Multistate Models for Colorectal Cancer Screening Evaluation

L. Briollais[1,2]

[1] *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada*

[2] *Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada*

ABSTRACT

The evaluation of screening procedures for cancer detection in high-risk families raises many statistical challenges due to the complexity of disease and screening processes. We discuss the application of multi-state (MS) models in this context of screening by colonoscopy and sigmoidoscopy in Lynch Syndrome families, characterized by very high-risk of colorectal cancer (CRC) and other cancers. The states are defined based on individuals' screening results, whose values are measured intermittently at each screening visit and vary across individuals. We propose a novel Markov-renewal MS model to assess how screening intervals affect different transition intensities and the probability of transitioning to CRC by assuming that the transition intensity functions are related to the elapsed time since the last screening visit. This model allows the transition intensities to depend on an individuals baseline characteristics and time-dependent covariates such as the type of the polyps.

We discuss here the estimation of transition intensities, covariate effects, and transition probabilities for the proposed Markov-renewal model. An ascertainment-corrected likelihood is proposed to deal with the non-random sampling of families. Simulation study results indicate that our approach generally provides unbiased log relative risk and probability estimates with small standard errors. The methodology is applied to a series of 18 large LS families from Newfoundland harbouring a founder MSH2 gene mutation. Our main result shows that the probability of transitioning to CRC is reduced by half for an individual with a polyp detected vs. no polyp detected, provided that individuals are screened every 2 years.

Key words: Multi-state model; Markov-renewal process; Intermittent observation; Cancer screening; Lynch syndrome family; Colonoscopy.

RESUMÉ

L évaluation des procédures de dépistage de cancers dans les familles à risque élevé entraînent de multiples défis statistiques résultant de la complexité des processus de maladie et de dépistage. On discute l'application des modèles multi-états (MME) dans le cadre du dépistage par colonoscopie et sigmoïdoscopie dans les familles ayant le syndrome de Lynch, qui se caractérisent par un très haut risque de cancer coloréctal et d'autres cancers. Les états sont définis à partir des résultats de dépistage, évalués de manière intermittente à chaque visite de dépistage et qui varient entre individus. On propose un nouveau modèle MME de renouvellement Markovien afin d'évaluer l'influence des intervalles de dépistage sur les intensités de transition et la probabilité de transitionner à létat de cancer, en supposant les fonctions de transition d'intensité dépendantes du temps survenu depuis la dernière visite de dépistage. Ce modèle permet aux transitions d'intensité de dépendre des charactéristiques d'un individu à l'origine et de variables dépendantes du temps comme le type de polypes.

On discute ici l'estimation des transitions d'intensité, de l'effet des covariables et des probabilités de tran-

sition du modèle de renouvellement Markovien proposé. Une correction de la vraisemblance pour le biais de sélection est proposée, prenant en compte la sélection non-aléatoire des familles. Les résultats de l'étude de simulations indiquent que notre approche donne en général des estimés non-biaisés du log des risques relatifs et des probabilités de transition avec de petites erreurs standards. La méthode est ensuite appliquée à une série de familles de Terre-Neuve porteuses d'une mutation fondatrice du gène MSH2. Nos résultats principaux montrent que la probabilité de transitionner à l'état de cancer est réduite par deux pour un individu qui a eu un polype détecté comparé à un quelqu'un sans polype détecté, en supposant que les individus sont dépistés tous les deux ans.

Mots clés: Modèles multi-états; Processus de renouvellement de Markov; Observation intermittente; Dépistage du cancer; Familles avec syndrome Lynch; Colonoscopie.

1. INTRODUCTION

The goal of this paper is to develop a statistical framework for CRC screening evaluation in Lynch Syndrome (LS) families based on a new Markov-renewal illness-death model and to be able to: 1) Account for the natural history of CRC and its development either through an intermediate polyp state or directly from a "no-polyp" state; 2) estimate transition intensities and transition probabilities to the polyp state and to a primary CRC state for someone with a prior history of adenoma or no adenoma; 3) analyze event times that are only observed intermittently at screening visits; 4) incorporate the effect of elapsed time since the last visit time; 5) assess the effect of screening for individuals entered the prescheduled CRC screening program vs those who did not enter this program; 6) be applicable to family data collected through non-random ascertainment.

2. MULTISTATE (MS) MODELS AND INTERMEDIATE OBSERVATIONS

In observational studies, individuals are seen intermittently at follow-up visits (e.g. colonoscopies), and clinical information about disease states and covariate values are recorded. The clinical information on events between visits is usually missing or incomplete. In fitting MS models, the difficulty with intermittent observations is that the exact transition times, and sometimes the states visited between successive observation times, are unknown. Irregular intermittent observations refers to the case where the visit times vary across individuals. Many researchers have studied Markov MS models with intermittent observations (see [2, 3]), and others have studied semi-Markov models in this context [6, 7].

A commonly-used MS model is the illness-death model, which is used in medical research for describing terminal and intermediate events. In Section 3 we consider an illness-death model shown in Figure 1 to analyze the occurrence of polyps and CRC for the family members in the Newfoundland (NL) data. We define the three states in the model as: state 1, never had any polyp detected, state 2, had at least one polyp detected (benign or adenoma), and state 3, CRC detected.

Continuous-time MS models with R states can be specified in terms of transition intensity functions. Let $Y(t)$ denote the state occupied at time $t \geq 0$ and $Z(t)$ represent a relevant covariate vector at time t. The

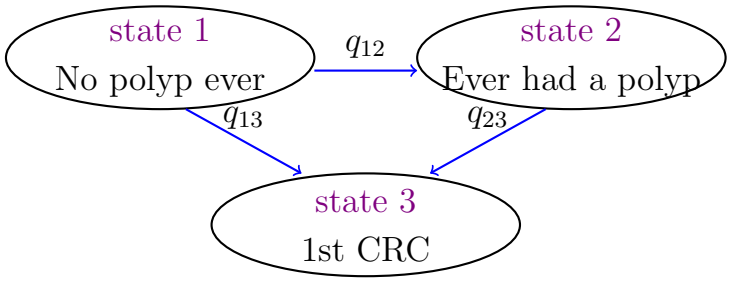

Figure 1: Three-state model for polyp and CRC occurrences.

transition intensity from state r to state s is given by

$$q_{rs}(t|H(t)) = \lim_{\Delta t \to 0} \frac{P\{(Y(t+\Delta t^-) = s|Y(t^-) = r, H(t))\}}{\Delta t}, \quad r \neq s, \tag{1}$$

where $H(t)$ denotes the history $\{Y(u), Z(u), 0 \leq u < t\}$ of the covariates and multistate process.

In observational studies, where individuals are seen intermittently at visit times, the information about the disease states and time-dependent covariates might be unknown at times between visits. In these situations, it is often assumed that the MS model is Markov; that is $q_{rs}(t|H(t))$ depends only on t and $Y(t^-)$ (e.g. [4] and [3]) if there are no covariates. With covariates, transition intensities are often taken to be of the form

$$q_{rs}(t|H(t)) = q_{0rs} exp(\beta_{rs}^T Z(t)), \quad r \neq s, \tag{2}$$

where β_{rs} is a vector of regression coefficients for the transition intensity from state r to s. We note that covariates $Z(t)$ are often also observed only at visit times, so even though models without covariates can be easily fitted, time varying covariates pose problems. In this paper we focus on non-Markov models.

3. ANALYSIS UNDER MARKOV-RENEWAL ILLNESS-DEATH MODEL

3.1. A Markov-renewal model for screening and disease outcomes

Consider the illness-death model displayed in Figure 1 with states $1-3$ defined as follows: State 1, an individual never had any polyp; State 2, an individual has had at least one polyp during their lifetime; and State 3 an individual has been diagnosed with a primary CRC. Let $N(t)$ represent the number of observation times by time t for an individual. Further define the indicator functions $B(t)$ and $C(t)$ as follows:

$$B(t) = \begin{cases} 0, & \text{if } (t - t_{N(t^-)}) \leq \text{b}, \\ 1, & \text{if } (t - t_{N(t^-)}) > \text{b}, \end{cases} \tag{3}$$

and

$$C(t) = \begin{cases} 0, & \text{if } (t - t_{N(t^-)}) \leq \text{c}, \\ 1, & \text{if } (t - t_{N(t^-)}) > \text{c}, \end{cases} \tag{4}$$

where $(t - t_{N(t^-)})$ is the elapsed time since the last visit, and b and c are pre-specified cut points. For the transition intensities in our illness-death model, we assume the following forms:

$$q_{12}(t; Z) = q_{012} exp(\beta_{12}^T Z(t)), \tag{5}$$

$$q_{13}(t; Z) = q_{013} exp(\beta_{13}^T Z(t) + \gamma B(t)), \tag{6}$$

$$q_{23}(t; Z) = q_{023} exp(\beta_{23}^T Z(t) + \eta C(t)), \tag{7}$$

where γ and η are regression coefficients associated with $B(t)$ and $C(t)$, respectively. In this proposed Markov-renewal model, the transition intensities $1 \to 3$ and $2 \to 3$ depend on the history of the joint disease-visit process through the present state and the time since last visit. This Markov-renewal model is different from the semi-Markov model for which the sojourn times depend on the history of the disease process through the present state and the time since entry into that state (e.g. [5, 1]).

3.2 Transition probabilities for the Markov-renewal model

We here address the calculation of transition probabilities for the model (5)-(7). For simplicity, we suppose that $Z(t) = Z$ in (5)-(7). Let t_{j-1} and t_j be the times for two successive visits. The transition probabilities of starting from state r at age t_{j-1} and being in state s at age t_j given the covariates values Z, are represented as $P_{rs}(t_{j-1}, t_j; Z), \quad r, s = 1, 2, 3$. The probability of being observed in state 2 at age t_j given the person was seen in state 2 at age t_{j-1} is given by

$$P_{22}(t_{j-1}, t_j; Z) = \begin{cases} exp\{-q_{023} exp(\beta_{23}^T Z)(t_j - t_{j-1})\}, & \text{if } (t_j - t_{j-1}) \le c, \\ exp\{-q_{023} exp(\beta_{23}^T Z) \times c\} \times exp\{-q_{023} exp(\beta_{23}^T Z + \eta)(t_j - t_{j-1} - c)\}. & \text{if } (t_j - t_{j-1}) > c, \end{cases} \tag{8}$$

Since there is only one state to reach from state 2, we have

$$P_{23}(t_{j-1}, t_j; Z) = 1 - P_{22}(t_{j-1}, t_j; Z). \tag{9}$$

Similarly $P_{11}(t_{j-1}, t_j; Z)$ is given by

$$\begin{cases} exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z))(t_j - t_{j-1})\}, & \text{if } (t_j - t_{j-1}) \le b, \\ exp\{-(q_{012} exp(\beta_{12}^T Z + q_{013} exp(\beta_{13}^T Z)) \times b\} & \text{if } (t_j - t_{j-1}) > b, \\ \times exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z + \gamma))(t_j - t_{j-1} - b)\}. \end{cases} \tag{10}$$

Similarly, again for $P_{13}(t_{j-1}, t_j; Z)$ we have

$$\begin{cases} q_{013} exp(\beta_{13}^T Z) \int_{t_{j-1}}^{t_j} exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z))(w - t_{j-1})\} dw, & \text{if } (t_j - t_{j-1}) \le b, \\ q_{013} exp(\beta_{13}^T Z) \int_{t_{j-1}}^{t_{j-1}+b} exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z))(w - t_{j-1})\} dw & \text{if } (t_j - t_{j-1}) > b, \\ + \int_{t_{j-1}+b}^{t_j} exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z + \gamma))(t_j - t_{j-1} - b)\} dw \\ \times exp\{-(q_{012} exp(\beta_{12}^T Z) + q_{013} exp(\beta_{13}^T Z)) \times b\} \times q_{013} exp(\beta_{13}^T Z + \gamma). \end{cases} \tag{11}$$

The probability of being observed in state 2 at age t_j given the person was seen in state 1 at age t_{j-1} is given by

$$P_{12}(t_{j-1}, t_j; Z) = 1 - P_{11}(t_{j-1}, t_j; Z) - P_{13}(t_{j-1}, t_j; Z). \tag{12}$$

3.3 Measuring screening effect under the Markov-renewal model

The proposed Markov-renewal MS model can be used to measure screening effect and also to evaluate the impact of the last gap time on the transition to the CRC state among individuals who have been screened. We can measure the impact of screening on transition intensities and transition probabilities by allowing the transition intensities to depend on an individual's screening status. The primary aim of screening is to reduce the rate of transitioning to the cancer state and consequently to increase the probability of staying in state two. If the effect of screening is associated with a higher rate of polyp detection and lower CRC risk, then it is also of interest to estimate $P_{13}(t)$ for the two possible paths by which individuals can enter state 3; a direct transition from state 1 to state 3, or state 1 to state 3 but passing through state 2. In this situation, we expect to see a larger proportion of $1 \to 3$ transitions and consequently a smaller proportion of $1 \to 2 \to 3$ transitions among non-screened individuals compared to screened individuals.

4. APPLICATION TO NEWFOUNDLAND LYNCH SYNDROME FAMILIES

The Newfoundland (NL) data consists of 322 MSH2 gene mutation carriers in 18 large LS families. Our analyses focused on 253 individuals who were non-probands and had one of the CRC tests among colonoscopy, sigmoidoscopy, and proctoscopy before a first CRC or last follow-up time. The data includes 119 males and 134 females. There were 533 colonoscopies, 56 sigmoidoscopies, and 1 proctoscopy tests reported. Of the 253 study participants, 151 (60%) entered the CRC screening program while the remaining 102 (40%) did not. The majority of non-screened individuals had only one CRC test, and among the 23 individuals who had at least two CRC tests, the reasons for screening were mostly for symptoms (56%), follow-up (24%) or unknown (20%).

One of the main interests is to sub-divide the probability $\hat{P}_{13}(t)$ into probability of a direct transition $1 \to 3$ vs. passing through state 2 ($1 \to 2 \to 3$), as displayed in figure . Indeed, if screening is efficient, we expect individuals who were screened and have a polyp detected during the course of their visit history to have a lower cancer incidence than those individuals who were screened with no history of polyp detection. Our results demonstrate that the risk to develop CRC is almost reduced by half for screened individuals who had a polyp detected (path $1 \to 2 \to 3$) compared to those without a polyp detected (direct transition $1 \to 3$), with corresponding probabilities by age 70 of 15.2% and 39.9% in females and 25.1% and 65.9% in males, respectively.

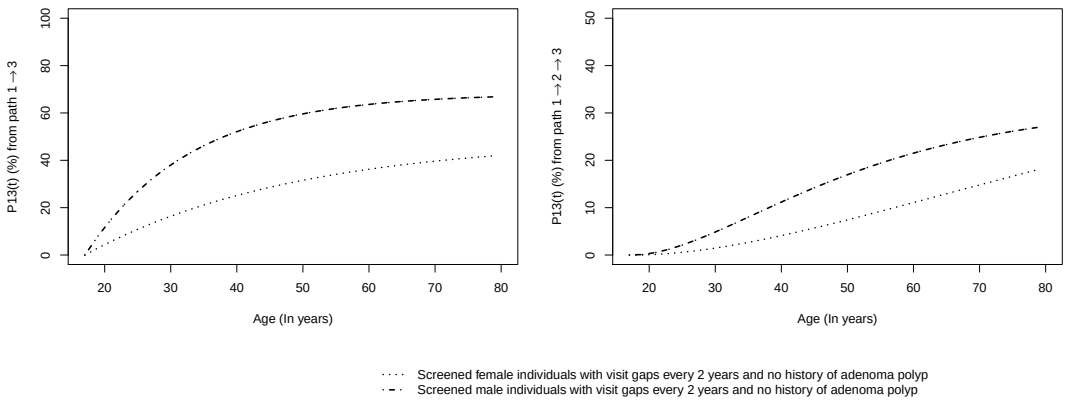

Figure 2. The Markov-renewal multistate model $P_{13}(t)$ estimates considering different paths.

References

[1] Choi YH, Briollais L, Green J, Parfrey P, Kopciuk K. Estimating successive cancer risks in Lynch Syndrome families using a progressive three-state model. *Statistics in Medicine* 2014; *33*: 618638.

[2] Cook RJ, Lawless JF. Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences* 2014; **6**: 127-161.

[3] Gentleman RC, Lawless JF, Lindsey JC, Yan P. Multistate Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* 1994; **13**: 805-821.

[4] Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* 1985; **80**: 863-871.

[5] Titman AC, Sharples LD. Model diagnostics for multi-state models. *Statistical Methods in Medical Research* 2010a; **19**: 621-651.

[6] Titman AC, Sharples LD. Semi-Markov models with phase-type sojourn distributions. *Biometrics* 2010; **66**: 742-752.

[7] Yong Y, Nair VN. Parametric inference for time-to-failure in multi-state semi-Markov models: A comparison of marginal and process approaches. *Canadian Journal of Statistics* 2011; **39**: 537-555.