet, B., J. Timsit, and V. Rondeau (2012). Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC Medical Research Methodology* 12(1), 79.
- Ma, T., I. Joly, and C. Raux (2010). A shared frailty semi-parametric markov renewal model for travel and activity time-use pattern analysis.
- Molina, K. C., V. F. Calsavara, V. D. Tomazella, and E. A. Milani (2021). Survival models induced by zero-modified power series discrete frailty: Application with a melanoma data set. *Statistical Methods in Medical Research* 30(8), 1874–1889.
- Pintilie, M. (2006). *Competing risks: a practical perspective*, Volume 58. John Wiley & Sons.
- Putter, H., M. Fiocco, and R. B. Geskus (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* 26(11), 2389–2430.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rotolo, F. (2013). *Frailty multi-state models for the analysis of survival data from multicenter clinical trials*. Tese (doutorado em ciências estatísticas), Università degli Studi di Padova, Pádua - Itália.
- Rotolo, F. and G. Horny (2012). `mlfm`: Semiparametric multilevel frailty models, estimated via `empl` method. In *R package*.
- Vasquez, J. K., J. Rodrigues, and N. Balakrishnan (2022). A useful variance decomposition for destructive waring regression cure model with an application to hiv data. *Communications in Statistics-Theory and Methods* 51(20), 6978–6989.

Table 1: Regression parameter estimates($\hat{\beta}_q$) and standard error of the regression parameters $SE(\hat{\beta}_q)$ for the MMWF, MMSF1, MMSF2 and MMNF.

	MMWF			MMSF1		
	$\hat{\beta}_q$	$\exp(\hat{\beta}_q)$	$SE(\hat{\beta}_q)$	$\hat{\beta}_q$	$\exp(\hat{\beta}_q)$	$SE(\hat{\beta}_q)$
$\beta_{1,1}$	0.0211	1.0213	0.0156	0.0278	1.0282	0.0163
$\beta_{1,2}$	0.0933	1.0978	0.0624	0.0978	1.1027	0.0632
$\beta_{1,3}$	0.0840	1.0876	0.1079	0.1003	1.1055	0.1108
$\beta_{1,4}$	-0.0610	0.9409	0.0552	-0.0535	0.9479	0.0561
$\beta_{1,5}$	0.0114	1.0115	0.0373	0.0184	1.0186	0.0378
$\beta_{1,6}$	0.0356	1.0362	0.0267	0.0431	1.0441	0.0271
$\beta_{1,7}$	0.0135	1.0136	0.0364	0.0217	1.0220	0.0371
$\beta_{1,8}$	-0.0430	0.9579	0.1001	-0.0675	0.9347	0.1004
$\beta_{1,9}$	-0.1586	0.8533	0.1202	-0.1758	0.8388	0.1291
$\beta_{1,10}$	0.0413	1.0422	0.0517	0.0462	1.0473	0.0526
Variance of random effects						
Hospital	-	-	-	-	0.0471	-
Transition	-	-	-	-	-	-
	MMSF2			MMNF		
	$\hat{\beta}_q$	$\exp(\hat{\beta}_q)$	$SE(\hat{\beta}_q)$	$\hat{\beta}_q$	$\exp(\hat{\beta}_q)$	$SE(\hat{\beta}_q)$
$\beta_{1,1}$	0.0398	1.0406	0.0170	0.0272	1.0276	0.0159
$\beta_{1,2}$	0.0953	1.0999	0.0628	0.0974	1.1023	0.0631
$\beta_{1,3}$	0.0840	1.0876	0.1079	0.0987	1.1037	0.1101
$\beta_{1,4}$	-0.1062	0.8992	0.0611	-0.0541	0.9474	0.0560
$\beta_{1,5}$	0.0114	1.0115	0.0373	0.0181	1.0182	0.0377
$\beta_{1,6}$	0.0356	1.0362	0.0267	0.0426	1.0436	0.0270
$\beta_{1,7}$	0.0244	1.0247	0.0377	0.0212	1.0214	0.0370
$\beta_{1,8}$	-0.0430	0.9579	0.1001	-0.0657	0.9364	0.1000
$\beta_{1,9}$	-0.1586	0.8533	0.1202	-0.1742	0.8401	0.1282
$\beta_{1,10}$	0.1269	1.1353	0.0808	0.0463	1.0474	0.0522
$\beta_{2,1}$	-0.0217	0.9785	0.0149	-0.0148	0.9854	0.0138
$\beta_{2,2}$	0.0027	1.0027	0.0515	0.0017	1.0017	0.0520
$\beta_{2,3}$	-0.0133	0.9868	0.0869	-0.0188	0.9814	0.0880
$\beta_{2,4}$	0.1122	1.1187	0.0417	0.0967	1.1015	0.0479
$\beta_{2,5}$	0.0198	1.0200	0.0392	0.0176	1.0177	0.0399
$\beta_{2,6}$	-0.0205	0.9797	0.0268	-0.0239	0.9764	0.0273
$\beta_{2,7}$	-0.0319	0.9686	0.0374	-0.0308	0.9697	0.0368
$\beta_{2,8}$	0.0808	1.0842	0.0856	0.0794	1.0827	0.0840
$\beta_{2,9}$	0.1483	1.1599	0.1770	0.1606	1.1742	0.1743
$\beta_{2,10}$	-0.1037	0.9015	0.1260	-0.0367	0.9640	0.1036
Variance of random effects						
Hospital	-	-	-	-	0.0401	-
Transition	*	*	*	*	2.1381	*

Vaupel, J., K. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3), 439–454.

Willekens, F. (2014). *Multistate analysis of life histories with R*. Springer.

Yen, A., T. Chen, S. Duffy, and C.-D. Chen (2010). Incorporating frailty in a multi-state model: application to disease natural history modelling of adenoma-carcinoma in the large bowel. *Statistical Methods in Medical Research* 19(5), 529–546.