総 説

生存時間解析における競合リスクモデル

Competing Risks Model in the Analysis of Survival Data

西川正子

Masako Nishikawa

国立保健医療科学院・技術評価部

Department of Technology Assessment and Biostatistics,
National Institute of Public Health,
3-6 Minami 2 chome Wako Saitama-ken 351-0197 Japan
e-mail:mnishikawa@niph.go.jp

The statistical analysis of survival time or failure time data is an important topic in many areas of research. In ordinary survival data, there is a single, possibly right-censored failure time for each individual. However, in a number of medical applications whereby data is collected an individual may experience one or more events but the first event will preclude the occurrence of another event under investigation. As a result, he/she can experience only one of several types of events. Such data are commonly referred to as competing risks data. Censoring due to such an event is generally not independent of the time to the event of interest.

This paper reviews statistical methods for analyzing a competing risks model. Both conceptual considerations and common approaches to one-sample inference; two sample comparison; and covariate effect modeling are discussed. The theory for the analysis of ordinary right-censored survival data can be applied under certain circumstances. Standard statistical software package can perform the necessary analysis, although interpretation of results will vary.

Key words: competing risks, cumulative incidence function, cause-specific hazard, non-informative censoring.

1. はじめに

ある個体あるいは対象を経時的に観察し、注目している事象が発生するまでの時間、あるいは経時的な累積発生割合を推定する問題は、医療において特定の原因による死亡や有害事象の発生等を扱う生存時間解析、ある製品の寿命に関する信頼性研究などをはじめとして、疫学・薬剤疫学、公衆衛生、人口問題、保険統計、生物・医学研究、経済学などの様々な領域で発生する。そのような場面では注目している事象以外の原因、すなわち競合リスク(競合危険, competing risks)による観察打ち切りの発生が問題になる。医学・生物学データの解析で競合リスク要因間の独立性

を仮定しているのをよく見かけるが、この仮定は多くの場合データから確認できない。生存時間関数の推定では Kaplan-Meier 推定量 (Kaplan and Meier, 1958) がよく用いられるが、競合リスク要因が存在する場合、Kaplan-Meier 推定量にはバイアスが入ることが知られている (e.g. Gaynor et al., 1993; Schwarzer et al., 2001; Southern et al., 2006)。また、たとえ独立性を仮定できたとしてもその解釈には問題がある (e.g. Kalbfleisch and Prentice, 1980; Gaynor et al., 1993; Pepe and Mori, 1993)。この数十年の間に、競合リスク要因の独立性を仮定しない様々な解析法が提案されている。

医学分野における生存時間解析の対象となるデータは、文字どおり、死亡までの時間であるほか、ある起点から特定の事象(イベント)が起こるまでの時間、例えば、治療法の比較では治療開始からある病気の治癒までの時間、副作用や合併症が発現するまでの時間、疾患が増悪するまでの時間(無増悪時間)等である。本稿では「生存時間」という用語はこれらのイベント発現までの時間を総称した一般的な意味で用いる。生存時間解析の特徴は、イベントが起こらないままで観察を終了することになる、観察打ち切りが発生するということである。

1つのイベントに対して複数の原因やイベントタイプなどで分類し、1つの原因又はイベントタイプが観測されれば他のそれらは観測できない(観察打ち切りが起こる)場合がある。このとき、ある原因又はイベントタイプは他の原因又はイベントタイプとリスクを競合し、競合リスク要因と呼ばれる。例えば、心血管系の疾患の治療法又は予防法の比較において、心血管系の疾患による死亡までの時間に関心を持つが、他の理由(がん、事故など)で死亡してしまう場合があり、それにより「心血管系の疾患による死亡」は観測できなくなる。また、心血管系の疾患により死亡した症例においては「他の理由による死亡」は観測できない。つまり、「心血管系の疾患による死亡」と「他の理由による死亡」は互いに競合リスク要因である。別な例として、心血管系の疾患の治療法又は再発予防法の安全性を、有害事象により治療を中止するまでの時間、あるいは有害事象による治療中止率で評価する場合 (e.g. Pitt et al., 2000)、治療中止前に死亡すれば、「有害事象により治療を中止」するまでの時間は観察打ち切りになる。もし、有害事象により治療中止になる前に「別な理由により治療を中止」になれば、同様に、有害事象により治療を中止するまでの時間は観察打ち切りになる。イベントを「有害事象により治療を中止」または「別な理由により治療を中止」または「治療を中止する前に死亡」と定義すれば、それぞれの「括弧」はイベントタイプを示し、いずれかのイベントタイプが発現すれば、そのほかのイベントタイプは観測できなくなる。したがって、それぞれのイベントタイプは互いに競合リスク要因である。

以降、まず、競合リスクモデルの解析に関連が深い一般的な生存時間解析の方法を概観し、競合リスクモデルと推測について解説する。また、適応例などを紹介する。

2. 観察打ち切り

イベントが確認されずに観察が打ち切りになった個体は観察打ち切り例、またはセンサー (censor) 例と呼ばれる。生存時間解析では、通常、センサー例を含む標本を対象とする。推定に偏りを与えない観察打ち切りは無情報センサー (non-informative censoring) であり、典型的なものを以下に挙げる。

(1) 試験の観察終了時点または観察終了期間があらかじめ決められているために、その時点までフォローしたがイベントが確認されずに観察打ち切りになる場合で、タイプ 1 センサー (type-1 consorig, 中村, 2001), 時間打ち切り (type I 打ち切り, 大橋, 浜田, 1995) などと呼ばれる。例えば、観察期間 24 週の試験、または○年○月○日終了の試験などが、これらに該当する。

(2) あらかじめ決められた最初のある例数のイベントが確認された時点で試験を終了することが決められていて、そのためにその時点までにイベントが確認されなかった個体についての観察を打ち切る場合で、タイプ 2 センサー (type-2 consorig, 中村, 2001), 個数打ち切り (type II 打ち切り, 大橋, 浜田, 1995) などと呼ばれる。

このほか、観察中に、イベントの発現について予測可能となるような情報を与えない理由により無作為な観察打ち切りが発生する場合なども無情報センサーである。Kalbfleisch and Prentice (1980), Cox and Oakes (1984), Lee and Wang (2003) などにも詳しい説明があるので参照して欲しい。無情報センサーであれば、観察打ち切りになった個体におけるそれ以降の将来のイベントの発現可能性は、今観察打ち切りにならずに観察を継続している個体の将来のイベントの発現可能性と同じであることを仮定できる。また、このことが仮定できるのは無情報センサーのときのみである。

追跡不能となり観察が継続できずに観察打ち切りになることもある。追跡不能となる理由は一般的に確認できないであろうが、治療の継続を被験者の判断などで恣意的にやめるのは、治療があまり好ましくない(または期待したほどには良くない)結果をもたらす治療の継続をやめたい・必要としない場合や、被験者の状態が良くなり治療の継続を必要としない場合などがあるであろう。解析上の取り扱いが簡単であるため、多くの場合で無情報センサーを仮定した取り扱いがなされているようであるが、それら追跡不能となった被験者は無情報センサーであるか否かをデータにより確認することができない。

注目するイベントよりも先にそれを妨げるイベントが発現し、観察が継続できずに観察打ち切りになることもある。注目するイベントの発現を妨げるイベントは無情報センサーではなく、情報を持つセンサー (informative censoring) であることが多い。これについては 4 章以降に詳しく述べる。ただし、次章に述べる一般的な生存時間解析の方法では、観察打ち切りは無情報センサーであることを仮定する。

3. 一般的な生存時間解析の方法

以下のように記号を定義する。 T をイベント発現までの時間 (failure time) とし、ここでは連続変数であると仮定する。 C を観察打ち切りまでの時間 (potential censoring time) とする。 $t_i, i = 1, \dots, n$, を被験者 i のイベント発現までの時間、 $c_i, i = 1, \dots, n$, を被験者 i の観察打ち切りまでの時間とする。実際には、被験者 $i, i = 1, \dots, n$, について観測されるデータは (u_i, δ_i) のペアとなっている。ここに、 $u_i = \min(t_i, c_i)$, $\delta_i = I(t_i \leq c_i)$, I は観察打ち切りの表示子 (censoring indicator) で、() 内の関係が成立すれば 1, そうでなければ 0 をとる関数である。次の 2 つを仮定する。

- i) $t_i, i = 1, \dots, n$, は独立に同じ生存時間関数 $S(t) = P(T > t), 0 < t < \infty$, に従う。
- ii) 各被験者について、観察打ち切りはイベント発現までの時間とは独立である (無情報センサー)。

すなわち、観察打ち切りになった被験者のそれ以降の将来のイベントの発現可能性は、今観察打ち切りにならずに観察を継続している被験者の将来のイベントの発現可能性と変わらないことを仮定する。

ハザード (hazard) 関数 $\lambda(t)$ は、時間 $T = t$ まで生存していた ($T = t$ まで観察が継続されていて、まだイベントが発現していない) ときに、次の瞬間に死亡 (イベント) が起こる率を意味し、

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

と定義される。 T が連続変数の場合のハザードは「 $T = t$ まで生存した者のうち、 $T = t + \Delta t$ までに死ぬ者の割合を、単位時間当たりの量に換算し、 $\Delta t \rightarrow 0$ としたときの極限值」 (中村, 2001) と解釈できる。日本語の「率」は比率、割合、変化率などの意味で使われ、これらの違いを厳密に区別していないことも多いが、これらの違いは佐藤 (2005) などに解説されている。 Δt が小さいときは、 $\lambda(t)\Delta t$ は $T = t$ まで生存した者が次の Δt の瞬間に死亡する確率の近似値となる。累積ハザード (cumulative hazard) 関数は

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

と定義される。累積分布関数と確率密度関数はそれぞれ

$$F(t) = P(T \leq t) = 1 - S(t), \quad 0 < t < \infty$$

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

のように定義され、これらの間には、以下の関係が成り立つ。

$$S(t) = \exp(-\Lambda(t))$$

$$f(t) = \lambda(t)S(t) = \lambda(t)\exp(-\Lambda(t))$$

$$\lambda(t) = \frac{-d(\log S(t))}{dt}$$

3.1 生存時間関数の推定

生存時間関数は Kaplan-Meier 推定量 (KM estimator; Product-Limit estimator, 以下, KME と略記) により求められる ($\hat{S}(t)$)。 u_i がイベント発現までの時間 ($\delta_i = 1$) であるものについて、タイがある場合には 1 つを残して時点間で重複を許さずに昇順に並び替えて添え字を $t_{(1)} < t_{(2)} < \dots < t_{(q)}, q \leq n$ のように付け直す。 T が連続変数の場合を考えているが、観測値のまるめの誤差などによりタイも起こりえる。

$$\hat{S}(t) = \prod_{i: t_{(i)} \leq t} \left(1 - \frac{d_i}{N_i}\right)$$

ここに、 N_i は時点 $t_{(i)}$ の直前までまだ観察が継続されていて、まだイベントが発現していない (つまり、リスクに曝されている, at risk) 被験者の集合 $\mathfrak{R}_i$ に属する人数で、 $\mathfrak{R}_i$ は $t_{(i)}$ でのリスク集合と呼ばれている。 d_i は時点 $t_{(i)}$ でのイベント数である。 $\hat{S}(t)$ はノンパラメトリック最尤推定量である。詳細は Kalbfleisch and Prentice (1980, 2002), Cox and Oakes (1984) などを参照

して欲しい。 $\hat{S}(t)$ は常に階段関数で、上の式では右連続である(左連続に定義することもある)。観察打ち切りとイベント発現までの時間が同じ時間である場合、慣例(e.g. Cox and Oakes, 1984; Lawless, 2003; 大橋, 浜田, 1995)はイベント発現を先として取り扱う。観察打ち切りとイベント発現までの時間が同時点であれば、観察打ち切り例のイベント発現はそれ以降であるという考えから、観察打ち切り例が後と考える立場からはこの取り扱いは当然と考えられる(e.g. Hosmer and Lemeshow, 1999; Lee and Wang, 2003)。

$\hat{S}(t)$ の各時点 t ごとの分散は、近似的に次の式で表現できる (Greenwood 式, 1926)。 t が $t_{(j)} < t \leq t_{(j+1)}$ であるとき、

$$Var(\hat{S}(t)) = (\hat{S}(t_{(j)}))^2 \left(\sum_{t_{(i)} \leq t} \frac{d_i}{N_i(N_i - d_i)} \right).$$

観察打ち切り例が全くなければ上の式を変形すれば次の式になり、成功確率を $\hat{S}(t)$ とする 2 項分布の分散として表現できる。

$$Var(\hat{S}(t)) = \frac{\hat{S}(t)(1 - \hat{S}(t))}{n}$$

各時点 t ごとの $S(t)$ の信頼係数 $100(1 - \alpha)\%$ の信頼区間は $\hat{S}(t)$ の漸近正規性により次のように近似できる。

$$\hat{S}(t) \pm \Phi_{\alpha/2} \sqrt{Var(\hat{S}(t))}.$$

ここに、 $\Phi_{\alpha/2}$ は標準正規分布の上側 $\alpha/2$ 点を示す。信頼区間が必ず 0 以上にはいるようにするために、 $\log(-\log \hat{S}(t))$ の漸近分散をもとに $S(t)$ の信頼区間を次式で求めることもある。このような変換を行ったあとで正規近似をしたほうがよい場合もある (Kalbfleisch and Prentice, 1980)。

$$\hat{S}(t)^{\exp[\pm \Phi_{\alpha/2} \sqrt{var(\hat{S}(t)) / [(\log \hat{S}(t)) \hat{S}(t)]^2}]}$$

SAS ver 9.0 から、LIFETEST プロシジャーに SURVIVAL ステートメントが追加され、 $S(t)$ の各時点ごとの信頼区間の計算方法が CONFTYPE オプションで指定できるようになった。CONFTYPE = LOGLOG により $\log(-\log \hat{S}(t))$ の漸近分散を利用した信頼区間が得られる。計算結果は OUT = SAS データセット名 または OUTSURV = SAS データセット名 を指定してファイルに出力できる。ただし、OUT オプション及び OUTSURV オプションは同時には有効にならない。

3.2 2 群間での生存時間関数の差の検定

ログランク (logrank) 検定 (Mantel, 1966) は 2 群の生存時間分布が同じという帰無仮説の検定に最も頻繁に用いられる検定方法である。ログランク検定は正確には、ハザード (hazard) の違いを比較している。2 群を併合してイベント発現までの時間を昇順に並べたときの i 番目のイベント発現が観測された時点 $t_{(i)}$ の直前の被験群 (T) 及び対照群 (A) における例数をそれぞれ n_{Ti} , n_{Ai} , 合計 N_i 例、時点 $t_{(i)}$ のイベント数をそれぞれ d_{Ti} , d_{Ai} , 合計 d_i 例とする ($i = 1, 2, \dots, q$)。いずれかの群でイベントが観測された時点ごとに、群とイベントの状態 (有・無) の分割表を作成する。

各群の生存時間関数が同じという帰無仮説の下、被験群 (または対照群) のイベント数の期待値と分散は超幾何分布のそれらであり、被験群について見るとすれば、期待値と分散はそれぞれ Jpn J Biomet Vol.29, No.2, 2008

$\hat{e}_{Ti} = (n_{Ti}d_i)/N_i$, $\hat{v}_{Ti}^2 = (n_{Ti}n_{Ai}d_i)(N_i - d_i)/\{N_i^2(N_i - 1)\}$ と推定される. 検定統計量 Q は次のように書ける.

$$Q = \frac{\{\sum_{i=1}^q (d_{Ti} - \hat{e}_{Ti})\}^2}{\sum_{i=1}^q \hat{v}_{Ti}^2}$$

帰無仮説の下, イベント数の合計と期待値の合計が十分大きいときに Q は自由度 1 の χ^2 分布に従う.

表 1. イベントが観測された時点 $t_{(i)}$ での分割表

群	イベント発現数	イベント未発現数	合計
被験群	d_{Ti}	$n_{Ti} - d_{Ti}$	n_{Ti}
対照群	d_{Ai}	$n_{Ai} - d_{Ai}$	n_{Ai}
リスク集合全体	d_i	$N_i - d_i$	N_i

層別ログランク検定は, 共変量(離散型)のそれぞれの水準が 2 群の生存時間関数 S_T , S_A に異なる影響を与えるが, 共変量のそれぞれの水準の中では群(処理)の影響は同程度であることが仮定できる状況で, 群(処理)の影響が同じという帰無仮説 $H_{0h}: S_{Th} = S_{Ah}$, $h = 1, \dots, L$ (L は共変量の水準数), の検定に用いられる検定方法である. 共変量(離散型)のそれぞれの水準により層別して, 層ごとに表 1 の分割表を作り, 層ごとに通常のログランク検定統計量の分子 (Q_{Nh}) と分母 (Q_{Dh}^2) を計算する.

$$Q_{Nh} = \sum_{i=1}^{qh} (d_{Thi} - \hat{e}_{Thi})$$

$$Q_{Dh}^2 = \sum_{i=1}^{qh} \hat{v}_{Thi}^2$$

層別ログランク検定の検定統計量 Q_{strat} は次のように書ける.

$$Q_{strat} = \frac{\{\sum_{h=1}^L Q_{Nh}\}^2}{\sum_{h=1}^L Q_{Dh}^2}$$

帰無仮説の下, イベント数の合計と期待値の合計が十分大きいときに Q_{strat} は自由度 1 の χ^2 分布に従う. 詳しくは Lawless (2003), Andersen et al. (1993) などを参照して欲しい.

3.3 比例ハザードモデル

医学生物分野での生存時間解析において, イベント発現までの時間に与える共変量(環境条件などの外的要因, 個体の特性などの内的要因)の影響を探索する, 又は調整するモデルとして, 最も頻繁に用いられるモデルは比例ハザードモデル (proportional hazard model, Cox, 1972) である. 話を簡単にするために共変量は時間に依存しない場合を考える. 共変量ベクトル $\mathbf{z}$ (p 次元) を持つ被験者の時間 t におけるハザード関数 $\lambda(t, \mathbf{z})$ は次の式で表現される.

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta' \mathbf{z})$$

ただし, $\lambda_0(t)$ は基準ハザード関数(共変量ベクトル $\mathbf{z} = \mathbf{0}$ の被験者のハザード関数), β は共変量の寄与を示す回帰係数 $\beta' = (\beta_1, \dots, \beta_p)$ である. Cox の比例ハザードモデルは比例ハザード性と対数線形性を仮定し, 共変量ベクトル $\mathbf{z}$ を持つ被験者のハザード関数は, 比例定数を $\exp(\beta' \mathbf{z})$ として基準ハザード(baseline hazard)関数に比例すると仮定したモデルである. このモデル下では, 共変量ベクトル $\mathbf{z}$ (p 次元)を持つ被験者の時間 t における累積ハザード関数及び生存時間関数についてそれぞれ以下の関係が得られる.

$$\Lambda(t, \mathbf{z}) = \Lambda_0(t) \exp(\beta' \mathbf{z})$$

$$S(t, \mathbf{z}) = S_0(t)^{\exp(\beta' \mathbf{z})}$$

ただし, $\Lambda_0(t)$ は基準累積ハザード関数(共変量ベクトル $\mathbf{z} = \mathbf{0}$ の被験者の累積ハザード関数), $S_0(t)$ は共変量ベクトル $\mathbf{z} = \mathbf{0}$ の被験者の生存時間関数である. 観測されるデータは $(u_i, \delta_i, \mathbf{z}_i)$ である. ここに, $\mathbf{z}_i$ は被験者 i ($i = 1, \dots, n$) の共変量ベクトルである. 観察打ち切りまでの時間は共変量を与えた下で生存時間と独立で, β に対して無情報センサーであることを仮定する. β は, 次の式で示す部分尤度 $L_p(\beta)$ (Cox, 1975) を最大化することにより推定する.

$$L_p(\beta) = \prod_i \left[\frac{\exp(\beta' \mathbf{z}_i)}{\sum_{k \in \mathfrak{R}_i} \exp(\beta' \mathbf{z}_k)} \right]^{\delta_i}$$

ここに, $\mathfrak{R}_i$ は死亡時間 u_i でのリスク集合 $\{k: u_i \leq u_k\}$ を示す. 部分尤度に基づいて得られる β の推定量は通常の尤度に基づく推定量と同様な良い性質を持っている. 詳しくは Kalbfleisch and Prentice (2002, 4 章)などを参照して欲しい. タイの扱いを含めた回帰係数の推定方法としては, タイになった被験者の間でのすべての起こりえる順序についての平均部分尤度を用いる方法 (Kalbfleisch and Prentice, 1980; DeLong et al., 1994), Breslow 法 (1974), Efron 法 (1977), 等がある. 平均部分尤度を用いる方法は, exact expression (e.g. Hosmer and Lemeshow, 1999; SAS Institute Inc., 2004) とも呼ばれている. Efron 法は, タイの調整時に Breslow 法よりも分母が小さくなる. Kalbfleisch and Prentice (1980, 2002) は, タイの割合 (d_i/N_i) が大きいときには, Breslow 法はバイアスが大きいかもしいこと, Efron 法は Breslow 法よりかなり良いこと, などにより, 平均部分尤度または Efron 法を使うのを薦めている. 一方, Hosmer and Lemeshow (1999) は, タイが多発した実際の試験のデータによりこれらの三つの方法による相違を検討し, 大きな差はなかったのでシンプルな Breslow 法を薦めている. 基準累積ハザード関数の推定には Breslow 法 (1974) などがある.

層ごとに基準ハザード関数は異なり共変量ベクトルの寄与は層を通して同じであるようなモデルを仮定することもできる. h 層に属する被験者の共変量ベクトルが $\mathbf{z}$ であるときのハザード関数は

$$\lambda_h(t, \mathbf{z}) = \lambda_{0h}(t) \exp(\beta' \mathbf{z})$$

と書ける. $\lambda_{0h}(t)$ は h 層の基準ハザード関数(h 層に属する被験者の共変量ベクトルが $\mathbf{z} = \mathbf{0}$ であるときのハザード関数)である. 回帰係数 β の推定方法については Kalbfleisch and Prentice (2002, 4 章)などを参照して欲しい.

4. 競合リスクが存在するときの解析

2章で説明したように、推定に偏りを与えない観察打ち切りは無情報センサーであり、情報を持つセンサーを無情報センサーとして取り扱えば推定値に偏りをもたらす。逆に、無情報センサーであるときのみ、観察打ち切りを受けた個体のそれ以降の将来のイベントの発現可能性は、今観察打ち切りを受けずに継続している個体の将来のイベントの発現可能性と変わらないことを仮定できる。追跡不能となり観察が継続できずに観察打ち切りになる場合や注目するイベントよりも先にその発現を妨げるイベントが発現して、観察が続けられずに観察打ち切りになる場合、観察を継続している個体よりもイベントの発現しやすい個体が観察打ち切りされた時にそれらを無情報センサーとして扱えば発現率は過小評価される。逆に、観察を継続している個体よりもイベントの発現しにくそうな個体が観察打ち切りされた時にそれらを無情報センサーとして扱えば発現率は過大評価される。無情報でない観察打ち切りを無情報センサーとして扱った場合のバイアスの大きさは Green et al. (2003, p. 189) などに例示されている。

1章に例示したような、あるイベントに対して複数の(重複しない)原因やタイプなどで分類し、1つの原因やイベントタイプが観測されれば他のそれらは観測できない場合のほか、あるイベントが注目するイベントの発現確率を変化させてしまうような場合も競合リスクイベントと呼ばれる(Gooley et al., 1999)。特に、あるイベントが注目するイベントの発現確率を変化させてしまうが、注目しているイベントが発現しても観察は継続することができ、終了イベント(terminal event)の発現により観察が終了する場合は、最近では準競合リスク(semi-competing risk)とも呼ばれる(Fine, et al., 2001; Katsahian et al., 2004)。これらの場合も注目するイベントの定義を工夫することにより、これまでに説明した競合リスクの枠組みで取り扱うことができる。これらの事例は以下のような場合に見られる。

(1) 白血病の治療としての骨髄移植後の治療の失敗までの時間を観察する場合、治療の失敗を白血病の再発又は死亡のうちいずれか先に起こるイベントと定義することが多い。再発した症例でも観測を続ければ死亡は観測可能なイベントであるが、死亡を「再発前の死亡」と定義することにより「再発」した症例においては「再発前の死亡」は観測できず、「再発前の死亡」症例においては「再発」は観測できない。つまり、「再発」は「再発前の死亡」の競合リスク要因であり、逆もまた競合リスク要因である。もし、ある再発予防方法についての効果を試験したい場合は「再発」により高い関心を持つかもしれない。

(2) 冠動脈狭窄のある治療を行った後の予後の検討において、死亡、冠動脈バイパス術(CABG)、経皮的冠動脈インターベンション(PCI)までの時間を見る場合、CABGを受ければそのあとにPCIを受ける可能性はCABGを受けなかった場合よりも低下する。また、死亡が先に起こればCABG、またはPCIまでの時間は観測できない。CABGによるリスクの変化を受けていない状況でのPCIまでの時間について高い関心がある場合、イベントを上記(1)のように、「(死亡や)CABG前に受けたPCI」などといずれか先に起こるイベントの定義を工夫すれば、死亡・CABG・PCIについて互いに競合リスク要因になるような関係に導くことができ、それらの発現リスクは他のイベントによるリスクの変化を受けていない。

(3) 臨床試験では、薬剤(一般的には治療方法)の安全性を有害事象(AE)の発現状況により評価す

る．被験者によっては同じ AE が何度も繰り返し発現したり，様々な種類の AE が発現したり，あるいは何の AE も発現しなかったりする．なんらかの理由により，計画された試験終了時点の前に薬剤の使用を中止することになった場合，中止の後には AE は起こらない(厳密には遅発性 AE が起こることはあるが)．AE を許容できなくなりそれ以上発現するのを回避したいために薬剤の使用を中止することもある．関心を持つイベントは AE の発現であり，薬剤の使用中止は競合リスク要因である．AE が発現した被験者でもそれが許容できれば薬剤の使用を継続することも多い．しかし，その後，何らかの理由で薬剤の使用を中止することもある．そこで，ある有害事象 AE1 の発現に関心を持つ場合，イベントを「初発 AE1 の発現」又は「初発 AE1 の発現前の中止」と定義することにより，実際にはいずれか先に起こるイベントのみ観測できることになる．

競合リスクが存在する下でのイベント発現までの時間の解析法では，競合リスク要因は同時に発現しないこと(ゆえに競合しているわけだが)を条件としている．競合するイベント発現までの時間の測定誤差やまるめの誤差などでタイが起こる場合の取り扱いについては，イベントのカテゴリのグループ化などを行い競合リスク要因は同時に起こらないようにイベントのカテゴリの定義をなおしたり，小さな数を一様乱数をもちいて発生させ，それをタイデータから増加または減少させてタイをなくす方法(jittering)なども提案されている(Tai et al., 2002)．そのようにして，イベントの原因またはタイプが，重複しない m 個に分類できたと仮定する．競合リスクの間は一般には独立ではない．競合リスクが存在する下でのイベント発現までの時間の解析に関しては Kalbfleisch and Prentice (1980, 2002), Lawless (2003), Klein and Moeschberger (1997), Marubini and Valsecchi (2004), Pintilie (2006) に詳しい解説がされている．ただし，Pintilie (2006) は記号の定義に矛盾も散見されわかりにくい部分があるので注意が必要である．これらの文献はおもに治療の有効性評価への応用という観点から説明がなされている．安全性評価への応用という観点からの解説は西川 (2006) などがある．

4.1 イベント発現までの時間の分布と推定

前章までの結果と対比させやすいように次のように 1 章の記号の定義を少し変更し，いくつかの新しい記号を定義する． T を観測の起点から互いに競合するイベントのうち，いずれか最初のイベントが起きるまでの時間とする． $J \in \{1, \dots, m\}$ を競合リスク要因の原因又はタイプのカテゴリとする．競合リスク要因は m 個の重複しないいずれか 1 つのカテゴリにはいるものとする． C はこれまでと同様の定義で，無情報センサーであることを仮定する．

競合リスクモデルにおいて重要な概念は，時間 $T = t$ まで生存していたときに，次の瞬間に原因 j によるイベントまたはタイプ j のイベントが起こる率を意味する原因別ハザード (cause-specific hazard) 関数であり，次式で定義される (Kalbfleisch and Prentice, 1980, p. 167; 2002, p. 251)．タイプ別ハザード (type-specific hazard, Kalbfleisch and Prentice, 2002, p. 251)，機序別ハザード (mode-specific hazard, Lawless, 2003) とも呼ばれる．ここでは原因別ハザードと呼ぶことにする．

$$\lambda_j(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, J = j \mid T \geq t)}{\Delta t}, \quad j = 1, \dots, m.$$

イベントの原因を区別しない全ハザード (overall hazard)，原因別の累積ハザード，全累積ハザード

ト (overall cumulative hazard) がそれぞれ以下のように定義される.

$$\begin{aligned}\lambda(t) &= \sum_j^m \lambda_j(t), \\ \Lambda_j(t) &= \int_0^t \lambda_j(s) ds, \\ \Lambda(t) &= \sum_j^m \Lambda_j(t).\end{aligned}$$

原因 j のイベントが時間 t までに発現する確率を意味する累積発生関数 (Cumulative incidence function, CIF) は次式で定義される (Kalbfleisch and Prentice, 1980).

$$I_j(t) = P(T < t, J = j) = \int_0^t \lambda_j(u) S(u) du, \quad j = 1, \dots, m,$$

ここに

$$S(t) = \exp \left\{ - \int_0^t \left(\sum_{j=1}^m \lambda_j(u) \right) du \right\}$$

であり, 時間 t までにいずれのイベントも発現していない被験者の割合である. 関数としては $S(t)$ は一般的な意味で全生存率 (全生存割合, overall survival) と呼ばれる. 4 章の例題 (3) では「初発 AE1 の発現」又は「初発 AE1 の発現前の中止」のいずれのイベントも発現していない被験者の時間 t における割合に相当する. 4 章の例題 (1) のようにイベントの定義を「再発」または「再発前の死亡」とするような物理的な死亡がイベントに含まれている場合, $S(t)$ は定義されたいずれのイベントも時間 t までに発現していない被験者の割合ということであるから, 医学的な $S(t)$ の意味は無イベント生存率 (event-free-survival) または無病生存率 (disease-free-survival) である. 例題 (1) のような場合には, 医学的な意味での全生存率は「再発」の有無に関係なく, 時間 t までに生存している被験者の割合を意味するので, これらを混同しないように注意が必要である. $I_j(t)$ は部分分布 (subdistribution) と呼ばれることもある. これは競合リスクの発生により $I_j(\infty) < 1$ となり, 厳密には分布関数にはならないことに由来する. 次の関係が成立する.

$$1 - S(t) = \sum_j^m I_j(t),$$

右辺は全 (事象) 累積発生関数 (overall incidence function) に相当する. そのほか, 競合リスクが存在しないときと類似の関数である原因 j の部分密度 (subdensity) と原因 j の部分分布のハザード (subdistribution hazard) がそれぞれ以下のように定義できるが, 競合リスクが存在しないときと解釈が異なってくる.

$$\begin{aligned}f_j(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, J = j)}{\Delta t} \\ &= \lambda_j(t) S(t) = \lambda_j(t) \exp(-\Lambda(t)), \\ \gamma_j(t) &= \frac{-d(\log(1 - I_j(t)))}{dt}.\end{aligned}$$

実際には、被験者 i ($i = 1, \dots, n$) について観測されるデータは (u_i, δ_i, j_i) の組となっている。ここに、 $u_i = \min(t_i, c_i)$, $\delta_i = I(t_i \leq c_i)$, $j_i \in \{1, \dots, m\}$ である。ただし、 $\delta_i = 0$ のときは j_i は定まらない。競合リスク要因相互の間は独立である必要はない。ここで、 b_{ji} を被験者 i の原因 j によるイベントについての観察打ち切りの表示子とする。この表示子に関する観察打ち切りは競合リスクのために原因 j のイベントが観測されない場合を含むので、無情報センサーとは限らない。

$$b_{ji} = \begin{cases} 1 & \text{if } j_i = j \text{ and } \delta_i = 1 \\ 0 & \text{otherwise.} \end{cases}$$

観測された死亡までの時間 u_i を昇順に並び替えて添え字をつけ、 $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(n)}$ とし、これらの時間に対応した被験者の観察打ち切りの表示子も並び替えて同様に $\delta_{(1)}, \delta_{(2)}, \dots, \delta_{(n)}$ と添え字をつけ直す。観察打ち切りとイベントが同じ時間である場合、3.1 章と同様にイベント発現を先として取り扱う。 $u_{(i)}$ に対応した被験者の、原因 j によるイベントの観察打ち切りの表示子も同様に $b_{j(i)}$ と添え字をつけ直す。 c_i は無情報センサーであることを仮定しているので、尤度関数は以下の式に比例する。

$$\begin{aligned} L &= \prod_{i=1}^n \prod_{j=1}^m [\lambda_j(u_i)]^{b_{ji}} S(u_i) \\ &= \prod_{i=1}^n \prod_{j=1}^m [\lambda_j(u_i)]^{b_{ji}} \prod_{j=1}^m \exp \left[- \int_0^{u_i} \lambda_j(x) dx \right] \\ &= \prod_{j=1}^m L_j \end{aligned}$$

ここに、 $L_j = \prod_{i=1}^n [\lambda_j(u_i)]^{b_{ji}} \exp(-\Lambda_j(u_i))$ である。

もし、競合リスク要因が独立であるならば、 L を最大化するには個別に L_j を最大化して $\lambda_j(t)$ の推測を行うことができるが、競合リスク要因は一般には独立ではない。イベントタイプ j_1 がイベントタイプ j_2 ($j_1 \neq j_2$) のパラメーターに関しての情報を持つ場合、 $\lambda_{j_1}(t)$ の推測にイベントタイプ j_2 の情報を利用しないことは推測の効率の損失を招く。 L_j は形式の上では j 以外のタイプのイベントを無情報センサー扱いをして尤度を構成したものと同様になっている。ただし、後述するが、 $\exp(-\Lambda_j(u_i))$ は一般には確率として解釈することが困難である。部分尤度 (Cox, 1975) は推測したいパラメーターに関する情報から構成されているが、推測したいパラメーターに関してのすべての情報から構成されているわけではない。それと同様に、 L_j はタイプ j のイベントに関する主な情報から構成されているが、タイプ j のイベントに関するすべての情報から構成されているわけではない。競合リスク要因にあるイベントタイプ j_1 を発現すれば、そのほかのイベントタイプを発現できない(もし観察を継続できたとすれば、という仮定は不成立)という点は、無情報センサーにより観察打ち切りになった(もし観察を継続できたとすればそのほかのイベントの将来の発現可能性は現在観察を継続している被験者と同じ)場合とは異なる。しかし、競合リスク要因にあるイベントタイプ j_1 を発現すれば、その時点以降にタイプ j_1 以外のタイプのイベントに関する情報を与えることはできない。そういう意味で、競合リスク要因が独立ではない場合

は, L_j を $\lambda_j(t)$ の推測に関しての部分尤度と解釈できるので, 原因別ハザード関数の推測については 3 章の理論が同様に適用できる. 詳しい説明は Kalbfleisch and Prentice (1980) などを参照して欲しい. 原因 j の CIF は次式で推定される.

$$\hat{I}_j(t) = \sum_{i: u_{(i)} \leq t} \frac{b_{j(i)}}{N_i} \hat{S}(u_{(i)}-), \quad j = 1, \dots, m.$$

3.1 章と類似の記号を使えば次のように表現される.

$$\hat{I}_j(t) = \sum_{i: t_{ji} \leq t} \frac{d_{ji}}{N_{ji}} \hat{S}(t_{ji}-), \quad j = 1, \dots, m.$$

ここに, t_{ji} は原因 j の i 番目のイベントが発現した時間, d_{ji} は時間 t_{ji} で原因 j のイベントを発現した被験者数, N_{ji} は t_{ji} の直前までいずれのイベントも発現せず, 観察を継続している被験者数(リスク集合の大きさ), $\hat{S}(t-)$ は時間 t の直前までにいずれのイベントも発現していない確率の推定値で, KME による左連続の生存時間関数推定値に相当する. $\hat{S}(t)$ を推定する際, $\delta_{(i)} = 0$ の被験者の $u_{(i)}$ が観察打ち切り(無情報センサー)として取り扱われる. 観察打ち切りがイベント発現までの時間とは独立であることは $\hat{S}(t)$ が不偏であるために必要な条件である. 競合リスク要因相互の間は独立である必要はない.

図 1 に, イベントをある疾患の再発(原因 1)または再発前の死亡(原因 2)とし, 治療方法の割り付け時点を起点とした累積発生率の推定値を示した(仮想データ). 4 章の例題(1)で説明したように, これらは互いに競合リスク要因である. それぞれ, 実線は全(事象)累積発生関数($1 - S(t)$), 太線は「再発」の CIF($I_1(t)$), 実線と太線の差は「再発前の死亡」の CIF($I_2(t)$), また, $1 - \text{実線}$ (probability = 1 と実線の差) は $S(t)$ の推定値を示す.

図 1. 経時的累積再生率と累積治療失敗率. 治療失敗は再発または再発前の死亡と定義する.

特別な場合として、計画された観察期間は全員に対して t_0 で固定である場合 (タイプ 1 センサーの 1 つ, 例えば, $t_0 = 24$ 週) を考える. そのとき次式が成立する.

$$\hat{I}_j(t) = \sum_{i: t_{ji} \leq t} \frac{d_{ji}}{n} \quad \text{for } t < t_0, \quad j = 1, \dots, m.$$

ここに, n は試験に組みこまれた全被験者数である. すなわち, 原因 j の CIF estimator (CIFE) は, 時間 t までに原因 j のイベントを発現した被験者の割合と等しくなる. 原因 j のいずれかのイベントを発現する前に試験の中止や脱落など無情報センサーといえない観察打ち切りが生じた場合, Nishikawa et al. (2006) はこれらを 1 つの競合リスク要因として取り扱い, 注目するイベントの CIFE とともに競合リスクの CIFE も報告することを薦めている.

$\hat{I}_j(t)$ の標準誤差 (SE) は Dinse and Larson (1986) の式など幾つかの方法により推定できるが, 一般に計算式は複雑になる. Dinse and Larson (1986) 式を以下に示す. $t_{(i-1)} \leq t < t_{(i)}$ である t に対して,

$$\begin{aligned} \text{var}(\hat{I}_j(t)) = & \sum_{r=1}^i \hat{\theta}_{jr}^2 \left\{ \frac{N_r - b_{j(r)}}{b_{j(r)} N_r} + \sum_{l=1}^{r-1} \frac{\delta_{(l)}}{N_l(N_l - \delta_{(l)})} \right\} \\ & + 2 \sum_{r=1}^{i-1} \sum_{r'=r+1}^i \hat{\theta}_{jr} \hat{\theta}_{jr'} \left\{ -\frac{1}{N_r} + \sum_{l=1}^{r-1} \frac{\delta_{(l)}}{N_l(N_l - \delta_{(l)})} \right\}, \end{aligned}$$

ここに,

$$\hat{\theta}_{ji} = \frac{b_{j(i)}}{N_i} \hat{S}(t_{(i)}).$$

もしイベントの原因を区別しない場合は上の式は Greenwood 式に帰着する. Choudhury (2002) は SE の推定方法 (Dinse and Larson 法, $\log(-\log)$ 変換, arcsine 変換) について simulation により比較 ($N = 100, 200, 500$) を行い, Dinse and Larson 法がよかったと報告している. Braun and Yuan (2007) は多項分布に基づいて導出された SE の推定方法 (Betensky and Schoenfeld, 2001 など), Martingale 理論 (e.g. Fleming and Harrington, 1991 など) に基づいて導出された SE の推定方法 (Lin, 1997 など), bootstrap SE (Efron, 1981) を simulation により比較し ($N = 20, 50, 100$), 多項分布に基づいて導出された SE の推定は小標本でも比較的性能が良いことを報告している. そのほかの SE の推定方法については Choudhury (2002), Marubini and Valsecchi (2004), Pintilie (2006) などを参照して欲しい. 各時点 t ごとの $I_j(t)$ の信頼係数 $100(1 - \alpha)\%$ の信頼区間は $\hat{I}_j(t)$ の漸近正規性により次のように近似できる.

$$\hat{I}_j(t) \pm \Phi_{\alpha/2} \sqrt{\text{Var}(\hat{I}_j(t))}$$

信頼区間が必ず 0 以上にはいるようにするために, $\log(-\log(1 - \hat{I}_j(t)))$ の漸近分散をもとに $I_j(t)$ の信頼区間を次式で求めることもある.

$$\hat{I}_j(t) \exp \left[\pm \Phi_{\alpha/2} \sqrt{\text{var}(\hat{I}_j(t)) / [(\log \hat{I}_j(t)) \hat{I}_j(t)]^2} \right]$$

Tai et al. (2001) には CIFE 計算などの SAS のコーディングが, Choudhury (2002) には CIFE 及び CIFE の SE 計算などの s-plus のコーディングが公開されている.

イベント発現までの時間の分布の見方として、そのほかに Pepe and Mori (1993) により提唱された条件付確率 $CP_j(t)$ がある。 $CP_j(t)$ は、競合リスクが存在するときに、原因 j の競合リスクである原因 j 以外のイベントを起こしていないという条件で、原因 j が時間 t までに発現する確率となる。別な言い方をすれば、 $CP_j(t)$ は、競合リスクであるイベントを起こさず観察を継続できている個体が原因 j のイベントを時間 t までに発現する確率を与える。次式で定義できる。

$$\begin{aligned} CP_j(t) &= P(T < t, J = j) / (1 - P(T < t, J \neq j)) \\ &= I_j(t) / (1 - \sum_{k=1, \dots, m, k \neq j} I_k(t)) \end{aligned}$$

$CP_j(t)$ の推定値は $\hat{I}_j(t) / (1 - \sum_{k=1, \dots, m, k \neq j} \hat{I}_k(t))$ により求められる。分散は Pepe and Mori (1993) を参照して欲しい。

関節リウマチの比較試験では被験治療間で治療の継続率(または中止率)の違いが大きいものも散見される(e.g. Lipsky et al., 2000; Klareskog et al., 2004)。そのような場合には単純な比率による有害事象の発現率の比較は不適切であり、注目する有害事象 AE1 の CIFE と AE1 発現前の中止の CIFE のみならず、AE1 の「条件付確率」を比較するのも良いであろう。

4.2 準競合リスクにおける推測

注目しているイベントが発現しても観察は継続することができ、終了イベントの発現により観察が終了する場合がある。例えば4章の競合リスクの例(1)では白血病を再発($J=1$)しても観察は継続することができ、死亡(終了イベントの発現, $J=2$)により観察が終了する場合がこれにあたる。競合リスクの例(3)では、有害事象($J=1$)が起こっても薬剤の使用を中止する(終了イベント, $J=2$)まで観察は継続することができる。治療効果の推定や比較に、このような、症状が相対的に軽い非致死性のイベント(以降、軽症イベントと呼ぶ)を経験した後にどれくらい経過すればどの程度の頻度で致死性の終了イベント(以降、終了イベントと呼ぶ)が起こるのか、または、軽症イベントを経験した後に終了のイベント発現を遅らせるかを見ることも有用であろう。終了イベントは注目している軽症イベントよりも先に発現することもありえる。終了イベントは軽症イベントの観測を妨げるが、軽症イベントは終了イベントの観測を妨げない。ここでは上記の例(1), (3)のように競合リスク要因は2つとし、軽症イベントを $J=1$, 終了イベントを $J=2$ とする。記号の定義を次のように追加する。 T_D 及び X を、観測の起点からそれぞれ終了イベント発現までの時間、及び軽症イベント発現までの時間とする。時間 t までに、軽症イベントのあとに終了イベントを発現する確率は次式で定義できる。これを軽症イベント後に終了イベントを発現する累積同時発生関数(Cumulative joint incidence function, CJIF)と呼ぶ(Nishikawa et al., 2006)。ここでの同時とは両方ともにの意味である。

$$I_D^{1+}(t) = P(X < T_D < t).$$

$t_{Di}, i=1, \dots, n$, を被験者 i の終了イベント発現までの時間とし、 $t_{Di}, i=1, \dots, n$, は独立に同じ分布関数 $F_D(t) = P(T_D < t)$ に従うと仮定する。このとき、生存時間関数は $1 - F_D(t)$ である。ここでのイベントである終了イベントについては競合リスクは存在しない。 $U_D = \min(T_D, C)$ を

観測の原点から、終了イベントまたは観察打ち切りのうち先に観測された時間とする。被験者 i の観察打ち切りまでの時間 $c_i, i = 1, \dots, n$, はこれまでのように、イベント発現までの時間とは独立(無情報センサー)であることを仮定する。被験者 $i, i = 1, \dots, n$, について観測されるデータは (u_{Di}, δ_{Di}) のペアとなっている。ここに、 $u_{Di} = \min(t_{Di}, c_i)$, $\delta_{Di} = I(t_{Di} \leq c_i)$ である。 $T \leq T_D$ 及び X, T_D は連続値であるから $P(X = T_D) = 0$ が成り立ち、

$$I_D^{1+}(t) = F_D(t) - I_m(t)$$

が導かれる。それぞれの推定量を代入して以下のように CJIFE (CJIF estimator) を求めることができる。

$$\widehat{I}_D^{1+}(t) = \widehat{F}_D(t) - \widehat{I}_m(t)$$

ここに、 $\widehat{F}_D(t)$ は $(u_{Di}, \delta_{Di}), i = 1, \dots, n$, を用いて $1 - F_D(t)$ を KME により推定し、1 からそれを減じることにより得られる。観察打ち切りデータがない状況では $\widehat{I}_D^{1+}(t)$ は時間 t までに軽症イベントを発現した後に終了イベントを発現した被験者の、試験に組み込まれた全被験者に対する割合と等しくなる。 $\widehat{I}_D^{1+}(t)$ は一致推定量であるが、有限の標本サイズで観察打ち切りデータが存在する場合は必ずしも単調非減小関数とならないことがある。多変量の生存時間分布の推定量はこのような直感に反する性質は珍しくない (Lin et al., 1999; Wang and Wells, 1998)。そこで $\widehat{I}_{D*}^{1+}(t) = \inf_{s \geq t} \{\widehat{I}_D^{1+}(s)\}$ と定義しなおす。このとき $\widehat{I}_{D*}^{1+}(t)$ は一致推定量であり単調非減小関数となる。軽症イベント後に終了イベントを発現する CJIFE の分散は Pepe (1991) の方法やブートストラップ法 (Efron, 1981) により推定できる。

4.3 Kaplan-Meier 推定量を CIF の推定に用いることについて

競合リスクであるイベントが発現し、注目しているイベントの観察が継続できずに観察が打ち切りになることは多くの場合無情報センサーではない。それにもかかわらず無情報センサー扱いをして(競合リスクであるイベントを発現した個体のそれ以降の将来のイベントがもし観察できると仮定してそのイベントの発現可能性は、今観察打ち切りを受けていない個体の将来のイベントの発現可能性と変わらないことを仮定し)、KME を用いて注目するイベントの累積発生率を推定すればバイアスがあることは良く知られている。KME はイベントを発現していない確率を推定するので、累積発生率は $1 - KME$ により推定される。過去には累積発生確率を $1 - KME$ により推定している文献が数多く見られる。 $1 - KME$ により推定している場合に net estimate, CIFE により推定する方法を crude estimate と呼び名を正確に使い分けている文献もあるが、これらの呼び名は文献によってはそのほかの用語を用いている場合もある (Farley et al., 2001)。1 章で KME を伝統的な表現で示したが、Gooley et al. (1999) は CIFE との違いがわかりやすい KME の表現を提唱した。以下にそれを概説する。話を簡単にするために互いに競合するイベントは 2 つとし、原因 1 のイベントを注目するイベントであると仮定する。以下の定義を追加する。

a_i : 時点 $t_{1(i-1)}$ と時点 t_{1i} の間に観察打ち切りされた(無情報センサー)数。

r_i : 時点 $t_{1(i-1)}$ と時点 t_{1i} の間に発現した原因 2 のイベント数。

$J_{CI}(t_{1i})$: 時点 t_{1i} に発現した原因 1 のイベントによる CIFE の増加量。

$J_{KM}(t_{1i})$: 時点 t_{1i} に発現した原因 1 のイベントによる $1 - KME$ の増加量。

すでに定義した d_{1i} は時点 t_{1i} に発現した原因 1 のイベント数を示している。もし、全ての被験者について原因 1 のイベントを発現する確率が等しいと仮定すれば、観察打ち切りがなく原因 1 のイベントを完全に観測できる限りは、一例の個体において原因 1 のイベントが発現するたびに $1/n$ だけ CIFE は増大する。ここで、時点 $t_{1(i-1)}$ と時点 t_{1i} の間に観察打ち切りを受けた個体があったとすれば、その個体の CIFE に対する増加分はまだ観察打ち切りを受けずに観測を継続している個体に等分される(右側への再分配, redistributing to the right (Efron, 1967))。したがって、

$$\begin{aligned} J_{CI}(t_{1i}) &= J_{CI}(t_{1(i-1)}) + J_{CI}(t_{1(i-1)}) \left(\frac{a_i}{N_{1i}} \right) \\ &= J_{CI}(t_{1(i-1)}) \left(1 + \frac{a_i}{N_{1i}} \right), \quad i = 2, \dots, n. \end{aligned}$$

ただし、 $J_{CI}(t_{1(1)}) = 1/n$ である。時点 $t_{1(i-1)}$ と時点 t_{1i} の間に競合リスク要因である原因 2 のイベントが発現したとしてもこれらの個体の $J_{CI}(t_{1i})$ に対する増加の寄与分は 0 である。一方、時点 $t_{1(i-1)}$ と時点 t_{1i} の間に発現した原因 2 のイベントを無情報センサー扱いをして $1 - KME$ により原因 1 のイベントの発現率を推定する場合、上記の右側への再分配により $J_{KM}(t_{1i})$ は以下のような形で表現できる。

$$\begin{aligned} J_{KM}(t_{1i}) &= J_{KM}(t_{1(i-1)}) + J_{KM}(t_{1(i-1)}) \left(\frac{a_i + r_i}{N_{1i}} \right) \\ &= J_{KM}(t_{1(i-1)}) \left(1 + \frac{a_i + r_i}{N_{1i}} \right), \quad i = 2, \dots, n. \end{aligned}$$

ただし、 $J_{KM}(t_{1(1)}) = 1/n$ である。上式は、原因 2 のイベントが発現した個体をそのあとも原因 1 のイベントを発現する可能性がある個体として原因 1 のイベントの発現率を推定している。 $J_{CI}(t_{1i})$ と $J_{KM}(t_{1i})$ は原因 2 のイベントが発現しない限りは一致している。原因 2 のイベントが発現したあと最初に原因 1 のイベントが発現した時点から $J_{KM}(t_{1i}) > J_{CI}(t_{1i})$ となる。したがって、原因 2 のイベントが沢山発現するほど $1 - KME$ と原因 1 の CIFE の差は増大する。原因 2 のイベントが発現しない間は $1 - KME$ と原因 1 の CIFE は一致している。なお、いくつかの文献には、解釈には問題があると書いた上で $1 - KME$ を次のような関数で示してある場合もある。

$$G_j(t) = \exp \left(- \int_0^t \lambda_j(x) dx \right) = \exp(-\Lambda_j(t))$$

$G_j(t)$ のかわりに $S_j(t)$ という記号で定義している文献も多いので、4.5.2 章の $S_j(t)$ と混同しないように注意が必要である。以下の関係が成り立つ。

$$I_j(t) \leq 1 - G_j(t) \leq 1 - S(t)$$

以上のように競合リスク要因が存在する場合にそれを無情報センサー扱いをすることには 2 つの問題がある。1 つは、無情報でないのを無情報センサーとして扱うことによる推定値への偏り。2 つ目は、こちらのほうがあまり知られていないようであるが、解釈において重要な点である。 $1 - KME$ による推定値を、もし競合リスク要因を排除できると仮定した場合の仮想的な想定での注目するイベントの発現確率とみなすこともありえるが、実際そういう状態にするには競合リス

ク要因を排除できるような治療や処置を行う必要がある。そしてそれがたとえできたとして、それでもなお注目するイベントの発現可能性には何の影響もあたえずに競合リスク要因を排除できるであろうか、という点である。Kalbfleisch and Prentice (1980, 2002) はこの点について真摯な議論を展開しているので参照して欲しい。1-KME は現実には存在しないような仮想的な集団を対象とする推測になっている (e.g. Gooley et al., 1999; Pepe and Mori, 1993)。Pepe and Mori (1993) にもわかりやすい議論がなされている。ただし、1-KME による推定値を確率として解釈するのではなく、原因別ハザードの違いを視覚的に表現する方法として見れば非常に有用かもしれない (Tai et al., 2001)。

治療域や試験の目的によって 1-KME と CIFE の使い分けを薦めている文献 (Chappell, 1996; Farley, et al., 2001 など) もあるが、1-KME については、データから確認できない競合リスク要因の間の独立性や競合リスク要因を排除できるような治療や処置を行ったあとのリスクの構造不変性を仮定しているため、1-KME については確率として解釈せずに Tai et al. (2001) のような解釈をするほうが有用であろう。これまで累積発生確率を 1-KME により推定している文献が非常に多かったのは、KME を備えているソフトウェアは豊富に存在するが、CIFE を備えているソフトウェアはそんなに豊富ではなかったことにも原因があるかもしれない。

4.4 多変量生存時間データとみなす方法について

かつては競合リスク要因の解析には潜在的な多変量生存時間データとみなす方法が用いられていたが、識別不能性 (nonidentifiability) の問題があり現在はこの方法はあまり用いられていない。簡単に紹介しておく。それぞれのイベントタイプごとに潜在的生存時間 $\tilde{T}_j (j = 1, \dots, m)$ があると仮定する。実際に観測されるデータは $\min(\tilde{T}_1, \dots, \tilde{T}_m)$ である。ここで $\tilde{T}_1, \dots, \tilde{T}_m$ の多変量生存時間関数 $P(\tilde{T}_1 > t_1, \dots, \tilde{T}_m > t_m)$ を考える。 j 番目の変数についての周辺生存時間関数は $S^j(t) = P(\tilde{T}_j > t) = P(0, \dots, 0, t, 0, \dots, 0)$ となる。このとき、観測されるデータからは多変量生存時間関数及び周辺生存時間関数は一意に定まらない。なぜなら、同一人において一変数のみ観測可能な場合は、どのような相関(従属)構造をもつ多変量生存時間関数についても観測されるデータから推測できる原因別ハザード関数と同一の原因別ハザード関数を持つような独立な多変量生存時間関数が存在するからである (e.g. Tsiatis, 1975)。詳細は Tsiatis (1975) や Kalbfleisch and Prentice (1980), Crowder (2001) などを参照して欲しい。

4.5 検定について

ここでは原因別ハザード関数を比較する方法、及び CIF を比較する方法について述べる。条件付確率の検定については Pepe and Mori (1993) を、累積同時発生関数の検定については Nishikawa et al. (2006) などを参照して欲しい。4.1 章で示したように原因別ハザード関数に関する推測については 3 章の理論が流用できる。2 群を併合して原因 j のイベント発現までの時間を昇順に並べたときの i 番目のイベント発現が観測された時点 $t_{j(i)}$ の直前の被験群 (T) 及び対照群 (A) における例数をそれぞれ n_{Tji} , n_{Aji} , 合計 N_{ji} 例, 時点 $t_{j(i)}$ での各群における原因 j のイベント数をそれぞれ d_{Tji} , d_{Aji} , 合計 d_{ji} 例とする ($i = 1, 2, \dots, q$)。

4.5.1 2 群間での原因 j のハザード関数の差の検定

原因 j のハザードについてのログランク検定を行う。被験群及び対照群における原因 j のハザード関数をそれぞれ $\lambda_{Tj}(t)$, $\lambda_{Aj}(t)$ とする。帰無仮説は $\lambda_{Tj}(t) = \lambda_{Aj}(t)$ である。原因別ハザード関数についての検定は、通常のログランク検定が流用できる。ログランク検定での分割表と同様にいずれかの群で原因 j のイベントが観測された時点ごとに、群と原因 j のイベントの状態(有・無)の分割表を作成する(表 2)。

表 2. 原因 j のイベントが観測された時点 $t_{j(i)}$ での分割表

群	原因 j の イベント発現数	原因 j の イベント未発現数	合計
被験群	d_{Tji}	$n_{Tji} - d_{Tji}$	n_{Tji}
対照群	d_{Aji}	$n_{Aji} - d_{Aji}$	n_{Aji}
リスク集合全体	d_{ji}	$N_{ji} - d_{ji}$	N_{ji}

これ以降の議論は、ログランク検定について各群の生存時間関数が同じという帰無仮説の下で展開される議論(3.2 章)と同様である。2 群の原因別ハザード関数についての層別ログランク検定も上記と同様にして 3.2 章の層別ログランク検定が流用できる。

4.5.2 2 群間での原因 j の CIF の差の検定

Gray (1988) は k 群の CIF の検定方法を提案した。ここでは 2 群比較について解説する。群 (T, A) の区別をするために各関数に上付で T, A を付与する。帰無仮説: $I_1^T(t) = I_1^A(t)$ の下でも $S^T(t) = S^A(t)$ や $\Lambda_1^T(t) = \Lambda_1^A(t)$ が成立するとは限らない。Gray 検定では部分分布のハザードの重み付き平均で検定統計量を構成する。重みの付け方により帰無仮説の正確な表現は変わってくる。この重みを変えることにより Gray 検定での直接の帰無仮説は $I_1^T(t) = I_1^A(t)$, $\gamma_1^T(t) = \gamma_1^A(t)$ または $I_1^T(t)/(1 - I_1^T(t)) = I_1^A(t)/(1 - I_1^A(t))$ と表現される。ここではログランク検定に対応する帰無仮説: $\gamma_1^T(t)/\gamma_1^A(t) = 1$ の場合を見る。群 A について以下の関数を定義する。群 T についての定義も同様である。

$$S_j^A(t) = 1 - I_j^A(t) = \exp\left(-\int_0^t \gamma_j^A(u)du\right) \quad (j = 1, \dots, m).$$

4.1 章で言及したように、 $I_j^A(t)$ は厳密な意味で分布関数になっていないが、以下のような T^* を考えることにより、 $I_1^A(t)$ は $T^* < \infty$ に対する T^* の分布関数になっているとみなせる。

$$T^* = \begin{cases} T & \text{if } J = 1 \\ \infty & \text{if } J \neq 1. \end{cases}$$

このとき T^* の分布関数は $T^* < \infty$ では $I_1^A(t)$, $T^* = \infty$ では 1 である。これより、以下の関係が成り立つ。

$$\begin{aligned} P^A(T^* > t) &= P^A(T > t) + I_2^A(t) + \dots + I_m^A(t) \\ &= 1 - I_1^A(t) \end{aligned}$$

したがって、部分分布のハザードに関するリスク集合はその時点でリスクに曝されている個体及び競合リスクを経験したことにより原因 1 のリスクから開放された個体から構成される。後者は原因 1 を経験できない個体に該当する。表 2 の分割表を作成する。原因 1 のイベントが観測された時間 t_{1i} のイベントタイプの添え字 1 はこの章の以降の式では省略する。関数の定義により

$$S_1^A(t_{i-1})/S^A(t_{i-1}) \geq 1$$

が成り立つ。B 群についても同様のことが成り立つので、それぞれの群での Gray の検定におけるリスク集合の大きさは n_{Tji} , n_{Aji} より一般に大きくなる。以下の式で t_i における調整されたりリスク集合の大きさを推定する。

$$\hat{R}_1(t_i) = n_{Aji} \frac{\hat{S}_1^A(t_{i-1})}{\hat{S}^A(t_{i-1})} + n_{Tji} \frac{\hat{S}_1^T(t_{i-1})}{\hat{S}^T(t_{i-1})}$$

帰無仮説の下では $\hat{\gamma}_1(t_i) = 1/\hat{R}_1(t_i)$ となり、表 2 の周辺で条件付けられた被験群の原因 1 のイベント数の期待値は

$$d_{ji} n_{Tji} \frac{\hat{S}_1^T(t_{i-1})}{\hat{S}^T(t_{i-1})} / \hat{R}_1(t_i)$$

となる。Gray の検定の検定統計量の分子における調整されたりリスク集合の大きさの数値例を表 3 に示す。表 3 において、原因 j のイベント発現数、原因 j のイベント未発現数、 $I_1^\bullet(t_{j(i-1)})$ 及び $S^\bullet(t_{j(i-1)})$ を所与の数値とし、そのほかのカラムはこれらの数値を用いて算出した。調整したりリスク集合の計算過程を脚注に示した。分散の計算式は複雑になるのでここでは省略する。詳細は Gray (1988) や Pintilie (2006) を参照して欲しい。

表 3. 時点 $t_{j(i)}$ での Gray 検定における調整リスク集合の計算例

群	原因 j のイベント発現数	原因 j のイベント未発現数	調整しないリスク集合	$I_1^\bullet(t_{j(i-1)})$	$S_1^\bullet(t_{j(i-1)})$	$S^\bullet(t_{j(i-1)})$	調整したリスク集合
被験群	2	24	26	1/35	34/35	26/35	34 ^a
対照群	0	27	27	2/35	33/35	27/35	33 ^b
全体	2	51	53	-	-	-	67

$$a: \frac{34/35}{26/35} 26 = (34/26)26$$

$$b: \frac{33/35}{27/35} 27 = (33/27)27$$

4.5.3 検定方法の選択について

競合リスクが存在する場合、2 群の比較に原因 j のハザードについてのログランク検定及び CIF についての Gray の検定のほかに全生存率 $S(t)$ を比較することが考えられる。話を簡単にするために、互いに競合するイベントは 2 つ(原因 1, 原因 2)とし、互いに競合する関係にあるいずれのイベントも発現を遅らせる処置の方が良いという場合を考えてみる。Marubini and Valsecchi (2004) は次のような検討を行っている。

被験治療(T)は対照治療(A)より原因 1 のイベント発現を遅らせることを仮定するが、原因 2 のイベント発現については次の 3 通りの場合がある。

- (1) 原因 2 のイベント発現については対照治療と同じ(不変)である。
- (2) 原因 2 のイベント発現も遅らせる。
- (3) 原因 2 のイベント発現については対照治療よりも早める(悪化)。

これらの場合に、原因 j のハザード関数(Marubini and Valsecchi (2004) で直接扱っているのは G_j であるが、検定方法との対応からここでは原因 j のハザード関数を用いて表現する.)、原因 j の CIF、全生存率の被験治療と対照治療との大小関係は次のようになる。導出は省略するので原文を参照して欲しい。

- (1) $I_1^T(t) < I_1^A(t)$, $S^T(t) > S^A(t)$ が成立. $\lambda_1^T(t)$ と $\lambda_1^A(t)$ の大小は一般には不明。
- (2) $S^T(t) > S^A(t)$ が成立. $I_1^T(t)$ と $I_1^A(t)$ の大小, 及び $\lambda_1^T(t)$ と $\lambda_1^A(t)$ の大小は一般には不明。
- (3) $I_1^T(t) < I_1^A(t)$ が成立. $S^T(t)$ と $S^A(t)$ の大小, 及び $\lambda_1^T(t)$ と $\lambda_1^A(t)$ の大小は一般には不明。

「一般には不明」というのはすべての時点でその関係が成立するとは限らないという意味である。何を検定すればいいか、という問題は治療のどのような側面を主な評価・要約とするかという問題に対応する。試験の計画の時点で何を推測して比較したいのか、発現確率であるのか原因別ハザードであるのかなど試験の目的が明確であれば、それを比較することが主な解析で、それ以外は副次的な解析となるであろう。

Marubini and Valsecchi (2004) は原因 j のハザード関数は原因 j のイベントの発現確率とは直接結びついていないのに対して、全生存率と CIF はイベントの発現確率と直接結びついているのでこれらの比較を薦めている。原因 j のハザード関数の違いは生物学上の反応性の違いを反映していて、CIF は原因 j のイベントの発現確率と直接結びついている (e.g. Pintilie, 2006; Pepe and Mori, 1993)。Kalbfleisch and Prentice (1980) はイベントの発現確率と直接結びついている CIF を薦め、CIF について原因別累積ハザード関数は補足的な情報を持つということで薦めている。原因 j の CIF には競合するイベントの CIF からくる制約条件があるので、競合するイベントの発現確率が高くなっているために原因 j のイベントの発現確率が低いように見えることがある。原因 j の CIFE を適切に解釈するには競合するイベントの CIFE を一緒に見ておく必要がある。「条件付確率」は、競合するイベントの発現確率で調整したあとの原因 j の発現確率ともみなせる。Klein and Moeschberger (1997) は発現確率を要約する方法として、CIF または「条件付確率」を薦め、 G_j の解釈の問題点を指摘している。

Caplan et al. (1994) は進行性の骨盤内悪性腫瘍に対する緩和的放射線療法における後期の合併症の経時的発現率の要約方法として「条件付確率」及び累積ハザードの有用性を図示し、生存している患者において後期の合併症発現のリスクはかなり高いと解釈した。これについて Bentzen et al. (1995), Caplan et al. (1995), Denham, et al. (1996), Caplan et al. (1996), Chappell (1996) らにより議論がなされているが、一部誤解も見受けられる。発現確率を主要な評価とする場合に G_j を比較することが、Chappell (1996) は別として、データから確認できない競合リスク要因の間の独立性や競合リスク要因を排除できるような治療や処置を行ったあとのリスクの構造不変性を仮定しないといけないことが、当時の医学分野ではあまり認識されていなかったのかもしれ

ない。Klein and Moeschberger (1997) が薦めているように、「条件付確率」はもっと頻繁に利用されてよいであろう。

原因別ハザード関数は競合リスクの発現の影響を受けないという説も見受けられるが、それは正確な表現ではない。推定量としては競合リスクを経験するか否かは明示的に関係ないように見えるが、競合リスクの発現状況が変化すれば、全体としてのリスク集合(推定量の分母に相当)は変化するわけであり、変化したゆえにリスク集合に含まれるような個体はもともとリスク集合に含まれているような個体と同程度に原因 j のイベントが発現しやすいと仮定することは、競合リスク要因の間の独立性を仮定することに相当する。この仮定は一般に成り立たないことはこれまでに述べたとおりである。

原因別ハザード関数の検定は競合リスクであるイベントの累積発生率が群間で同程度であれば Gray の検定とほぼ同様の結果になる。そうではない場合に、Gray の検定と原因別ハザード関数の検定結果が異なる例は Gray (1988) や Pintilie (2006) などで紹介されている。Gray の検定の応用は Martelli et al. (2005) や Oka et al. (1995) 等に見られる。全生存率や原因 j のハザード関数の比較は汎用のログランク検定のソフトを利用して容易に行える。Gray の検定は R の library から利用できるようになった (cmprsk パッケージの cuminc 関数)。

4.6 共変量のモデル化

4.6.1 原因別ハザード関数のモデル化

共変量ベクトル $\mathbf{z}$ (p 次元) を持つ被験者の時間 t における原因 j のハザード関数 $\lambda_j(t, \mathbf{z})$ に対して比例ハザード性及び対数線形性を仮定したモデル化は以下の式で表現される。

$$\lambda_j(t, \mathbf{z}) = \lambda_{j0}(t) \exp(\beta_j' \mathbf{z})$$

ここに、 $\lambda_{j0}(t)$ は原因 j のハザードについての基準ハザード関数(共変量ベクトル $\mathbf{z} = \mathbf{0}$ の被験者の原因 j のハザード関数)、 β_j は原因 j のハザード関数への共変量の寄与を示す回帰係数 $\beta_j' = (\beta_{j1}, \dots, \beta_{jp})$ である。 β_j の推定には、競合するイベントを発現した被験者を形式的にはその時点での観察打ち切り例として扱うような、3 章と同様な部分尤度の最大化を行う。リスク集合はいずれのイベントも発現していない被験者から構成される。

原因 j のハザード関数に、層ごとに基準ハザード関数は異なり共変量ベクトルの寄与は層を通して同じであるようなモデルを仮定することもできる。そのとき、 h 層に属する被験者で共変量ベクトルが $\mathbf{z}$ であるときのハザード関数は 3.3 章と同様に表現できる。解析においては原因 j のイベントのみをイベントとし、それ以外のイベントを観察打ち切りとして取扱った場合の通常のプロポーションハザードモデルのように見なせるが、解釈は通常のプロポーションハザードモデルとは異なり、イベント発現確率に関する直接の推測はできない。 $I_j(t)$ は原因 j のハザードの影響のほかに全生存率の影響を受けるので、 $I_j(t)$ への共変量の影響は原因 j のハザード関数を通してのみではなく原因 j 以外のハザードを通しての影響としても表れてくる。更に、全ての原因別ハザードの基準ハザード関数の影響も受けることになる。このため、 $I_j(t)$ への共変量の影響は原因 j のハザードへの共変量の影響とは必ずしも同じではない。

4.6.2 CIF の直接の回帰モデル

Fine and Gray (1999) は CIF への直接の回帰モデルを提案した．共変量ベクトル $\mathbf{z}$ (p 次元) を持つ被験者の時間 t における原因 j の部分分布のハザード関数 $\gamma_j(t, \mathbf{z})$ に対し，比例ハザード性を仮定したモデル化で，以下の式で表現される．

$$\gamma_j(t, \mathbf{z}) = \gamma_{j0} \exp(\beta_j' \mathbf{z})$$

ただし， $\gamma_{j0}(t)$ は原因 j の部分分布の基準ハザード関数(共変量ベクトル $\mathbf{z} = \mathbf{0}$ の被験者の部分分布のハザード関数)， β_j は原因 j の部分分布のハザード関数への共変量の寄与を示す回帰係数 $\beta_j' = (\beta_{j1}, \dots, \beta_{jp})$ である．推定には通常の Cox 回帰と同様に部分尤度を最大化する方法を用いるが，リスク集合の定義は通常の部分尤度とは異なる．Gray の検定と同様の理由で，いずれのイベントも発現していない被験者のほかに，競合するイベントを発現した被験者はリスク集合に残り続ける．時間依存共変量に関する取り扱いが明確ではない(Putter et al., 2007)．

CIF を直接モデル化する別な方法も Klein and Andersen (2005) により提案された．推定には，部分尤度とは異なり，擬似値(pseudovalues, Andersen et al., 2003)と一般化推定方程式(generalized estimating equation (GEE), Liang and Zeger, 1986)を利用した．

4.6.3 そのほかの回帰モデル

原因別ハザードについての比例ハザードモデルや Fine and Gray (1999) モデルでは原因 j の部分分布のハザードごとにモデル化し推定を行うが，Klein (2006) は原因別ハザードの間や，部分分布の間に暗示的に要求される制約条件を反映した加法モデルを提案した．推定には，擬似値(Andersen et al., 2003; Klein and Andersen, 2005)と GEE (Liang and Zeger, 1986)を利用した．

4.7 視覚的な要約表示

通常の生存時間解析において分布の要約やモデルの適合度の確認のために様々な視覚的な表示方法が工夫されている(e.g. Lee and Wang, 2003, 8 章; Lawless, 2003, 3 章, Andersen et al., 1997)．原因別ハザードについての比例ハザード性の確認などには通常の生存時間解析における視覚的な表示方法が競合リスクモデルの解析においてもそのまま流用できる．競合リスクモデルでの解析の視覚的な要約として CIFE のみを表示している文献は数多いが，結果の解釈には競合リスクであるイベントの発現状況を見ておくことが重要である．この観点から有用な表示方法が Pepe and Mori (1993), Betensky et al. (2001), Nishikawa et al. (2006) などにより工夫されている．Betensky et al. (2001) の方法は Nishikawa et al. (2006) と類似しているので，Nishikawa et al. (2006) の表示方法を彼らの表 1 のデータを用いて紹介する．競合するイベントとして「初発 AE1 の発現」と「初発 AE1 の発現前中止」の 2 つを考え，それぞれタイプ 1 ($J=1$)，タイプ 2 ($J=2$) と呼ぶことにする．表 4 に彼らのデータの一部を簡素化して再現した．注目しているイベントはタイプ 1 のイベントであり，タイプ 2 のイベントは終了イベントである．25 人中 11 人に注目しているイベントが発現しているので，単純な割合により計算した発現率は 0.44 である．

図 2 にタイプ 1 及びタイプ 2 のイベントの CIFE を示した．図 2 では注目しているタイプ 1 のイベントの累積発生率は左側の縦軸で読むが，タイプ 1 と競合するタイプ 2 イベントの累積発生率は右側の縦軸で読む．右側の縦軸は下方向へ向かって増加を示す．競合するイベントの累積発

表 4. イベント発現までの時間

被験者 番号	イベント ^{a)} (タイプ, 時間)	被験者 番号	イベント ^{a)} (タイプ, 時間)	被験者 番号	イベント ^{a)} (タイプ, 時間)	被験者 番号	イベント ^{a)} (タイプ, 時間)
1	2,60	7	2,133	13	2,35	21	1,140
2	1,119	8	1,63	14	2,12	21	0,143
2	2,133	8	2,73	15	1,58	22	1,144
3	1,100	9	1,69	15	2,73	22	0,150
3	2,112	9	2,100	16	2,43	23	1,147
4	2,30	10	1,70	17	2,36	23	0,150
5	2,61	10	2,77	18	0,142	24	1,130
6	2,31	11	2,93	19	2,120	24	0,137
7	1,105	12	2,83	20	2,80	25	0,141

^{a)}: イベントのタイプ: 1, タイプ 1; 2, タイプ 2(終了イベント); 0, 観察打ち切り(無情報センサー)
時間の単位は「日」.

生率曲線を左側の軸で読めば, それは注目しているイベントを発現する可能性を持つ被験者の, 全被験者に対する経時的な割合となる. 左縦軸での破線と実線の距離は, ある時点 t までに観察を継続していて注目しているイベントも競合するイベントもまだ発現していない被験者の, 全被験者に対する経時的な割合 $\hat{S}(t)$ を示す. $I_1(t) + I_2(t) \leq 1$ という制約があるので, $I_2(t)$ が増大すれば $I_1(t)$ は増えることができない. すなわち, 競合するイベントが多発すれば注目しているイベン

図 2. タイプ 1 イベント, タイプ 2 イベント及びタイプ 1 イベント後の終了イベントの CIFE.

トの発現は多くならない. 図 2 では $t = 147$ で $I_1(t)$ が上昇し, $I_2(t)$ と結合している. $t = 147$ までは観察を継続すれば被験者の全員において注目しているイベントまたは競合するイベントが発現していることがわかる. 図 2 では左縦軸で読む破線に対する実線の比率が「条件付確率」 $\widehat{CP}_1(t)$ となる. 競合するイベントが発現せずに観察を継続できた場合, $t = 147$ までは全員において注目しているイベントが発現していることがわかる. 表 4 のデータを用いて, タイプ 1 イベント(軽症イベント)後に終了イベントを発現する累積同時発生関数の CJIFE を図 2 に太線で示している. CJIFE は左側の縦軸で読む. 図 2 の例では, 注目しているイベントを発現すればそのあとほどなく終了イベントを発現することがわかる. Pepe and Mori (1993) では複数の要約グラフを示しながら, それらの使い分けが検討されているのでそちらも参照して欲しい.

4.8 文献での適用例

これまでに解説した方法の適用例を紹介する.

(1) Martelli et al. (2005) の比較試験

従来, 乳がん患者に対して, 胸部手術と一緒に腋窩リンパ節郭清(AD)がなされていた. AD は治療のためではなく手術の後の治療方針を決めるためになされてきたが, 侵襲があり, AD による AE のために QOL を低下させることもある. 高齢の乳がん患者に対して AD が有効であるのかデータが不足していた. そこで, 腋窩リンパ節転移のない初期の高齢の乳がん患者に対して, AD を省略しても安全ではないかという仮説を立て, 従来の胸部手術及び AD(AD 群)と従来の胸部手術のみ(AD 省略群)とを無作為化比較試験により比較した. 主要評価指標を腋窩リンパ節に関するイベント(AD 省略群のみ), 全死亡, 乳がんによる死亡, 乳がんに関連したイベントとして, 5 年間追跡調査した. Tamoxifen は全員に処方された. 全死亡はログランク検定, 乳がんによる死亡及び乳がんに関連したイベントは Gray 検定により比較した.

AD 群及び AD 省略群に, それぞれ 109 及び 110 例が割り付けられた. 5 年時点の乳がんによる死亡の累積発生率は AD 群において 3.95%, AD 省略群において 3.93%であった. 全死亡の累積発生率は AD 群において 13.1%, AD 省略群において 7.8%であり, 原因別死亡の累積発生率の和となっている(原著論文の表 2). 全死亡の差は AD 省略群が AD 群よりも 5.3% (95% 信頼区間, $-3.3 \sim 13.9\%$)良好であった. 乳がんに関連したイベントは, AD 群及び AD 省略群においてそれぞれ 10.1% 及び 8.6% (原著論文の表 2 より計算)であった. 計画した主要評価指標の解析においていずれにおいても群間に有意差は認められなかった.

論文は CIFE で推測した乳がんに関連したイベント累積発生率, 及び乳がんによる死亡の累積発生率を図示し, 上述の CIFE とその 95% 信頼区間を表に示している. KME による全生存率も図示した. 乳がんによる死亡はそのほかの原因による死亡という競合リスクに曝されている. 対象が高齢者であるから乳がんに関連したイベントが起こる前に死亡する可能性もあり, 乳がんに関連したイベントは死亡という競合リスクに曝されている. CIFE はそのような状況での累積発生率と解釈できる.

(2) 抗精神病薬の大規模市販後試験(Lieberman et al., 2005)

5 種類(olanzapine, perphenazine, quetiapine, 及び risperidone, ziprasidone を承認後に追加)の抗精神病薬の有効性を比較する二重盲検無作為化比較試験で米国の国立精神保健局(National

Institute of Mental Health)が実施した大規模な試験である。ここでは有効性の主要評価指標を、あらゆる理由による試験薬の中止(投与薬剤の変更も含む)とし、中止までの時間(試験薬継続期間)を解析変数とした。中止の理由は効果不十分、安全性の問題など4種類にカテゴリー化され、副次解析として理由別に中止までの時間を解析し薬剤間で比較した。いずれかの理由による試験薬の中止が発現していない割合(試験薬継続率)、及びそれぞれの中止理由ごとに試験薬の中止が発現していない割合をKaplan-Meier法(4.2章の G_j)で推定し、Cox回帰における回帰係数を薬剤間で比較した。

1493例が割り付けられたが、ある一施設のデータは質の問題から盲検解除前に除外された。1460例が解析対象とされた。いずれかの理由により試験薬を中止した割合はolanzapine群において64%、perphenazine群において75%、quetiapine群において82%、risperidone群において74%、及びziprasidone群において79%であった。理由別中止は1章で説明した原因別死亡と同様に競合リスク要因になる。1-KMEで推測した理由別中止率が原著論文の図2に示されている。例えば、図2から目視により判読したquetiapine群における理由別中止率は、効果不十分: 45%、安全性の問題: 22%、患者の決定: 47%であり、第4のカテゴリー(そのほかの理由)については表示されていないが、これらの3つの理由による中止率の合計はすでに1を越える。ただし、論文にはKMEで推測した理由別継続率を図示してはいるが、表2には単純な割合として計算した理由別中止率を示している。表2においては、quetiapine群における理由別中止率は、効果不十分: 28%、安全性の問題: 15%、患者の決定: 33%である(第4のカテゴリーについては表示されていない)。また、最終時点での理由別中止率についての χ^2 検定も行っている。「いずれかの理由による試験の中止が発現していない割合」は4.1章の $S(t)$ を推定することであるからKMEで全く問題はないが、4.3章の G_j を理由別中止率として解釈するための問題点は4.3章に詳しく述べた。理由別中止の間に独立性を仮定しなくても原因別ハザードに関する推測が行えることは、4.1章及び4.6章などに詳しく述べた。手法の適用とその結果の解釈は、通常の競合リスクが存在しない場合のそれとは異なることに注意が必要である。

5. おわりに

競合リスクモデルは多段階遷移モデル(multistate models, e.g. Therneau and Grambsch, 2000; Lawless, 2003)、生涯(イベント)履歴過程(life (event) history processes, Kalbfleisch and Prentice, 2002; Lawless, 2003)の一部として見ることもできる(Klein, 2005; Andersen et al., 1993)。準競合リスクは多段階遷移モデルの考えを一部反映している。競合リスクモデルや多段階遷移モデルについては様々な観点からの解析手法の研究が増加している。競合リスクモデルについては自習のための本(Kleinbaum and Klein, 2005)でも解説されるほどに広く応用されるようになってきた。本文では紹介しなかったが、 CP_j についての k 群間の比較はLunn (1998)を参照して欲しい。ある母集団での原因別死亡を比較するような1標本内での原因別ハザードの間の比較はLam (1998)を参照して欲しい。

競合リスクの問題に関する研究が増加している一方、競合リスク特有の解析手法の汎用ソフトはまだあまり多くないようである。本文で紹介した文献のほか、Pintilie (2006)にR及びSASの

競合リスク解析手法のコーディングが公開されている。Putter et al. (2007) には競合リスク解析に関して s-plus, R, Stata などの説明がされている。競合リスク解析についてユーザーが必要に応じて、適切な手法が応用できる環境にだんだんなりつつあるようである。

従来、競合リスクの問題があまり意識されていない、または、競合リスクの手法があまり知られていないためにイベント発現までの時間で観察打ち切りがあればいつも KME, という短絡的な解析方法がとられる傾向も見られていたが、競合リスクが存在する場合の KME, 及び原因別ハザードなどの意味合いを理解したうえで適切な解釈がなされることが望まれる。

謝 辞

貴重なコメントを下さった査読者の方々、編集委員長の松山 裕博士、医学用語について助言を頂きました国立保健医療科学院横山徹爾博士に心より感謝致します。本研究の一部は、科学研究費基盤研究(C) No.18500224, No.20500267 の支援を受けて行ったものである。

参考文献

- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Process*. New York; Springer-Verlag.
- Andersen, P. K., Klein, J. P., and Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, **90**, 15-27.
- Bentzen, S. M., Vaeth, M., Pedersen, D. E., and Overgaard, I. (1995). Why actuarial estimates should be used in reporting late normal-tissue effects of cancer treatment. *International Journal of Radiation Oncology, Biology, Physics*, **32**, 1531- 1534.
- Betensky, R. A. and Schoenfeld, D. A. (2001). Nonparametric Estimation in a Cure Model with Random Cure Times. *Biometrics*, **57**, 282-286.
- Braun, T. M. and Yuan, Z. (2007). Comparing the small sample performance of several variance estimators under competing risks. *Statistics in Medicine*, **26**, 1170-1180.
- Breslow N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89-100.
- Caplan R. J., Pajak T. F., and Cox J. D. (1994). Analysis of the probability and risk of cause-specific failure. *International Journal of Radiation Oncology, Biology, Physics*, **29**, 1183-1186.
- Caplan R. J., Pajak T. F., and Cox J. D. (1995). In response to Bentzen et al. *International Journal of Radiation Oncology, Biology, Physics*, **32**, 1547.
- Caplan R. J., Pajak T. F., and Cox J. D. (1996). In response to Dr. Denham et al. *International Journal of Radiation Oncology, Biology, Physics*, **32**, 1547.
- Chappell, R. (1996). Re: Caplan et al. IJROBP 29:1183-1186; 1994, and Bentzen et al. IJROBP 32:1531-1534; 1995. *International Journal of Radiation Oncology, Biology, Physics*, **36**, 988-989.

- Choudhury, J. B. (2002). Non-parametric confidence interval estimation for competing risks analysis: application to contraceptive data. *Statistics in Medicine*, **21**, 1129-1144.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.
- Cox, D. R. and Oakes D. (1984). Analysis of survival Data. Chapman and Hall.
- Crowder M. (2001). *Classical Competing Risks*. Chapman & Hall/CRC: Boca Raton.
- DeLong, D. M., Guirguis, G. H., and So, Y. C. (1994). Efficient Computation of Subset Selection Probabilities with Application to Cox Regression. *Biometrika*, **81**, 607 - 611.
- Denham, J. W., Hamilton, C. S., and O'Brien, P. (1996). Regarding actuarial late effect analysis: Bentzen et al., IJROBP 32:1531-1534; 1995 and Caplan et al., IJROBP 32:1547; 1995. *International Journal of Radiation Oncology, Biology, Physics*, **35**, 197.
- Dinse, G. E. and Larson, M. G. (1986). A note on semi-Markov models for partially censored data. *Biometrika*, **73**, 379-386.
- Efron, B. (1967). The two sample problem with censored data. *Proceeding of Fifth Berkeley Symposium*, **4**, 831-853.
- Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72**, 557-565.
- Efron, B. (1981). Censored data and bootstrap. *Journal of the American Statistical Association*, **76**, 312-321.
- Farley, T. M. M., Ali, M. M., and Slaymaker, E. (2001). Competing approaches to analysis of failure times with competing risks. *Statistics in Medicine*, **20**, 3601-3610.
- Fine, J. P., and Gray, R. J. (1999). A Proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496-509.
- Fine, J. P., Jiang H., and Chappell, R. (2001). On Semi-Competing Risks Data. *Biometrika*, **88**, 907-919.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis*. John Wiley & Sons.
- Gaynor, J., Feuer, E. J., Tan, C. C., et al. (1993). On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. *Journal of the American Statistical Association*, **88**, 400-409.
- Gooley, T. A., Leisenring, W, Crowley, J, and Storer, B. E. (1999). Estimation of failure probabilities in the presence of Competing risks: New representation of old estimators. *Statistics in Medicine*, **18**, 695-706.
- Jpn J Biomet Vol.29, No.2, 2008

- Gray, R. J. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*, **16**, 1141-1154.
- Green, S., Benedetti, J., and Crowley, J. (2003). *Clinical Trial in Oncology, Interdisciplinary Statistics*. Second edition. Chapman & Hall.
- Greenwood, M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects*. **33**, 1-26. Her Majesty's Stationery Office, London.
- Hosmer, D. W. and Lemeshow, S. (1999). *Applied Survival Analysis*. John Wiley & Sons.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley: New York.
- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Second Edition. John Wiley & Sons: New Jersey.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457-481.
- Katsahian1, S., Porcher, R., Mary, J-Y. and Chevret, S. (2004). The graft-versus-leukaemia effect after allogeneic bone-marrow transplantation: assessment through competing risks approaches. *Statistics in Medicine*, **23**, 3851-3863.
- Klareskog, L, van der Heijde, D, de Jager, J. P., et al. (2004). Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: Double-blind randomised controlled trial. *Lancet*, **363**, 675-681.
- Klein, J. P. (2005). Book review. Classical competing risks. *Statistics in Medicine*, **24**, 158-159.
- Klein, J. P. (2006). Modelling competing risks in cancer studies. *Statistics in Medicine*, **25**, 1015-1034.
- Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag.
- Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, **61**, 223-229.
- Kleinbaum, D. G. and Klein, M. (2005). *Survival Analysis, A Self-learning Text*. Second edition. Springer.
- Lam, K. F. (1998). A class of tests for the equality of cause-specific hazards rates in a competing risks model. *Biometrika*, **85**, 179-188.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Second edition. John Wiley & Sons.
- Lee, E. T. and Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis*. Third edition. John Wiley & Sons.

- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.
- Lieberman, J. A., Stroup, T. S., McEvoy, J. P., et al. (2005). Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *New England Journal of Medicine*, **353**, 1209-1223.
- Lin, D. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine*, **16**, 901-910.
- Lin, D. Y., Sun, W, and Ying, Z. (1999). Non-parametric estimation of the gap time distributions for serial events with censored data. *Biometrika*, **86**, 59-70.
- Lipsky, P. E., van der Heijde, D. M., St Clair, E. W., et al. (2000). Infliximab and methotrexate in the treatment of rheumatoid arthritis. *New England Journal of Medicine*, **343**, 1594-1602.
- Lunn, M. (1998). Applying k-sample tests to conditional probabilities for competing risks in a clinical trial. *Biometrics*, **54**, 1662-1672.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163-170.
- Martelli, G., Boracchi, P., De Palo, M., et al. (2005). A randomized trial comparing axillary dissection to no axillary dissection in older patients with T1N0 breast cancer results after 5 years of follow-up. *Annals of Surgery*, **242**, 1-6.
- Marubini, E. and Valsecchi, M. G. (2004). *Analysing Survival Data from Clinical Trials and Observational Studies*. Reprinted in paperback, Wiley: New York.
- Nishikawa, M., Tango, T., Ogawa, M. (2006). Non-Parametric inference of adverse events under informative censoring, *Statistics in Medicine*, **25**, 3981-4003.
- 中村 剛. (2001). *Cox 比例ハザードモデル*. 朝倉書店.
- 西川正子. (2006). 有害事象の経時的発現状況の推測. 丹後俊郎, 上坂浩之編集. 臨床試験ハンドブック. 朝倉書店 p. 607-616.
- 大橋靖雄, 浜田知久馬. (1995). 生存時間解析. 東京大学出版会.
- Oka, H., Yamamoto, S., Kuroki, T., et al. (1995). Prospective study of chemoprevention of hepatocellular carcinoma with Sho-saiko-to (TJ-9). *Cancer*, **76**, 743-749.
- Pepe, M. S., and Mori, M. (1993). Kaplan-Meier, marginal, or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, **12**, 737-751.
- Pepe, M. S. (1991). Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, **86**, 770-778.
- Pintilie, M. (2006). *Competing Risks, a Practical Perspective*. John Wiley & Sons.
- Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics; Competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389-2430.
- Jpn J Biomet Vol.29, No.2, 2008

- Pitt, B., Poole-Wilson, P. A., Segal, R., Martinez, F. A., Dickstein, K., Camm, A. J., Konstam, M. A., Riegger, G., Klinger, G. H., Neaton, J., Sharma, D., and Thiyagarajan, B. on behalf of the ELITE II investigators. (2000). Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial — the Losartan Heart Failure Survival Study ELITE II. *Lancet*, **355**, 1582-1587.
- SAS Institute Inc. (2004). SAS OnlineDocR 9.1.3. Cary, NC: SAS Institute Inc.
- Southern, D. A., Faris, P. D., Brant, R., Galbraith, P. D., Norris, C. M., Knudtson, M. L., Ghali, W. A. for the APPROACH Investigators. (2006). Kaplan-Meier methods yielded misleading results in competing risk scenarios. *Journal of Clinical Epidemiology*, **59**, 1110-1114.
- Schwarzer, G., Schumacher, M., Maurer, T. B., Ochsner, P. E. (2001). Statistical analysis of failure times in total joint replacement. *Journal of Clinical Epidemiology*, **54**, 997-1003.
- 佐藤俊哉. (2005). 宇宙怪人しまりす医療統計を学ぶ. 岩波書店.
- Tai, B. C., Machin, D., White, I., and Gebiski, V; On behalf of the EOI (The European Osteosarcoma Intergroup). (2001). Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Statistics in Medicine*, **20**, 661-684.
- Tai, B. C., Peregoudov, A. and Machin, D. (2001). A competing risk approach to the analysis of trials of alternative intra-uterine devices (IUDs) for fertility regulation. *Statistics in Medicine*, **20**, 3589-3600.
- Tai, B. C., White, I. R., Gebiski, V., and Machin, D.; On behalf of the EOI (The European Osteosarcoma Intergroup). (2002). On the issue of 'multiple' first failures in competing risks analysis. *Statistics in Medicine*, **21**, 2243-2255.
- Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data. Extending the Cox Model. Statistics for Biology and Health*. Springer: New York.
- Tsiatis, A. A. (1975). A non-identifiability aspect of the problem of competing risks. *Proceeding of the National Academy of Science*, **72**, 20-22.
- Wang, W. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, **85**, 561-572.