

費用対効果評価における生存時間解析の諸問題

日本製薬工業協会

医薬品評価委員会 データサイエンス部会

2019 年度継続タスクフォース 5

「医療経済評価に関連する統計的手法の整理と紹介」

Ver 1.0

2020 年 2 月

本報告書の中で **Technical Support Document 14, 16, 及び 19 (Decision Support Unit, National Institute for Health and Care Excellence)** を引用/参考にした記載は、本タスクフォースの解釈であり、NICE には一切の責任はない。また各 TSD のすべてを紹介したわけではなく、一部を抜粋し紹介している。それらの紹介にあたり、タスクフォースでの追加の解釈を加えている。これらは本タスクが独自にまとめたものであり、NICE から見解を得たものではない。

目次

はじめに	1
1 生存時間のモデリング	2
1.1 生存時間解析の基礎	3
1.2 従来のパラメトリックモデルによるモデリング	4
1.2.1 指数分布	5
1.2.2 ワイブル分布	5
1.2.3 ゴンペルツ分布	6
1.2.4 対数正規分布	6
1.2.5 対数ロジスティック分布	6
1.2.6 一般化ガンマ分布	7
1.3 従来のパラメトリックモデル以外によるモデリング	7
1.3.1 区分モデル	7
1.3.2 Royston-Parmar model	8
1.4 生存時間のモデリングのパターン	10
1.5 モデルの適合性評価：内的妥当性と外的妥当性	10
1.6 内的妥当性の評価方法	11
1.6.1 目視確認	11
1.6.2 2重対数プロット	11
1.6.3 AIC, BIC	13
1.6.4 その他の方法	14
1.6.5 内的妥当性の評価の留意点	15
1.7 外的妥当性の評価方法	15
1.8 感度分析の実施	16
1.9 事例紹介	16
1.10 生存時間のモデリングの手順	19
1.11 生存時間のモデリングのまとめ	21
2 分割生存時間モデル	22

2.1	分割生存時間モデルの手順	23
2.2	分割生存時間モデルの適用事例の紹介	24
2.3	NICE TA での分割生存時間モデルの使用実態	29
2.4	状態推移モデル	31
2.5	過去の文献での分割生存時間モデルと状態推移モデルの比較	31
2.6	分割生存時間モデルと状態推移モデルの相違点	33
2.7	分割生存時間モデル適用時の利点と留意点	35
2.8	分割生存時間モデルの推奨使用方法のまとめ	37
2.8.1	モデリング手法の選択, 記述, 正当化	37
2.8.2	外挿に関連する不確実性の表現方法	38
2.8.3	モデリングにおける代替手段の使用	40
3	スイッチングとはなにか?	43
3.1	スイッチングがもたらす問題	43
3.2	スイッチングの調整法(シンプルな方法)[51].....	45
3.2.1	ITT 解析	45
3.2.2	Per protocol analysis(スイッチングのある患者を除外/打ち切りとする)	45
3.2.3	スイッチング後の費用を含める方法	45
3.2.4	PFS のみに基づくモデリング	46
3.2.5	疾患増悪後の死亡リスクを両群同一とみなす解析	46
3.2.6	両群で同一の OS を設定する方法	46
3.2.7	逐次モデルの使用	47
3.3	スイッチングの調整法(高度な方法)	47
3.3.1	Inverse Probability of Censoring Weights(IPCW 法).....	47
3.3.2	Rank Preserving Structural Failure Time Model (RPSFTM)	48
3.3.3	Iterative Parameter Estimation algorithm(IPE)[53].....	49
3.3.4	Alternative “two-stage” method	50
3.3.5	Using external data	51
3.4	費用対効果評価への適用	51
3.4.1	理論的限界	51

3.4.2	実務上の限界	52
3.4.3	方法とプロセスガイダンス	53
3.5	ディスカッション	58
おわりに		59
引用文献		60

はじめに

“どれほど費用がかかろうともその生命が危機にある特定の人物を助けようと試みる強力な人間の衝動が存在する。これは、「救助原則（rule of rescue）」として知られている。保健医療のために限られた資源しか存在しない場合、「救助原則」を適用すれば、必要とするケアや治療をそれ以外の人々が受けられなくなることを意味する可能性がある。” [1]

本邦では令和元年より医薬品及び医療機器の価格調整を目的として、費用対効果評価制度が本格導入された[2]。英国では約 20 年前より NICE（National Institute for Health and Care Excellence）が医療技術等の保険償還の可否、及び新薬の薬価に関連した費用対効果評価及び総合的評価（アプレイザル）を実施しており、制度のみならず、評価の事例や技術的側面についても参考になるものが多い[3]。NICE の DSU（Decision Support Unit）は、様々な技術的文書（Technical Support Document, TSD）を作成しており、現在までに 20 の TSD が発行されている[4]。本タスクフォースでは NICE の TSD で扱われるトピックスを整理して、国内で費用対効果評価を実践する際に有用と思われる情報を提供する活動を行ってきた。

費用対効果評価においても、使用するデータの多くは臨床試験等から得られるものであり、臨床試験の解析に用いられる手法の応用が基本的には可能である。一方で費用対効果評価では、原理原則は共有しながらも、一見独特とも見える手法により分析を行うことが多い。本報告書では、生存時間解析の費用対効果評価への応用に着目し、生存時間のモデル化、観察期間を超える予後の予測に関してその方法と留意点をまとめた。各章の内容は以下のとおりである。

- 1 章（TSD14 に対応）：費用対効果評価においては、多数のシナリオ、すなわち十分に長い分析期間における予後や諸費用を予測するモデルを構築することが必要である。このため、臨床成績を数学的に簡潔なモデルで示すことが必要となる。また、臨床試験の観察期間を大幅に超えるような長期間にわたる費用対効果評価がしばしば求められるが、数学モデルで示すことにより観察期間外への外挿が可能となる。
- 2 章（TSD19 に対応）：予後のモデル化の手法の 1 つとして分割生存時間モデルと呼ばれる方法を紹介する。多くの費用対効果評価で採用されているマルコフモデルと異なり、分割生存時間モデルでは臨床試験から実際に得られた生存時間曲線の曲線下面積をもとに各状態メンバーシップの患者割合を推定する。
- 3 章（TSD16 に対応）：実医療、実際の臨床試験のいずれでも発生する、治療スイッチング、すなわち対照治療を使用していた患者が新治療に切り替える事象が臨床成績に与えるバイアスについて論じ、その調整法をまとめる。

なお、本報告書は費用対効果評価の基本的な知識を前提にしている。背景や費用対効果評価の基本的な考え方等については、保健医療経済評価研究センター（CORE 2 Health）のサイト、中医協ガイドライン第 2 版、日本製薬工業協会の過去の報告書、成書等を参照されたい[5] [6] [7] [8] [9] [10] [11]。

1 生存時間のモデリング

がん領域や心疾患領域等の臨床研究では特定のイベント発現までの時間（time to event）を評価することが多い。例えば、がん領域であれば PFS（Progression Free Survival Time, 無増悪生存時間）、OS（Overall Survival Time, 生存時間）、心疾患であれば心筋梗塞発生までの時間、脳梗塞までの時間等を評価する。そして、多くの場合、臨床研究では全被験者のイベント発現までの時間を完全には追跡せず、追跡終了時点でイベントが発現していない被験者¹が多く存在する。臨床研究では評価期間においてイベント発現までの時間の分布（生存時間曲線）が治療群間で異なるかを評価することで研究目的を達成するため、全てのイベントを完全に追跡する必要はない。一方、費用対効果評価では生存時間曲線の比較よりもイベント発現までの時間自体に興味がある（生存時間の平均値に興味がある）ことから、イベント発現までの時間の平均値を評価するために臨床研究の追跡期間よりもさらに長い期間の評価が必要になる。このような場合、臨床研究の追跡期間を越えた部分に関しては外挿による推定が必要となる（図 1-1）。

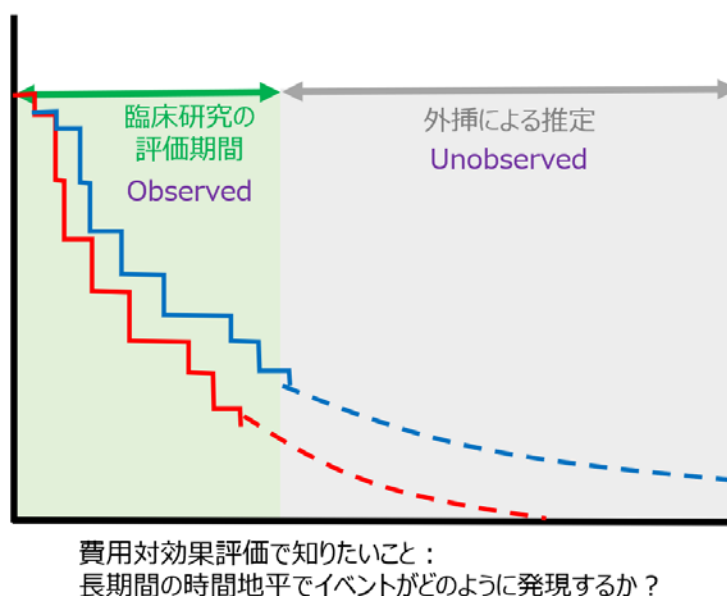


図 1-1 生存時間曲線の外挿による推定のイメージ

臨床研究から得られた Kaplan-Meier 曲線（実線）と臨床研究評価期間後の生存時間の推定（点線）

少々古い調査にはなるが、2011 年 6 月から 2013 年 8 月に報告書が公開された抗がん剤に関する NICE の技術評価（Technology Appraisal）[12]では、28/29（96.6%）で外挿による推定が行われていた。外挿による推定は費用対効果評価の結果に大きな影響を与える重要なプロセスだが、生存時間のモデリングに関して NICE による技術評価²では ERG（Evidence Review Group）から下記に示すような多くの指摘が

¹ 生存時間解析で「打ち切り」とよぶ。

² NICE は主に公衆衛生、診療、医療技術に関するガイダンスを作成する責任を持つ。医療技術については、企業が提出した有効性・安全性・費用対効果評価の結果と、それに対する ERG の批判的吟味等を参考に NICE が総合的評価（アプレイザル）を行い、医療技術の推奨可否を決定する。このプロセスが STA（Single Technology Appraisal）であり、検討結果は技術評価ガイダンスとして報告される。

挙げられている。

- ✓ 生存時間のモデリングに用いたモデルの選択が妥当ではなかった
- ✓ 生存時間のモデリングに用いたモデルはハザードの変化を捉えることができなかったため、区分モデルを優先すべきであった
- ✓ 生存時間のモデリングに用いたモデルに臨床的根拠が無かった
- ✓ 生存時間の長期的な外挿は非常に不確実性が高かった
- ✓ データの詳細を検討せずに過去の STA (Single Technology Appraisal) で適用された方法を用いて外挿していた

上記の指摘は様々な STA で挙げており、ときに ERG は適切な外挿方法に関して強い見解を示している。その一方で、2011 年～2013 年において ERG 間でもしばしばモデリングの考え方に整合性が取れていないケースも散見される。

生存時間のモデリングには様々な方法が提案されているため、これらの方法を一貫性と透明性を持って選択するために NICE は 2011 年 (2013 年に更新) に生存時間のモデリングに関する Technical Support Document 14 (TSD 14) として「Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data」を公開した[13]。これ以降もいくつかの論文で生存時間のモデリング方法の手順が提案されているが[14][15]、本章では TSD 14 を中心に、費用対効果評価を行うために必要な生存時間のモデリングについて紹介する。なお、本報告書では TSD 14 と同様に患者レベルのデータ (IPD, Individual patient-level data) が入手可能であり、試験間のエビデンスの統合が必要とならない状況、すなわち、1 つの主要な臨床研究に基づいて費用対効果を評価できる状況を想定する。利用できる情報が患者レベルのデータではなく論文等に記載された Kaplan-Meier 曲線及び要約統計量のみであり、そこから患者レベルのデータを再現したい場合は Guyot et al (2012) ³[16]を参照されたい。

1.1 生存時間解析の基礎

ある基準の時点からあるイベントが起きるまでの時間を対象とした解析方法を生存時間解析と呼ぶ。生存時間解析の詳細は成書を参照されたい[17][18]が、ここでは次項で様々なモデルを紹介する上で最低限必要と考える生存関数、ハザード関数、累積ハザード関数について説明する。

生存時間を表す非負の確率変数を T とする。この生存時間の分布の表現方法の 1 つに生存関数 $S(t)$ があり、生存関数は生存時間 T が時点 t を超える確率 (t までイベントが発生しない確率) である。

$$S(t) = \Pr(T \geq t)$$

³ イベント発現までの時間を評価したランダム化比較試験では、通常、Kaplan-Meier 曲線、イベント発現までの時間の中央値、コックス比例ハザードモデルのハザード比が報告される。Guyot et al は Kaplan-Meier 曲線から時点ごとのイベント数、生存確率、リスク集団の数を読み取り、患者レベルの生存時間データを再現するアルゴリズムを提案した。

一方、ハザード関数 $h(t)$ は、 $T \geq t$ となる条件の下で、次の瞬間に死亡が起こる確率（ある瞬間 t のイベント発現率）を意味し、ハザード関数と累積ハザード関数 $H(t)$ には以下の関係がある。また、生存関数とハザード関数は図 1-2 の関係がある。

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} = -\frac{d S(t)}{dt} \frac{1}{S(t)} = -\frac{d(\log S(t))}{dt}$$

$$H(t) = \int_0^t h(u) du = -\log S(t)$$

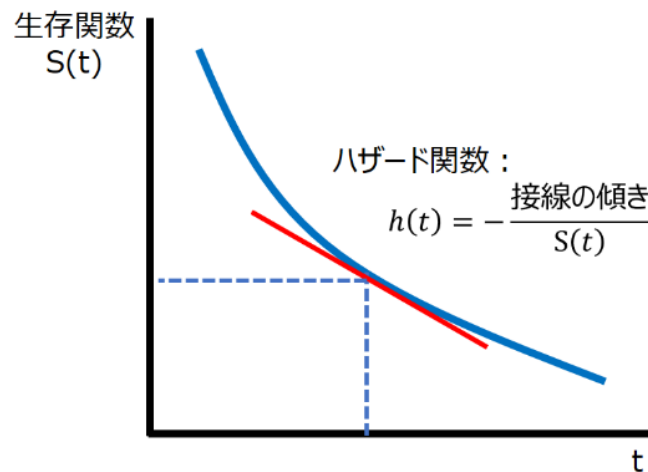


図 1-2 生存関数とハザード関数

このように、生存関数、ハザード関数、累積ハザード関数のいずれかが決まれば、他の関数が決まるため、生存時間のモデリングは得られたデータを生存関数又はハザード関数を用いてどのように表現するかの問題といえる。

1.2 従来のパラメトリックモデルによるモデリング

よく用いられる生存時間のモデリング方法として、（生存時間の分布がある確率分布に従うことを前提とした）パラメトリックモデルを当てはめる方法がある。2011 年 6 月から 2013 年 8 月に報告書が公開された抗がん剤に関する NICE の技術評価でも 22/29（75.9%）でパラメトリックモデルによるモデリングが行われていた（表 1-1）[19]。様々なパラメトリックモデルが存在し、パラメトリックモデルごとに特徴が異なる。そのため、経時的なハザードの変化を確認し、どの確率分布を当てはめるとあてはまりが良いかを検討する必要がある（図 1-3）。以下に代表的なパラメトリックモデルを紹介する。

表 1-1 抗がん剤に関する NICE の技術評価（2011/6～2013/8）における生存時間のモデリング方法

	PFS	OS
Non-parametric techniques only	1 (3.4%)	1 (3.4%)
Parametric techniques	22 (75.9%)	22 (75.9%)
Mix of non-parametric and parametric approach	6 (20.7%)	6 (20.7%)

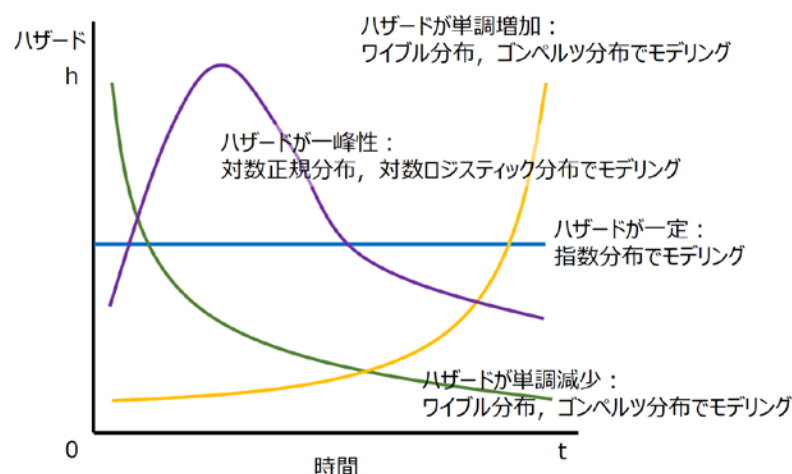


図 1-3 各パラメトリックモデルの特徴

1.2.1 指数分布

指数分布は、最もシンプルなパラメトリックモデルであり、ハザードが一定であることを仮定している。そのため、ハザードが一定と仮定できる場合に最も適したモデルである。

$$h(t) = \lambda$$

$$S(t) = \exp\left\{-\int_0^t \lambda \, du\right\} = \exp(-\lambda t)$$

1.2.2 ワイブル分布

ワイブル分布は、単調なハザードの増加又は減少が仮定できる場合に適したモデルである。形状パラメータである γ が 1 より大きいときハザードは単調に増加し、 γ が 1 より小さいときハザードは単調に減少する。なお、 γ が 1 のときは指数分布に帰着する。

$$h(t) = \lambda \gamma t^{\gamma-1}$$

$$S(t) = \exp\left\{-\int_0^t \lambda \gamma u^{\gamma-1} \, du\right\} = \exp(-\lambda t^\gamma)$$

1.2.3 ギンペルツ分布

ギンペルツ分布は、ワイブル分布と似ており、単調なハザードの増加又は減少が仮定できる場合に適したモデルである。形状パラメータである θ が 0 より大きいときハザードは単調に増加し、 θ が 0 より小さいときハザードは単調に減少する。

$$h(t) = \lambda \exp(\theta t)$$
$$S(t) = \exp\left(\frac{\lambda}{\theta}(1 - \exp(\theta t))\right)$$

ワイブル分布とギンペルツ分布の違いは、対数ハザード関数に関して、ギンペルツ分布は時間に対して線形であり、ワイブル分布は時間の対数に対して線形であるという点である。

$$\text{ギンペルツ分布 : } \log h(t) = \log \lambda + \theta t \quad \text{ワイブル分布 : } \log h(t) = \log \lambda + \log \gamma + (\gamma - 1) \log t$$

1.2.4 対数正規分布

対数正規分布は、2つのパラメータ (μ , σ) を持った対数をとると正規分布になる分布であり、時間に対して一峰性であり、早い時点でハザードが最大値に増加し、時間が経つにつれてハザードは減少する。

$$h(t) = \frac{\exp\left[-\frac{1}{2\sigma^2}(\log t - \mu)^2\right]}{\sqrt{2\pi}\sigma t \left[1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right]}$$
$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$$

Φ : 標準正規分布の分布関数

1.2.5 対数ロジスティック分布

対数ロジスティック分布は、2つのパラメータ (θ , κ) を持った対数をとるとロジスティック分布になる分布であり、対数正規分布と似ており、時間に対して一峰性になることが特徴である。対数正規分布と対数ロジスティック分布をデータから目視で見分けることはほとんど不可能である。

$$h(t) = \frac{e^{\theta \kappa t^{\kappa-1}}}{1 + e^{\theta t^{\kappa}}} \text{ for } \kappa > 0$$
$$S(t) = \{1 + e^{\theta t^{\kappa}}\}^{-1}$$

1.2.6 一般化ガンマ分布

一般化ガンマ分布は、3つのパラメータ (λ , ρ , θ) を持った柔軟なモデルであり、2つのガンマ分布のパラメータを一般化し、特殊なケースとしてワイブル分布、指数分布、対数正規分布に帰着する。

$$h(t) = \frac{f(t)}{S(t)}$$
$$S(t) = 1 - \Gamma_{(\lambda t)}^\theta(\rho)$$

$\Gamma_{(\lambda t)}(\rho)$: 不完全ガンマ関数

1.3 従来のパラメトリックモデル以外によるモデリング

前述した従来のパラメトリックモデル以外に区分モデル[20], Royston-Parmer model[21], Generalised F, Bayesian semi-parametric models[22]のような、より柔軟なモデル（より柔軟にハザードの変化を表現できるモデル）がある。このような柔軟なモデルは TSD14 作成時にはあまり使用されていなかったが、近年、免疫チェックポイント阻害剤のようにある時点でハザードが変わる事例で実際に使われ始めている（1.9 項参照）。本項では区分モデル、Royston-Parmer model に関して簡単に紹介するが詳細は原著を参照されたい。なお、従来のパラメトリックモデル又はより柔軟なモデルのいずれを選択するかは節約の原理⁴やオーバーフィッティング⁵も勘案する必要がある。

1.3.1 区分モデル

区分モデルは、前述のパラメトリックモデルよりも柔軟でかつ可変的なハザード関数をモデリングすることが可能である。代表的な区分モデルとして区分指数モデルがあり、区間ごとに指数分布を適用し、それぞれの期間は一定のハザードを有すると仮定するモデルである。以下に生存関数を示す。

既知の時点：

$$t_{(1)}, \dots, t_{(k)}$$

時間の区分：

$$(t_{(0)} = 0, t_{(1)}], (t_{(1)}, t_{(2)}], (t_{(2)}, t_{(3)}], \dots, (t_{(k)}, t_{(k+1)} = \infty]$$

⁴ 予測力が同等のモデルが複数存在する場合、モデルのパラメータの数がより少ないものを優れているとみなすモデル選択の原則。

⁵ モデリングに用いるデータは母集団からの標本抽出の結果に過ぎないが、このデータに対して過度に複雑なモデルをあてはめてしまうことで、未知のデータに対する汎用性が低くなってしまうこと。過剰適合。

時間の区分幅：

$$\delta_{j+1} = t_{(j+1)} - t_{(j)}$$

生存関数：

$$S_{j+1}(t) = \exp(-H_j) \exp(-h_{j+1}(t - t_{(j)}))$$

$$h_{j+1} : \text{区間}(t_{(k)}, t_{(k+1)}] \text{のハザード}$$

$$H_j = \sum_i h_i \delta_i : \text{時間} t_{(j)} \text{までの累積ハザード}$$

区分指数モデルは経時的にハザードが変化するようなデータのモデリングに有用であるが、外挿部分のハザードは観察されないため、外挿部分のハザードをどのように設定するかが問題となる。

1.3.2 Royston-Parmar model

2002 年に Royston と Parmar は比例ハザードモデルのベースライン対数累積ハザードを柔軟にモデル化するために制限付き 3 次スプラインを用いた Royston-Parmar model を提案した。

ワイブル分布の生存関数は $S(t) = \exp(-\lambda t^\gamma)$ であり、対数累積ハザード関数は以下の式になる。

$$\log\{H(t)\} = \log[-\log\{S(t)\}] = \log(\lambda) + \gamma \log(t)$$

$\log\{H(t)\}$ は傾き γ の $\log(t)$ の 1 次関数である。この式に共変量 $\mathbf{x}_i$ を追加すると以下の式になる。

$$\log\{H(t|\mathbf{x}_i)\} = \log(\lambda) + \gamma \log(t) + \boldsymbol{\beta}\mathbf{x}_i$$

第 1 項と第 2 項は対数時間に対する線形のベースライン関数（ベースライン対数累積ハザード）と考えると、下記の式に書き換えることができる。

$$\log\{H(t|\mathbf{x}_i)\} = \log\{H_0(t)\} + \boldsymbol{\beta}\mathbf{x}_i$$

Royston と Parmar は $\log\{H_0(t)\}$ に制限付き 3 次スプラインを用いることを提案した。

$$\log\{H(t|\mathbf{x}_i)\} = s(\log(t) | \eta, k_0) + \boldsymbol{\beta}\mathbf{x}_i$$

ここで $s(\log(t) | \eta, k_0)$ は制限付き 3 次スプラインである。スプラインとは区分多項式で定義された柔軟なパラメトリック関数で、多項式がつながる点をノットと呼ぶ。図 1-4 のスプライン関数のノットは 3 つ（破線）であり、4 つの多項式をつなげている。 $s(\log(t) | \eta, k_0)$ はノットの数 k_0 と η の関数である。

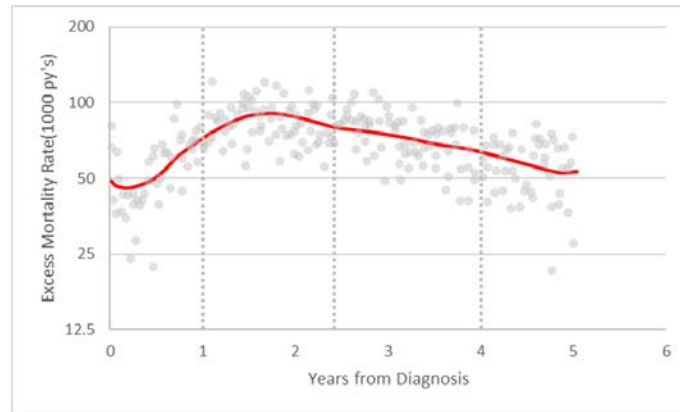


図 1-4 制限付き 3 次スプラインの例

制限付き 3 次スプラインは以下の式で表現される。

$$S(x) = \eta_0 + \eta_1 z_1 + \eta_2 z_2 + \cdots + \eta_{K-1} z_{K-1}$$

ここで K はノットの数, z_i は導出変数, η_i は導出変数の係数である。導出変数は以下の式により算出する。

$$z_1 = x$$

$$z_j = (x - k_j)^3 - \psi_j (x - k_j)^3 - (1 - \psi_j)(x - k_j)^3, \text{ for } j = 2, \dots, K - 1$$

x は前述の x_i とは異なり, k_j は j 番目のノットの位置, $\psi_j = (k_K - k_j)/(k_K - k_1)$ である。例えば, ノットが 4 つの場合, 累積対数ハザード関数は以下の式になる。

$$\log\{H(t|x_i)\} = \eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 z_3 + \beta x_i$$

実際に臨床試験の解析で活用された事例は Jacobs[23]等を参照されたい。

コラム：SAS を用いたパラメトリックモデルによるモデリング

(臨床試験の統計解析担当者には親しみがある) 統計解析ソフトウェアの SAS では LIFEREG プロシジャを用いて従来のパラメトリックモデルをあてはめることが可能である。プログラム例を以下に示す。

```
proc lifereg data = xxx outest = xxx ;
    model time * censor( 1 ) = treatment / dist = xxx ;
    output out= xxx xbeta = xxx ;
run ;
```

MODEL ステートメントの DISTRIBUTION オプションで確率分布を指定することで、本項で紹介した様々なパラメトリックモデルをあてはめた結果を確認することができる。また、デフォルトで後述する AIC 及び BIC も出力され、あてはめたモデルの生存時間曲線をプロットしたい場合には LIFEREG ステートメントの OUTEST オプションと OUTPUT ステートメントを用いればよい。より柔軟なモデルの区分指数分布, Royston-Parmar model を実行するためには ICPHREG, PHREG プロシジャを用いることになる。

1.4 生存時間のモデリングのパターン

評価対象技術と比較対照技術に対して生存時間のモデリングを行う際、大きく2つのパターンがある。治療群ごとに個別にモデリングするか、治療群間で比例ハザード性が成り立つと仮定し、共変量に治療群を含めて両治療群を同時にモデリング（比例ハザードモデル）するかである。なお、比例ハザード性とは治療群間のハザード比が時点によらず一定であることであり、比例ハザード性は二重対数プロット（後述）、残差プロット、比例ハザード性の検定⁶で確認することができる。なお、両治療群を同時にモデリングする方法として、比例ハザードモデルの他に加速モデルを用いることも可能である。各モデルの詳細は成書を参照されたい[17]。

1.5 モデルの適合性評価：内的妥当性と外的妥当性

生存時間をモデリングした後はモデルの適合性を評価する必要がある。生存時間のモデリングの適合性を評価するために様々な方法が提案されており、NICE TA では表 1-2 に示すように方法が用いられてきた。生存時間のモデリングの適合性は、当てはめたモデルがデータに良好に当てはまるかどうかの「内的妥当性」、外挿部分が臨床的及び生物学的に尤もらしいかどうかの「外的妥当性」の2つの観点から評価する必要がある⁷。

⁶ 時間に依存する変数を人工的に作成しモデルに含めることで、この変数のパラメータ推定値とその標準誤差を用いて検定することで比例ハザード性を検討する。詳細は成書[17]を参照。

⁷ 「中央社会保険医療協議会における費用対効果評価の分析ガイドライン」におけるモデルの妥当性とは定義が異なることに留意する。ガイドラインではモデルの妥当性について内的妥当性を「なぜそのような構造のモデルを構築したのか、病態の自然経過を十分にとらえられているか、使用しているパラメータは適切なものか等」とし、外的妥当性を「既存の臨床データ等と比較して、モデルから得られた推計が適切なものであるか等」としている。ガイドラインの“モデル”はマルコフモデル等の構造を意図し、疾患状態の定義の妥当性について記載している。そのため、生存時間のモデルとは用語の意味が異なる。

表 1-2 NICE TA におけるパラメトリックモデルの適合性確認方法 (TSD14 Table 4 を一部改変)

妥当性確認の方法	NICE TA の利用状況
内的妥当性の確認	
統計的な検討 AIC, BIC, 2 重対数プロット, 残差プロット等	比較的少なく, 他の妥当性確認の方法と組み合わせて体系的に行われているわけではない.
目視確認	よく利用される.
外的妥当性の確認	
外部データ	少ない.
臨床的妥当性	少ない.

1.6 内的妥当性の評価方法

内的妥当性の評価方法として, 以下でそれぞれ説明する目視確認 (visual inspection), 2 重対数プロット, AIC/BIC, 残差プロット等が提案されている。これらのうち 1 つを用いて内的妥当性を評価すればいいというわけではなく, これらは相互に補完し合う関係であるため, それぞれの方法で適合性を評価し, 総合的にモデルのあてはまりを確認する必要がある。

1.6.1 目視確認

目視確認は Kaplan-Meier 曲線と当てはめたモデルをオーバープロットし, 視覚的にモデルのあてはまりを確認することであり, 最もベーシックなモデルのあてはまりを評価する方法である。この方法はあてはまりの程度を把握しやすいが, 定性的なため, 不確実でかつ不正確な側面がある。そのため, パラメトリックモデルの適合性を評価する際には他の方法と合わせて用いる必要がある。

1.6.2 2 重対数プロット

前述した通り, 生存時間のモデリングは経時的なハザードの変化を把握することが重要である。2 重対数プロットは縦軸に $\log(-\log \hat{S}(t))$, 横軸 $\log t$ を取ったグラフであり, ハザードの経時的な変化を確認する際に有用なグラフである。2 重対数プロットを用いて, ハザードが時点を通じて非単調, 単調, 一定なのか, 比例ハザード性を仮定したモデルが妥当かを検討する。また, ある時点でハザードに顕著な変化が生じる部分には区分モデルのように区間ごとに異なるハザードを持つモデルをあてはめるといった検討も必要である。

2 重対数プロットでは傾きが 1 の場合は指数分布, 傾きが単調の場合はワイブル分布, ゴンペルツ分布, 傾きが非単調な場合は対数ロジスティック分布, 対数正規分布, それ以外の場合はより柔軟なモデ

ルが候補となる。また、比例ハザードモデルの場合は治療群間で垂直方向に平行になり、加速モデルでは水平方向に平行になる（図 1-5）。

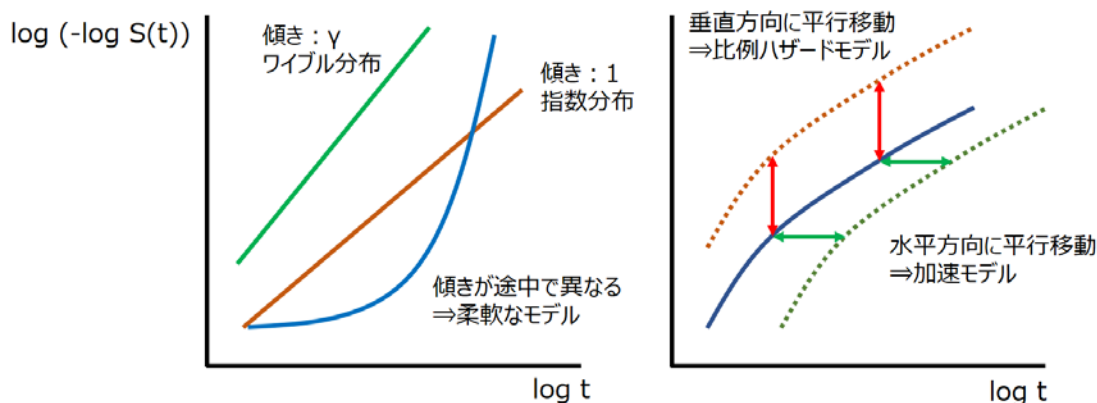


図 1-5 2重対数プロットと各モデルの関係

	指数分布	ワイブル分布
生存関数	$S(t) = \exp(-\lambda t)$	$S(t) = \exp(-\lambda t^\gamma)$
両辺の対数を取り，マイナス	$-\log S(t) = \lambda t$	$-\log S(t) = \lambda t^\gamma$
再度，両辺に対数を取ると，	$\log\{-\log S(t)\} = \log \lambda + \log t$	$\log\{-\log S(t)\} = \log \lambda + \gamma \log t$
2重対数プロット上での表現	傾きが 1	傾きが γ

	比例ハザードモデル	加速モデル
モデル式	$h(t; x) = h_0(t) \exp(\beta x)$ (ハザードについてモデル化)	$T = T_0 \exp(\beta x)$ $\log T = \log T_0 + \beta x$ (時間についてモデル化)
生存関数	$S(t) = S_0(t)^{\exp(\beta x)}$	$S(t) = S_0\{\exp(-\beta x) t\}$
両辺の対数を取り，マイナス	$-\log S(t)$ $= \exp(\beta x) (-\log S_0(t))$	—
再度，両辺に対数を取ると，	$\log\{-\log S(t)\} = \beta x$ $+ \log(-\log S_0(t))$	—
2重対数プロット上での表現	垂直方向に平行	水平方向に平行

例えば，図 1-6 の 2 重対数プロットでは 5 週 ($\exp(1.5)$) から重要なハザードの変化が見られるが，5 週以降は 2 群間のハザードは垂直方向で平行である。5 週以降のプロットの傾きは 1 未満であるように見えるため，5 週以降に関して指数分布を想定したモデルは適切ではなく，ワイブル分布が適切だろう。ただし，追跡調査終了時付近では試験薬群で急勾配があるため，この部分に関してはデータの詳細を確認する必要がある。

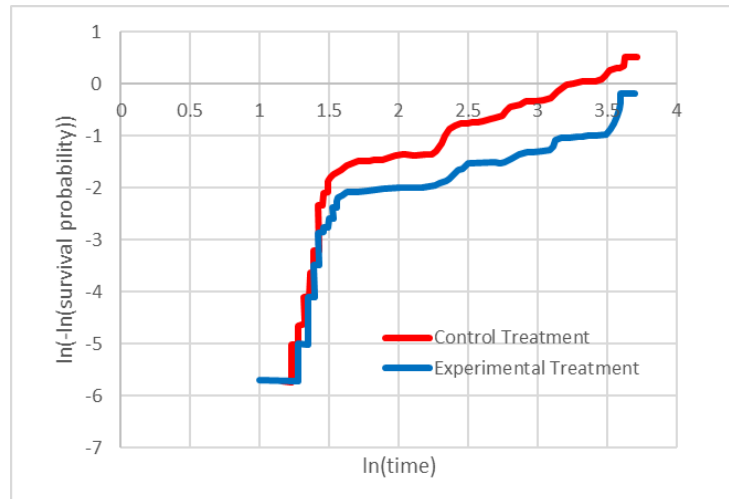


図 1-6 2重対数プロットの例（TSD14 Figure 2 を一部改変）

1.6.3 AIC, BIC

モデル選択の方法として尤度比検定があるが、尤度比検定はネスト構造⁸にある2つのモデルの逸脱度に関して統計的に有意差があるかを確認する方法であり、ネスト構造にないモデル同士の比較に用いることはできない（それに関わらず、過去のNICE TAsではネスト構造にない異なるモデル間で尤度比検定が誤用されているケースがあった）。本項で紹介するAIC（Akaike's Information Criterion）とBIC（Bayesian Information Criterion）はネストの構造にならないモデル間の比較が可能であり、パラメータ数に関してペナルティを設けている。例えば、指数分布は1つのパラメータだが、ワイブル分布、ゴンプエルツ分布は2つのパラメータを持つ。このような場合、ワイブル分布とゴンプエルツ分布は指数分布に比べてパラメータが多い分、ペナルティが適用される。AIC, BICはパラメトリックモデルの相対的な適合性に関する有用な統計的指標であり、様々な候補モデルがあった場合にAIC及びBICが小さいモデルを選択することになる（コラム参照）。

コラム：AIC と BIC

生存時間をモデリングする際、複数のモデルをあてはめ、得られたデータに対して最もあてはまりの良いモデルを選択する。良いモデルを選択するためには何らかの基準を導入し、あてはまりの良さを数値化し、その数値を複数のモデル間で比較することが考えられる。このモデルのあてはまり具合を判定するための基準の1つが情報量基準であり、代表的なものとしてAICとBICがある。

⁸ 2つのモデルがあるときに、一方のモデルに含まれる説明変数のセットが他方のモデルの説明変数の部分集合になっているとき、2つのモデルをネストしたモデルと呼ぶ。

【AIC】

AIC は以下の式で定義される。

$$AIC = -2 \log L(\hat{\theta}) + 2k$$

θ : モデルパラメータ, $L(\hat{\theta})$: 最大尤度, k : モデリングに必要なパラメータ数

ある単一のモデルに対して AIC の意味を解釈することは困難である。複数の候補モデルがある際に各モデルに対して AIC を計算し、AIC が最小のモデルを「最良の」モデルとする。第 1 項の対数尤度関数 $\log L(\theta)$ はモデルの下でデータが得られる確率の対数であり、第 2 項を無視すると尤度関数に -2 を掛けているため、AIC が最小のモデルが尤度関数の値が最も高いモデルとなる。AIC はモデルを作成する際に推定したパラメータの数で調整している（第 2 項）。つまり、推定したパラメータの数が多いほど、AIC が大きくなり、ペナルティが与えられている。これは節約の原理と整合する。

【BIC】

BIC は以下の式で定義される。

$$BIC = -2 \log L(\hat{\theta}) + k \log n$$

n はパラメータ推定に用いたデータ数である。右辺の第 1 項は AIC 基準の第 1 項と同じだが、第 2 項はモデリングに必要なパラメータ数が $\log n$ 倍されているため、AIC に比べてパラメータ数の増加に対するペナルティが強い。

1.6.4 その他の方法

AIC と BIC はモデルのあてはまりの良さを評価する指標として有用だが、あくまでモデリングに用いたデータに対するあてはまりの良さを評価している。モデリングに用いたデータは母集団から標本抽出により得られた標本であり、標本は不確実性を伴う。そのため、この不確実性を考慮するためにクロスバリデーションが提案されている。クロスバリデーションは、データをモデリング用のデータとモデルの検証用のデータにランダムに分割し、モデリングと検証を行う。

クロスバリデーションの手順（図 1-7）：

1. データを K 個のグループに分割し、 $K-1$ 個のデータを使ってモデリングを行い、残りの 1 つのデータで検証する
2. 検証に使うデータを変えて K 回モデリングと検証を繰り返し、あてはまりの良さを評価する

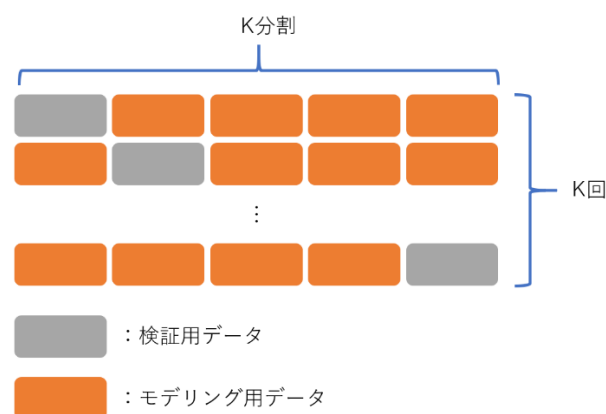


図 1-7 クロスバリデーション

クロスバリデーションではデータを分割し、あてはまりの良さを評価したが、統計的リサンプリングを用いたブートストラップ法による評価も提案されている[24]。その他にも残差プロット、DIC (deviance information criterion) を用いたあてはまりの評価も利用可能である。

1.6.5 内的妥当性の評価の留意点

目視確認、二重対数プロット、AIC 及び BIC はいずれもモデリングに用いたデータに対するモデルのあてはまりの良さを評価する点では有用だが、臨床研究の評価期間を越えた期間に関しては情報を与えてくれないことに留意する必要がある。これは生存時間曲線の外挿部分がイベント発現までの時間の平均値の推定に影響を及ぼす場合、とても重要な問題である。長期間にわたってイベントを追跡できている場合又はイベント発現までのデータが比較的完全に捕捉できている場合には二重対数プロット、AIC 及び BIC の結果は有用だろう。一方で、長期間の外挿が必要となる場合には、モデリングにより得られた外挿部分の妥当性（外的妥当性）を他の方法を用いて確認する必要がある。

1.7 外的妥当性の評価方法

外挿部分のモデリングの妥当性を評価する方法として、外部データの活用、臨床的妥当性の検討が考えられる。外部データに関しては、同様の患者集団に対してより長期の追跡を行った別の研究や当該患者集団を長期間追跡したレジストリーがあれば、これらの情報は強力な情報源となる。例えば、もしレジストリーにおいて特定の疾患に関する 5 年生存率が 10%であったときに、モデリングによる外挿結果が 0%や 40%となっていたら、そのモデルは適切とはいえないだろう。外部データを利用する際には、外部データと費用対効果評価の対象集団の違い等の外部データの特徴も考慮すべきである。評価対象技術が新規の技術であった場合には、比較対照技術にのみ長期間の外部データが存在するケースもあるだろう。このような場合、評価対象技術に対する長期のイベント発現までの時間を推定する上で外部データは役に立たない。したがって、評価対象技術の長期間のデータを外挿するためには、治療効果の持続

期間に関して臨床的に妥当な仮定を臨床専門家の意見や生物学的妥当性をもとに検討することになる。そして、このような仮定はシナリオ分析として感度分析の対象とすべきである。

Jackson et. al はランダム化比較試験からの短期間の個人レベルのデータと長期間の外部データを組み合わせて短期間の外挿する方法を提案しているため、必要に応じて参照されたい[25]。

1.8 感度分析の実施

費用対効果評価を行う際に生存時間のモデリングの不確実性を感度分析により検討することは重要である。具体的には、異なるモデルをあてはめた場合の影響や外挿に関する仮定（例えば、治療効果の持続期間に関する仮定）を変えた場合の影響を感度分析により確認することが考えられる。

1.9 事例紹介

生存時間のモデリングの具体的なイメージを掴むために、柔軟なモデリングとクロスバリデーションを用いた事例（NICE の技術評価の事例ではないが）を紹介する[14]。なお、あくまで一例であり、実際に生存時間のモデリングを行う場合にはケースバイケースで対応する必要がある。

近年、免疫チェックポイント阻害剤のように長期の予後ではハザードが小さくなるような薬剤が存在する。このようなケースに従来のパラメトリックモデルをあてはめるとイベント発現までの時間を過小評価してしまう可能性があるため、より柔軟なモデルをあてはめることが一案である。Gibson et al は未治療の転移性悪性黒色腫に対する治療を評価した CheckMate 067 試験の PFS に関して、生存時間のモデリングを行った。

<CheckMate 067 試験[26]の概要>

未治療の転移性悪性黒色腫に対する Nivolumab + Ipilimumab 併用療法、Nivolumab 単独療法、Ipilimumab 単独療法を比較したランダム化二重盲検第 III 相臨床試験

【対象】

根治切除不能又は転移を有する未治療の病期 III・IV の悪性黒色腫患者

【試験デザイン】

対象患者を 1:1:1 に層別ランダム化し（層別化因子：PD-L1 発現状況，BRAF 変異状況，AJCC M 分類），病勢進行又は毒性発現まで継続。

投与群：

- Nivolumab + Ipilimumab 群
- Nivolumab 群
- Ipilimumab 群

Nivolumab + Ipilimumab 群と Ipilimumab 群の Kaplan-Meier 曲線及び 2 重対数プロットを図 1-8 に示す。

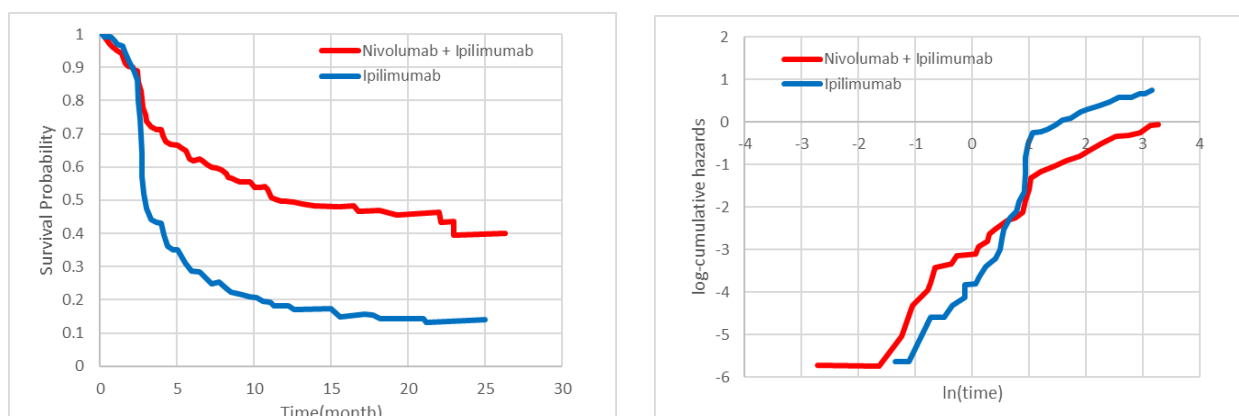


図 1-8 PFS の Kaplan-Meier 曲線 (左) と 2 重対数プロット (右)

(Gibson et al[14]の Fig.1 を一部改変, 打ち切り非表示)

2 重対数プロットを確認したところ, Nivolumab + Ipilimumab 群と Ipilimumab 群の間に比例ハザード性が成り立たないため, 比例ハザードモデルによるモデリングが適切でないと考え, 治療群ごとにワイブル分布, 指数分布, ギンペルツ分布, 対数正規分布, 対数ロジスティック分布, RCS (restricted cubic spline) モデル⁹[27]をあてはめている。

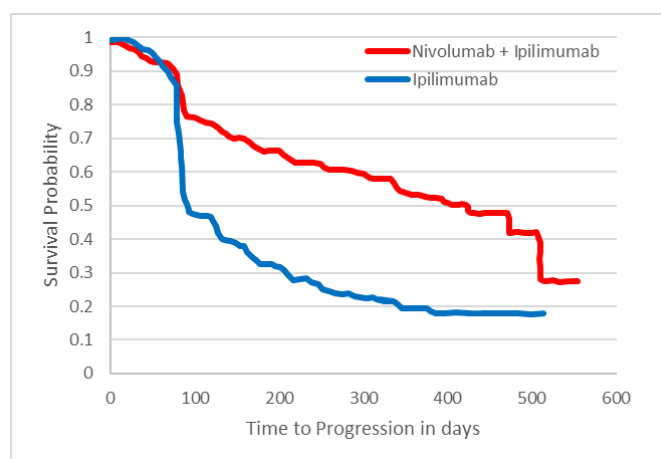
Kaplan-Meier 曲線と各モデルをオーバープロットし目視確認を行い, AIC 及び BIC を算出した (表 1-3)。なお, 内的妥当性を確認するためにクロスバリデーションを行い, 患者レベルのデータをランダムにモデリング用データと検証用データに 2 分割した (分割したデータ間で被験者背景と PFS に差が無いか確認)。モデリング用データであてはめた結果は検証用データでも同様の傾向が示され, 従来のパラメトリックモデルでは対数ロジスティック分布, RCS モデルではノットが 6 又は 7 の RCS モデルが目視確認及び AIC, BIC より支持された。外挿部分に関しては Ipilimumab 群を最長 4, 6 年間追跡した他試験のデータと比較して, RCS モデルが最も近い結果であることを確認した。ノットの数を増やすと全体的にあてはまりがよくなったが, 最終的に節約の原理も勘案し 1 つのノットの RCS モデルが最適なモデルと結論づけた。この点はオーバーフィッティングの観点からも支持されるだろう。

⁹ Royston-Parmar model は対数累積ハザードと対数時間の関係に線形性を仮定したワイブル分布がベースとなったモデルだが, RCS モデルはこの関係を非線形にすることでさらに柔軟性を高めた方法である。

表 1-3 トレーニングデータ及び検証用データに対する各モデルの AIC, BIC
(Gibsone et al[14]の Table 2 より引用)

	トレーニングデータセット				検証用データセット			
	Nivolumab + Ipilimumab (n = 154)		Ipilimumab (n = 135)		Nivolumab + Ipilimumab (n = 150)		Ipilimumab (n = 157)	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Weibull	449	455	398	403	428	434	467	473
Exponential	454	457	396	399	437	440	465	468
Gompertz	433	439	385	391	411	417	452	459
Log-logistic	438	444	341	347	421	427	415	421
Log normal	432	438	351	357	417	423	417	423
RCS 1 knot	417	426	301	310	412	421	371	380
RCS 2 knot	419	431	300	312	412	424	365	377
RCS 3 knot	405	420	245	259	414	429	345	361
RCS 4 knot	398	417	249	266	414	432	312	331
RCS 5 knot	400	422	209	230	416	437	332	353
RCS 6 knot	385	409	213	236	408	432	293	317
RCS 7 knot	394	422	226	252	399	426	304	331
AIC Akaike information criterion, BIC Bayesian information criterion, RCS restricted cubic spline								

なお、NICE TA400[28]において、同じ CheckMate 067 試験の TTP (time to progression, 図 1-9) についてモデリングを行っている。企業分析では Kaplan-Meier 曲線が開始時付近において両群で急激に落ちている理由として、CheckMate 067 試験では割付後の最初の腫瘍評価時間は 12 週後 (84 日後) になっており、この検査時点で Progression となっている患者が多かったためとしている。12 週以前の Progression の発現状況の詳細を確認することはできず、12 週以降に Kaplan-Meier 曲線が離れ始めていると考え、84 日より前は比例ハザードモデルを当てはめ、84 日以降は対数正規分布のあてはまりが良いと判断した¹⁰。



¹⁰ ERG は企業分析の TTP のモデリングに満足していないとコメントしている。具体的には、CheckMate 067 試験の 12 週時来院の許容範囲は±1 週間であったため、84 日で区切って前後でそれぞれモデルを当てはめるのではなく、許容範囲を考慮して 77 日又は 91 日で区切ることを提案している。詳細は TA400[28]をご確認いただきたい。

図 1-9 TTP の Kaplan-Meier 曲線（企業分析の Figure 25 を一部改変，打ち切り非表示）

なお，TA400 で企業分析は TSD14 にしたがって生存時間のモデリング行ったことが述べられている。

The best fitting models have been selected (in line with DSU TSD 14 guidance) by considering the visual fit of the parametric curves compared to the KM curves (separately for each trial/treatment), clinical plausibility of extrapolation and comparison of the Akaike information criteria (AIC) and Bayesian information criteria (BIC) values. To make consistent comparisons between fitted curves by treatments related to the KM curves, the parametric model estimates were applied to the covariate values as observed in the specific trial/treatment arm.

また，ERG は上記の手順で企業分析は適切に実施されたと判断したが，曲線の尾の方の評価では残差分析も有益とコメントしている。

The company carried out a parametric survival analysis. Model fit was assessed using relative fit statistics (i.e. AIC and BIC) and visual inspection. The ERG considers the analysis to have been carried out appropriately; however, residual analysis would have been informative to assess the behaviour of the parametric functions especially around the tails of the curves.

このように同じ Kaplan-Meier 曲線でも企業分析と ERG ではモデルのあてはまりの考え方が異なるケースがある。そのため，解析者はモデルを選択した根拠を明確に説明できるようにしておく必要がある。

1.10 生存時間のモデリングの手順

生存時間のモデリングには様々な方法があり，ケースバイケースで最適なモデルが異なる。そのため，どのモデルが最適というガイダンスを普遍的に提供することは困難である。しかし，モデリングのプロセスの透明性を確保するために，モデリングの手順を標準化することは可能である。NICE TSD14 では生存時間のモデリングの手順として図 1-10 を提案している。NICE TSD14 以外にも Treamblay ら[15]がモデリングの手順を提案しているため，必要に応じてご参照いただきたい。NICE TSD14 ではさらにモデリングプロセスをチャートで説明することを提案しており，仮想事例ではあるが Survival Model Selection For Economic Evaluations Process(SMEEP) Chart を例示している。

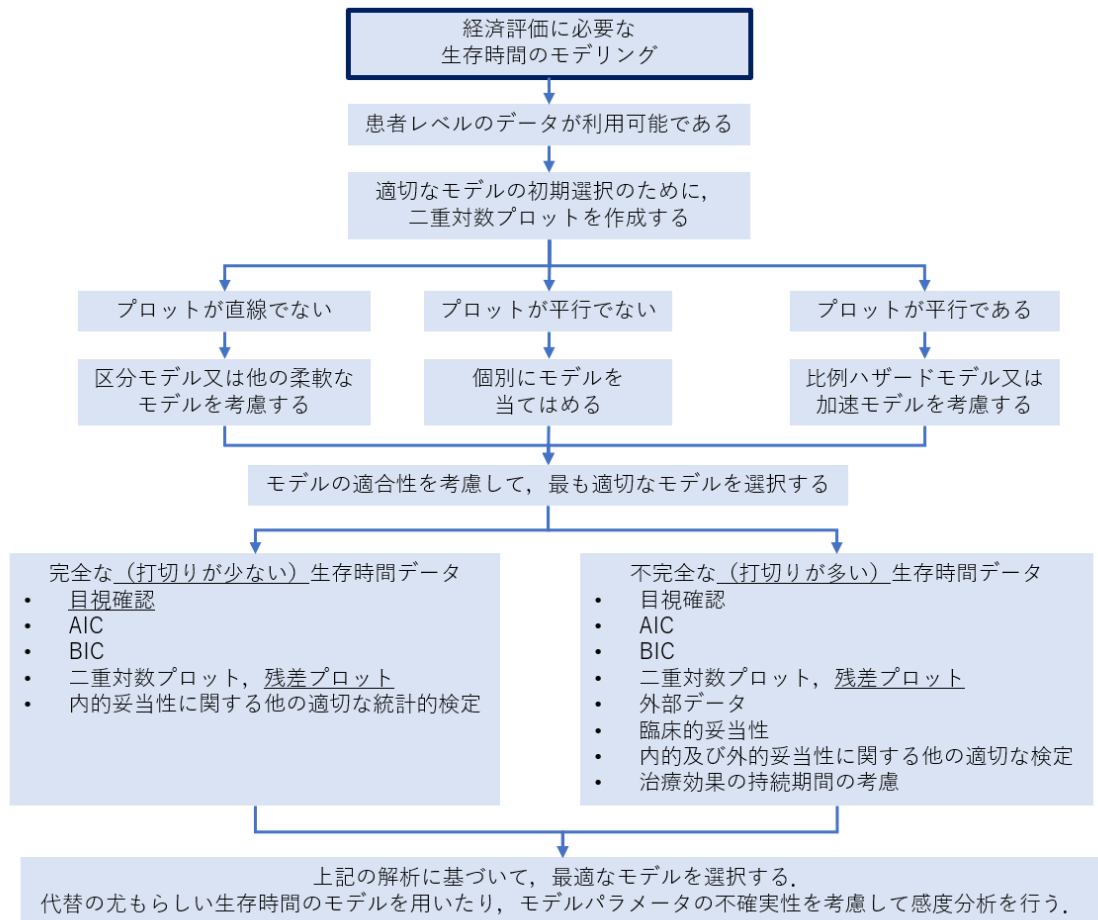


図 1-10 NICE TSD14 の Survival Model Selection Process Algorithm（和訳し、下線部を追記）

Step1

観察されたハザードの傾向を評価するために、2重対数プロットを作成する。ここでは以下の3点を主に検討する。

- ✓ どの種類のパラメトリックモデルが適しているか？
- ✓ 比例ハザード性を仮定できるか？
- ✓ ハザードの変化を区分ごとに考える必要があるか？

Step2

Step1の結果を踏まえ、2重対数プロットの形状によってどのモデルを用いるか検討する。

- ✓ 2重対数プロットがほぼ直線、単調な場合：

従来のパラメトリックモデルをデータに当てはめる。2つの治療群のハザードのプロットが平行である場合は比例ハザードモデル又は加速モデルを検討する。平行でない場合は治療群ごとに個別に適切なモデルをあてはめる。
- ✓ 2重対数プロットで直線でなく複雑な場合：

区分モデル、Royston-Parmarモデル等のより柔軟なパラメトリックモデリングを検討する。

Step3

Step2 で当てはめたモデルを AIC 及び BIC（又はその他の適切な指標）を用いて比較する。データが完全に近い（打切りが少ない）場合、内的妥当性を目視確認、AIC 及び BIC、2 重対数プロットを用いて確認し、モデルを選択できるだろう。一方、不完全なデータ（打切りが多い）の場合はさらに外部データ、臨床的妥当性、臨床専門家の判断を活用して、候補となるモデルの外的妥当性を評価する。

Step2 で比例ハザード性の仮定が妥当であり、比例ハザードモデルを用いる場合、ハザード比は共変量として治療群を含めたモデルから推定することになる。外挿部分は治療効果に関して異なるシナリオも考慮すべきである（例えば、外挿部分の治療効果の持続効果を何パターンか設定する）。

Step4

以上の手順に基づき、最も適したモデルを費用対効果評価の基本分析として選択する。モデルの妥当性を評価する際は内的妥当性と外的妥当性を考慮すべきである。また、異なる治療群に別のモデルがより妥当であるという強いエビデンスが無い限りは類似したモデルを使うべきである。もし、複数の尤もらしいモデルがあった場合、候補となるモデルを用いてシナリオ分析を行い、さらにそれぞれのシナリオについて、パラメータに関する不確実性を考慮した確率的感度分析¹¹で行うべきである。このシナリオ分析を行うことで、異なるモデル選択が費用対効果評価に及ぼす影響を評価できる。そして、これは費用対効果評価の意思決定をサポートする情報提供になるはずである。

1.11 生存時間のモデリングのまとめ

本章では生存時間のモデリングの方法とモデリングの適合性の評価方法を紹介した。NICE が提示したアルゴリズムはシンプルであり、透明性を示す上で役立つだろう。外挿部分のモデリングに関しては外部データや臨床家の意見を参考に検討し、不確実性についてシナリオを想定して感度分析を行う等慎重な対応が求められる。本章では詳細な説明や理論を割愛した部分が多いが、必要に応じて、原著で詳細をご確認いただきたい。

¹¹ パラメータに確率分布を当てはめること等により、多数回のシミュレーションを実施し ICER の確率分布を検討する方法。詳細は[10]参照。

2 分割生存時間モデル

分割生存時間モデル (Partitioned Survival Model, PSM) は Area Under Curve 法又は Area Under Curve モデルとも呼ばれる。本モデルは、進行がん又は転移性がんの治療の経済評価に適用されるモデル構造を用いて示される。典型的なモデルには、3つの健康状態 (Progression-free / 無再発, Progressed / 再発, Death / 死亡) が含まれる (図 2-1 参照)。各個人は複数ある健康状態のうちの1つに属すると仮定する。

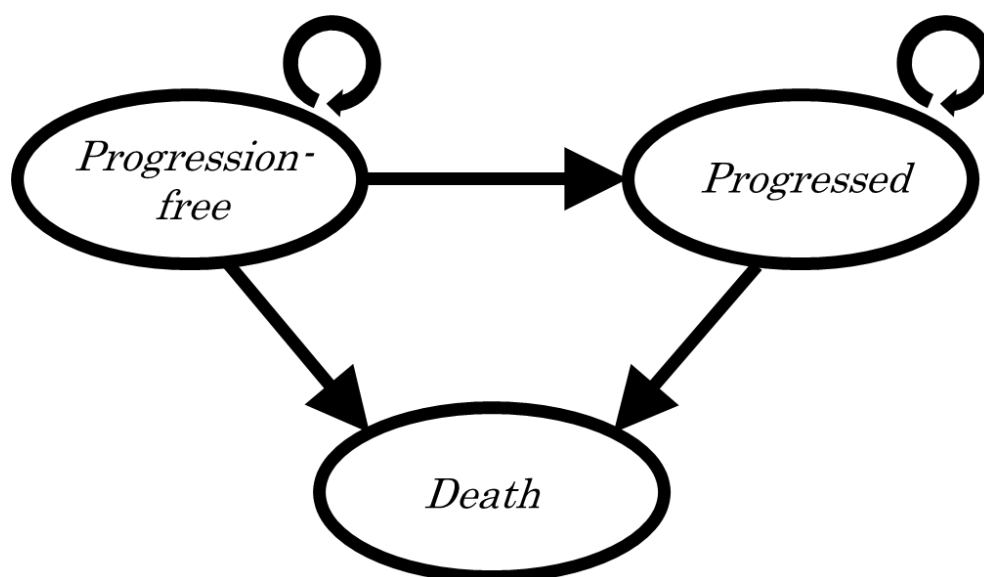


図 2-1 進行性がん又は転移性がんを用いられる典型的なモデル構造

本方法では、各時点の各健康状態の患者の割合、すなわち患者がどの健康状態にあるかを示す状態メンバーシップ (State membership) を、エンドポイントごとにそれぞれ別々にモデル化された生存曲線によって決定する。状態推移モデル (2.4 章参照) では、通常、状態メンバーシップは一定期間内に別の状態に移行する確率をもとに決定するが、分割生存時間モデルでは、状態メンバーシップは生存曲線から決定する。生存曲線はモデルの開始から各健康状態に移行するまでの時間を示すものを使用する。

NICE は 2017 年に分割生存時間モデルに関する Technical Support Document19 (TSD 19) として「Partitioned Survival Analysis as a Decision Modelling Tool」[29]を公開した。本 TSD の目的は、意思決定モデリングツールの 1 つである分割生存時間モデルの説明と批判的吟味であり、様々なステークホルダーがその適切性を判断する際の助けとなる文書である。

2.1 分割生存時間モデルの手順

分割生存時間モデルは推定とモデリングの二段階に分けることができる。前者は実際の臨床試験において PFS, OS の生存曲線を Kaplan-Meier 推定し, 1 章で示した生存時間のモデリングを実施することである。後者はモデリングの結果から, 各健康状態の患者割合を読み取り, 状態メンバーシップ (各状態の患者数の割合) の一覧表 (表 2-4 参照) を作成することである。

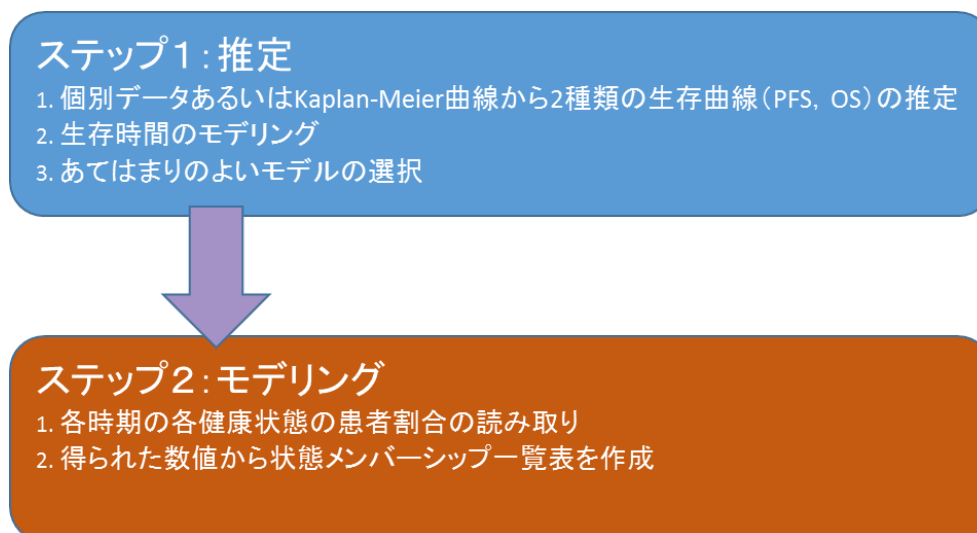


図 2-2 分割生存時間解析における 2 つのステップ

分割生存時間モデルでは, 通常以下の手順で 2 つの生存曲線を用いて状態メンバーシップを推定する。具体的な利用事例は次章で紹介する。

- PFS 曲線より, 各時点の Progression-free (Progressed でも Death でもない状態) の状態メンバーシップが得られる。
- OS 曲線より, 実際には 1 から各時点の OS の割合を引くことで, Death の状態メンバーシップが得られる。
- OS 曲線と PFS 曲線の差により, Progressed の状態メンバーシップが得られる。(図 2-3 参照)

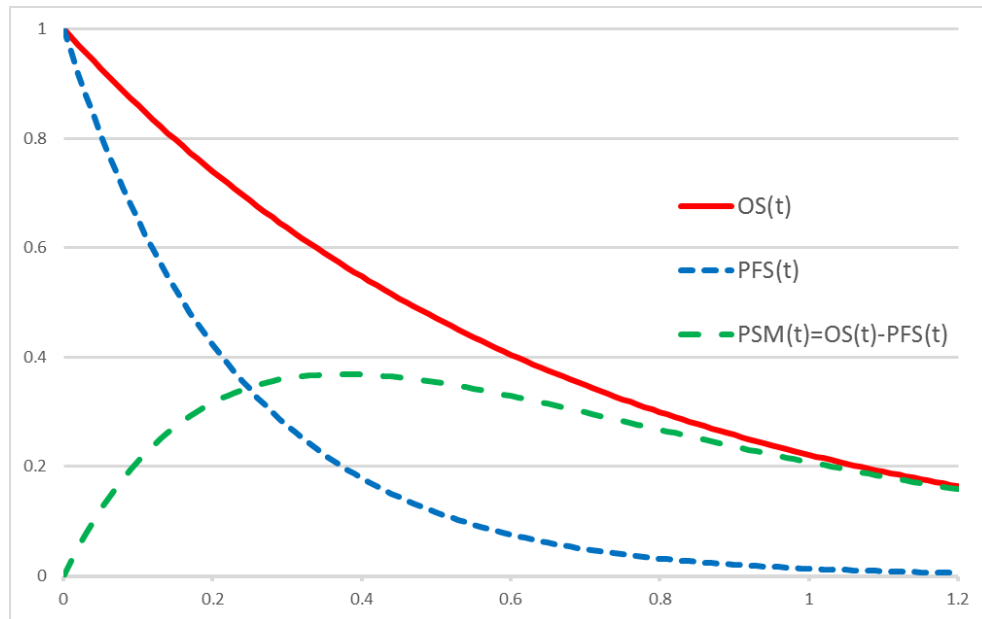


図 2-3 分割生存時間モデルにおける状態メンバーシップの決定

2.2 分割生存時間モデルの適用事例の紹介

NICE で公開されている Technology appraisal guidance TA447 ”Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer”の事例を以下に示す(TA447 は評価の見直しに伴い TA531 として公開されている[30])。

KEYNOTE-024 試験は、未治療の PD-L1 高発現 (TPS $\geq$ 50%) の転移を有する NSCLC 患者を対象にペムブロリズマブ単独投与群と標準治療のプラチナベース化学療法群を比較した試験である。本試験は転移を有する未治療の PD-L1 高発現 (TPS $\geq$ 50%) NSCLC 患者 (305 例) を対象とした試験群：ペムブロリズマブ 200mg 3 週ごと (154 例)、対照群：治験担当医が選択したプラチナベース化学療法 4~6 サイクル (151 例) の無作為化オープンラベル第 III 相国際共同試験である。

KEYNOTE-024 試験[31][32]の PFS, OS, 客観的奏効率 (Objective Response Rate, ORR) の情報, PFS, OS の Kaplan-Meier 曲線を次に示す。

表 2-1 KEYNOTE-024 試験の PFS, OS, ORR の要約ペムブロリズマブ（遺伝子組換え）
（申請資料概要[32] 表 2.5.4-5, 表 2.5.4-6, 表 2.5.4-7 を要約）

エンドポイント		ペムブロリズマブ群 n = 154	化学療法群 n = 151
主要エンドポイント	PFS		
	中央値(月) (95%CI)	10.3 (6.7, -)	6.0 (4.2, 6.2)
	HR (95%CI)	0.50 (0.37, 0.68) P < 0.001	
	イベント数(%)	73 (47.4)	116 (76.8)
	人月	1000.2	785.6
	100人月当たりのイベント発生率	7.3	14.8
	6カ月時点の発生率	62.1%	50.3%
	12カ月時点の発生率	47.7%	15.0%
副次エンドポイント	OS		
	中央値(月)	-	-
	HR (95%CI)	0.60 (0.41, 0.89) P = 0.005	
	イベント数(%)	44 (28.6)	64 (42.4)
	人月	1402	1227.5
	100人月当たりのイベント発生率	3.1	5.2
	6カ月時点の発生率	80.2%	72.4%
	12カ月時点の発生率	69.9%	54.2%
ORR	客観的奏効率(95%CI)	44.8% (36.8, 53.0)	27.8% (20.8, 35.7)
	群間差 (ペムブロリズマブ群 - 化学療法群)	16.6% (6.0, 27.0) P = 0.0011	

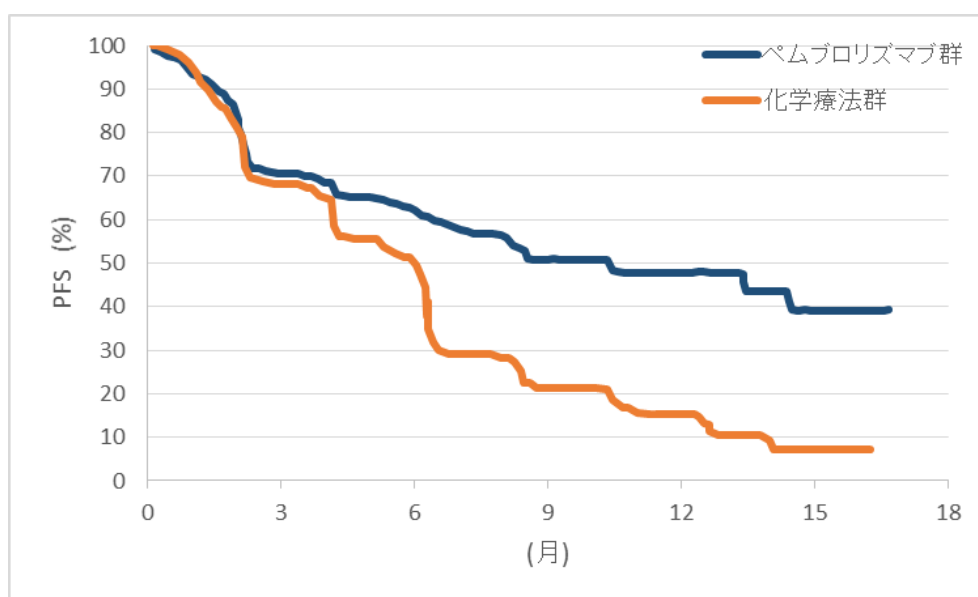


図 2-4 KEYNOTE-024 試験の PFS の Kaplan-Meier 曲線（ITT 集団）
（申請資料概要[32]図 2.5.4-5 を一部改変，打ち切り非表示）

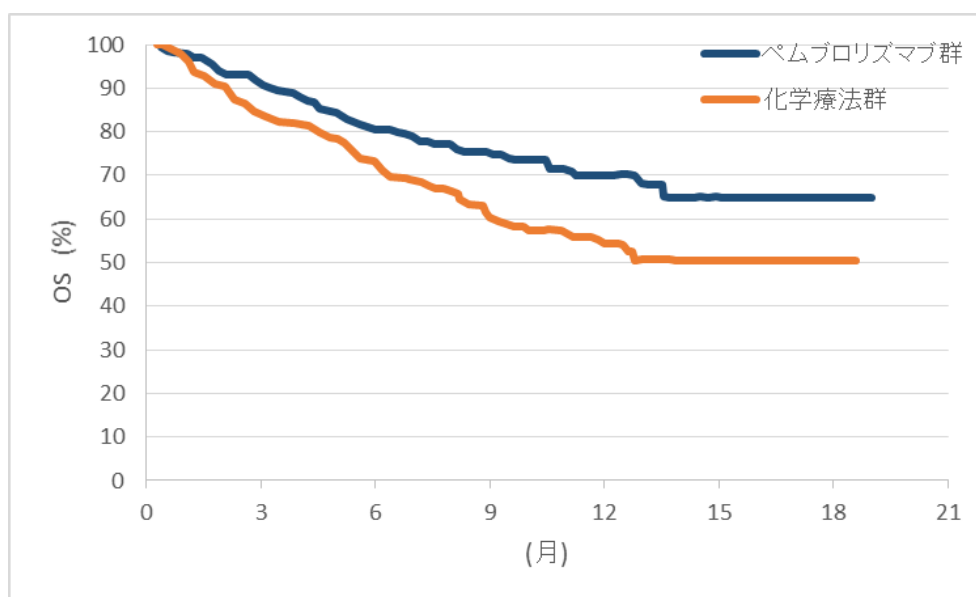


図 2-5 KEYNOTE-024 試験の OS の Kaplan-Meier 曲線 (ITT 集団)
(申請資料概要[32]図 2.5.4-6 を一部改変, 打ち切り非表示)

TSD14 で推奨されている手順 (1 章参照) に従い, PFS に関して 9 週前後での 2 区間の区分モデルを想定した。指数分布, ワイブル分布, 対数正規分布, 対数ロジスティック分布, ギンペルツ分布, 一般化ガンマ分布のそれぞれを用い 9 週後以降の生存時間分布を推定し, 各群であてはまりが良いモデルを AIC, BIC を用い選定した。AIC, BIC は値が小さいほどモデルのあてはまりの良いとされ, ペムブロリズマブ群ではワイブル分布, 化学療法群では BIC から指数分布が選択された。

表 2-2 KEYNOTE-024 試験の各群の PFS の Kaplan-Meier 曲線に対するモデル適合度一覧
(TA447 Committee papers (現在非公開) Table 58 を一部改変)

モデル	ペムブロリズマブ群		化学療法群	
	AIC	BIC	AIC	BIC
指数分布	430.1	432.8	718	720.8
ワイブル分布	424.1	429.6	719.4	724.9
対数正規分布	427.3	432.8	749.2	754.7
対数ロジスティック分布	425.4	430.9	735.1	740.6
ギンペルツ分布	430.1	435.6	719	724.4
一般化ガンマ分布	425.3	433.5	714.4	722.6

AIC : Akaike Information Criteria, BIC : Bayesian Information Criteria

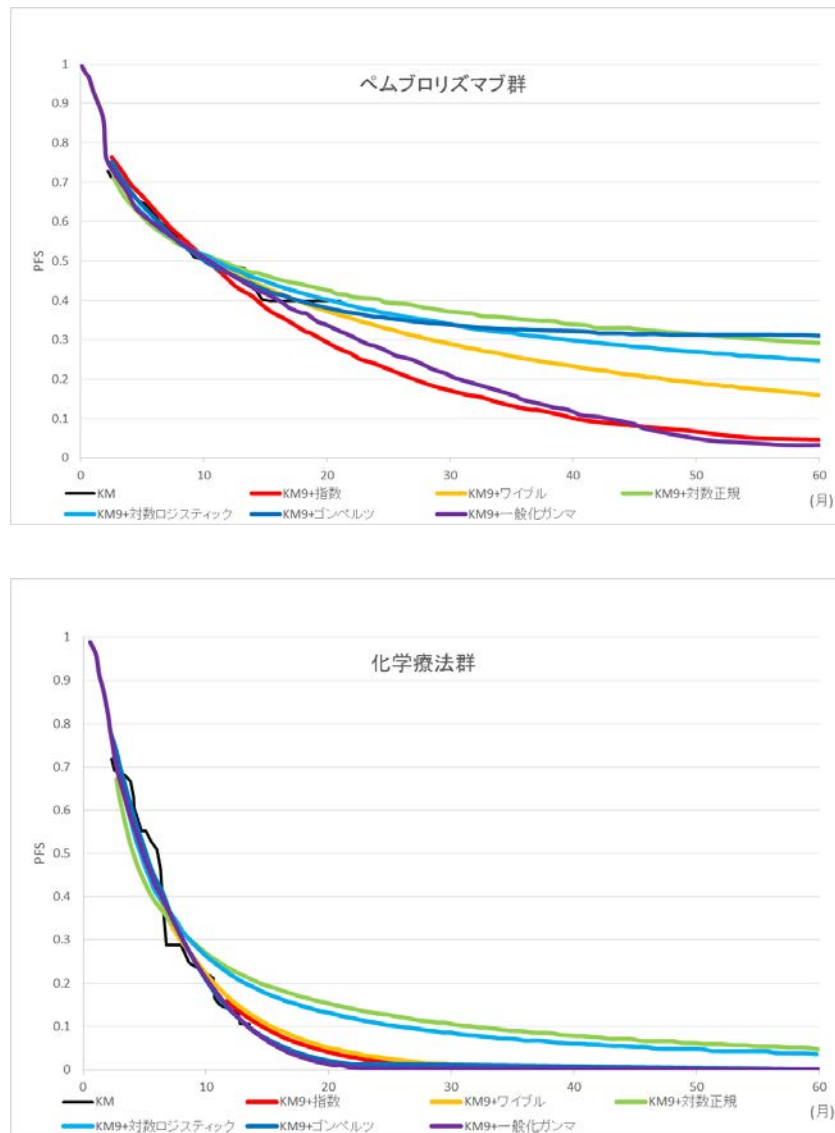


図 2-6 KEYNOTE-024 試験の各群の PFS の Kaplan-Meier 曲線に対するモデル推定
(TA447 Committee papers (現在非公開) Figure 37, 38 を一部改変)

OS に関しては PFS とは異なり，区分に分けずに推定した。ペムブロリズマブ群では指数分布，化学療法群では対数正規分布のあてはまりが最も良いという結果となった。

表 2-3 KEYNOTE-024 試験の各群の OS の Kaplan-Meier 曲線に対するモデル適合度一覧
(TA447 Committee papers (現在非公開) Table 56 を一部改変)

モデル	ペムブロリズマブ群		化学療法群	
	AIC	BIC	AIC	BIC
指数分布	523.9	527	670	673
ワイブル分布	525.6	531.7	671.8	677.8
対数正規分布	525.1	531.1	665.7	671.7
対数ロジスティック分布	525	531	668.5	674.5
ゴンペルツ分布	524.8	530.9	671.2	677.3
一般化ガンマ分布	526.8	535.9	666.7	675.7

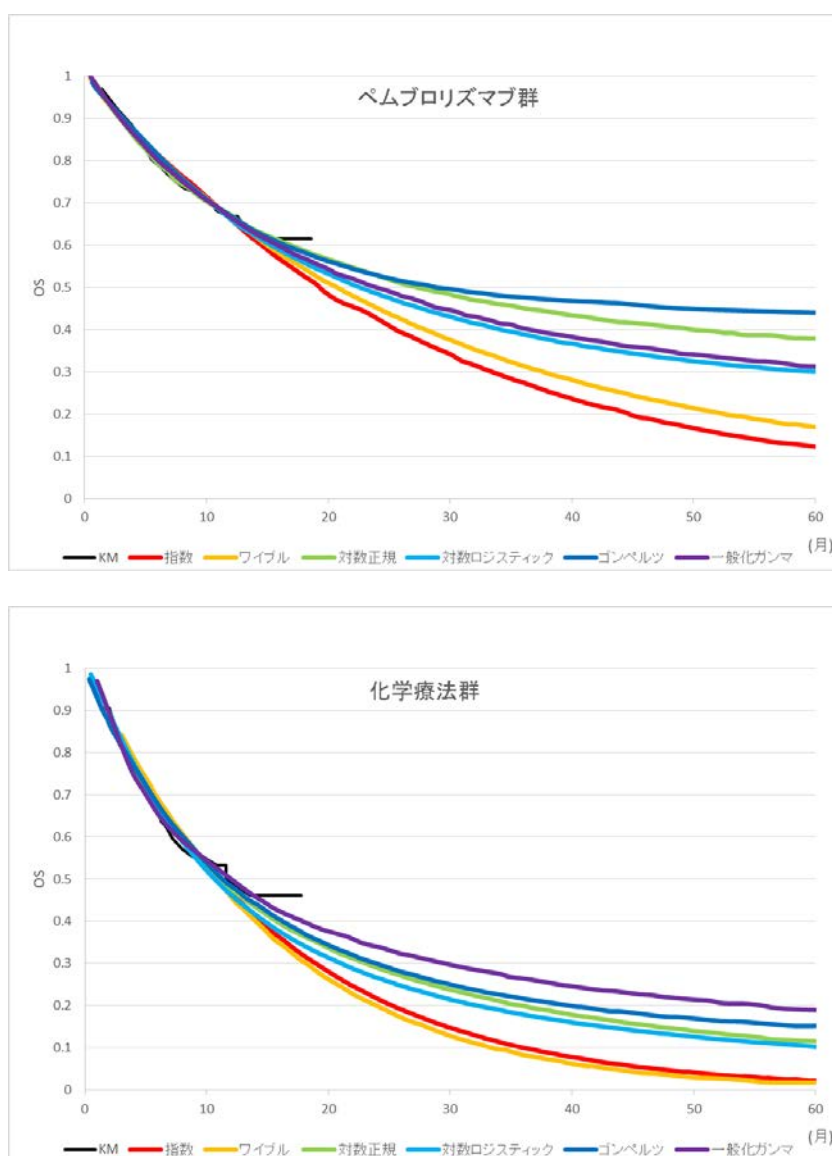


図 2-7 KEYNOTE-024 試験の各群の OS の Kaplan-Meier 曲線に対するモデル推定
(TA447 Committee papers (現在非公開) Figure 31 を一部改変)

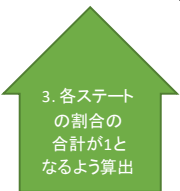
続いて、各群の PFS、OS のモデル推定の結果から、各状態メンバーシップを算出していく。まずは Progression-free の割合を読み取り一覧表に記入する。続けて Death の割合を読み取り一覧表に記入する。最終的に Progressed の割合を算出し、状態メンバーシップの推定が完了する。別途、各状態の費用と効用値を準備し、重み付け平均を算出することで各群の費用と効用値を推定することができる。

表 2-4 KEYNOTE-024 試験の各群の状態メンバーシップ一覧表（PFS、OS の推定モデルより算出）

時間	ペムプロリズマブ群			化学療法群		
	Progression-free	Progressed	Death	Progression-free	Progressed	Death
0	1	0	0	1	0	0
1	0.48	0.22	0.30	0.18	0.32	0.50
2	0.32	0.18	0.50	0.04	0.16	0.80
3	0.25	0.13	0.62	0.01	0.09	0.90
4	0.18	0.10	0.72	0.00	0.05	0.95
5	0.15	0.05	0.80	0.00	0.01	0.99



1. PFSの
モデル推定
から読み取り



3. 各ステート
の割合の
合計が1と
なるよう算出



2. OSの
モデル推定
から読み取り

本事例で取り上げたような薬効発現までに一定の時間がかかるがん免疫療法の分野では、これまでの3つの健康状態を仮定した分割生存時間モデルとは異なる、新たなアプローチも模索されている（例えば Gibson 2017 [33]）。Gibson の論文では通常の分割生存時間モデルに加えて、異なる健康状態を加えた2モデル（無再発の状態における反応の有無、及び疾患の進行の種類を反映）、免疫反応のメカニズムを反映させたマルコフモデルが検討された。その結果、これらの方法がより患者の健康状態を適切に反映できる可能性が示唆された。

2.3 NICE TA での分割生存時間モデルの使用実態

2013年5月から2016年2月の間にNICEによって評価された抗がん剤治療に関する30報のTreatment Assessment (TA) をレビューした。そのうち22報（73%）のTAで分割生存時間モデルが採用され、残り8報では採用されなかった。22報のうち、MTA（Multiple Technology Appraisal）は2報のみでそれ以外はSTAであった。また、進行がん及び転移がんに対するものがほとんどであったが、手術で切除可能な局所進行がんを含むものも一部あった。

- モデリング手法の説明
提出企業が分割生存時間モデルを正しく記載していたのは 22 報のうち、12 報のみであった。残りの 10 報は実際にはマルコフモデルやセミマルコフモデル (2.4 章参照) が用いられていた。ERG (Evidence Review Group) により正しくモデルが用いられていたのは (異なる) 12 報であった。残りの 10 報のうち、4 報はモデルの定義のみが記載され、6 報は ERG により不適切な手法が用いられていた。手法の妥当性については 22 報すべてで記載されていたが、詳細まで適切に説明されたのは 7 報のみであった。(表 2-5 参照)
- モデルの仮定に関する議論
エンドポイント間 (OS と PFS 間等) に相関がないこと (独立性) の仮定が分割生存時間モデルを使用する前提条件である (実際には PFS のイベントの定義には死亡が含まれていることが多い) が、4 報でのみその議論がなされていた。(表 2-5 参照)
- エンドポイント
22 報のうち 17 報では 2 つの生存エンドポイントが使用されていた。ほとんどは PFS 及び OS であり、これらを使用して 3 つの健康状態に分割した。そのうち 2 報では、治療の中止又は死亡までの時間 (Time to Treatment discontinuation, TTD) をモデル化し疾患の進行 (再発) の代わりとした。残る 5 報では 3 つの生存エンドポイント (TTD, PFS, 及び OS) が使用されていた。これらでは最初の治療に費やされた時間を説明するために TTD が用いられており、患者は、治療中 (無再発)、治療中止 (無再発)、再発、死亡の 4 つの状態に分けられた。
- ほとんどの場合において、NMA (Network Meta-Analysis[34][35][36][37][38][39][40][41][42]) 等の間接比較を用いて追加の比較情報を提示していた。
- ERG の懸念は、ほぼ OS のモデル推定による外挿の不確実性についてであった。
- 比較対照群に割り付けられた被験者の病状が悪化した場合には治験薬を用いる等、6 報でスイッチングに対する影響の調整が行われた。(詳細は 3 章を参照)

分割生存時間モデル以外の方法が採用された 8 報のうち、1 報のみ奏効率で層別した (response-stratified) 分割生存時間モデル (2.7 章参照) が用いられ、残りの 7 報は状態推移モデルが用いられた。

表 2-5 分割生存時間モデルが用いられた TA の使用実態の要約

分割生存時間モデルが用いられた TA	n/N (%)
提出企業が正しくモデルを記載	12/22 (55%)
ERGが正しくモデルを記載	12/22 (55%)
企業が手法の妥当性を詳細に説明	7/22 (32%)
企業/ERG間で独立性の仮定を議論	4/22 (18%)

2.4 状態推移モデル

状態推移モデルにおいて、各個人は複数ある健康状態のうちの1つに属する。状態メンバーシップは状態推移の内訳（どのような推移が考えられるか）とその推移確率（スピード）を設定した上で決定される。状態推移モデルを評価する方法として、例えば以下のような方法がある。

- マルコフモデル：状態メンバーシップは、離散的な時間（例：1 か月ごと等）でのコンピュータを用いたコホートシミュレーション¹²を用いて決定される。推移確率はすべて一定のこともあれば、カレンダータイム（すなわちモデルシミュレーション開始時点を0とした場合の時間）のみに依存することもある。
- セミマルコフモデル：状態メンバーシップは、離散的な時間でのコンピュータを用いたコホートシミュレーションを用いて決定され、1 つ以上の推移確率は、中間の健康状態に滞在している時間に依存する。推移確率はカレンダータイムにも依存することもある。
- 患者レベルでのシミュレーション[43]：生存曲線からある状態での滞在時間を算出し、モデルを用いて個々の患者を評価する。イベント発現率はカレンダータイムや、各状態に滞在している時間によって異なる。

状態推移モデルについての詳細は、成書[8][9]等を参照頂きたい。

2.5 過去の文献での分割生存時間モデルと状態推移モデルの比較

Williams ら [44]の研究では、慢性リンパ球性白血病におけるリツキシマブの費用対効果評価において、多状態（multi-state）モデリングを使用して推測された推移確率を用いた状態推移モデルと、分割生存時間モデルのアプローチを比較している。分割生存時間モデルと状態推移モデルの典型的なモデル構造をそれぞれ図 2-8、図 2-9 に示す。

¹² 患者集団（コホート）の分布の推移を追跡するシミュレーション

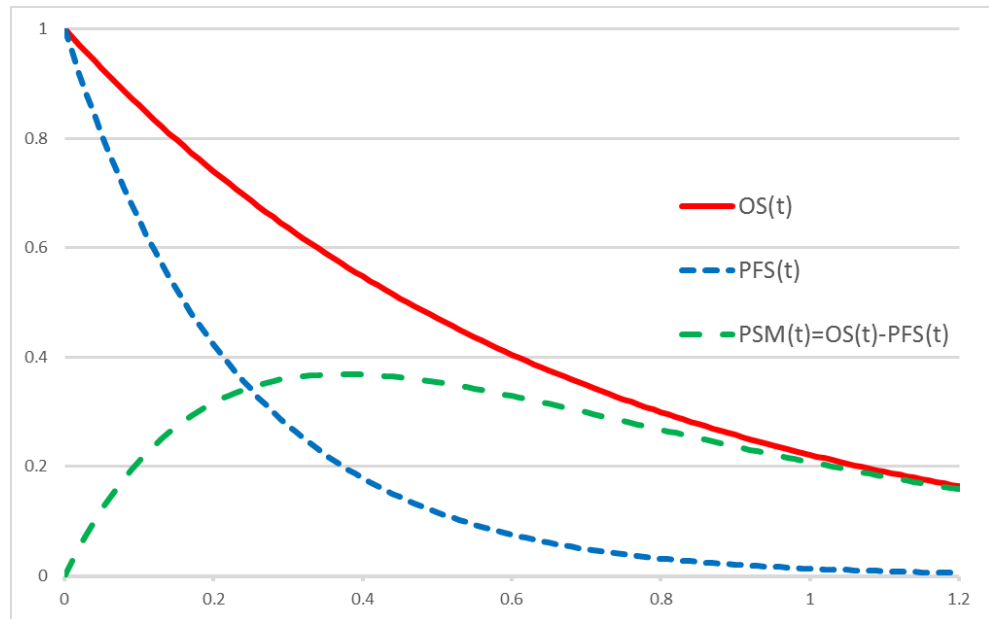


図 2-8 分割生存時間モデルを用いた状態メンバーシップの決定方法
(3つの健康状態を仮定したがんのモデル) (図 2-3 再掲)

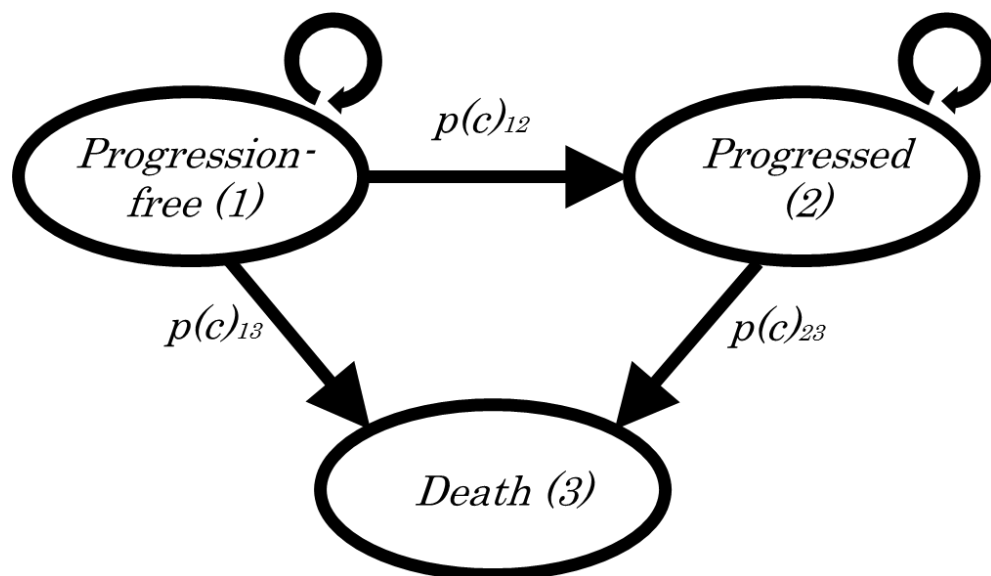


図 2-9 進行性がん及び転移性がんにおいて用いられる典型的な状態推移モデル (c : サイクル)

本研究では、推定対象期間における費用対効果評価の推定結果が解析手法間で異なり、状態推移モデルの QALY (Quality-adjusted life year, 質調整生存年) は、分割生存時間モデルの QALY よりも低く推定

された。その要因は明確でないものの、生存モデルの選択における手法の違いが推定結果に影響を与えている可能性がある。考え得る重要な要因の1つとして、各モデルで用いられた治療効果の推定結果の差異が考えられる。分割生存時間モデルでは RFC 群（試験群：rituximab + fludarabine + cyclophosphamide）の FC 群（対照群：fludarabine + cyclophosphamide）に対する PFS と OS のハザード比（ハザード比は一定と仮定）がそれぞれ 0.6, 0.8 と推定された。一方、状態推移モデルでは、各状態推移（状態 1 から状態 2 への推移、状態 1 から状態 3 への推移、状態 2 から状態 3 への推移）の治療間のハザード比がそれぞれ 0.6, 0.7, 1.4 と推定された。つまり再発後は、RFC 群は FC 群より死亡リスクが高いという推定結果となった。ハザード比 1.4 と推定された PPS（Post-progression survival, progression 後の生存率）は、状態推移モデル内の生存者における状態 2 の割合が高くなるにつれて、OS 曲線に収束していた。分割生存時間モデルでは、OS 及び PFS のハザード比を推定しており、ハザード比が 1 を超えるような結果はなかった。これにより、状態推移モデルで推定された RFC 群の FC 群に対する増分生存年数は分割生存時間モデルで推定されたその約半分となった。

本研究の結果は、解析手法間で著しく異なる結果が推定されることを示唆しているが、特定の手法がより信頼できるかは明確になってはいない。なお、エンドポイントの選択や打ち切りのもたらす影響については言及されていない。

2.6 分割生存時間モデルと状態推移モデルの相違点

以下に、分割生存時間モデルと状態推移モデルの違いをまとめた。

分割生存時間モデルは、意思決定モデルの1つであり、その状態メンバーシップは、エンドポイントごとに別々にモデル化された生存時間曲線によって決定されるという特徴を備えている。したがって、モデル化された臨床イベントに対する関連が明確である状態推移モデルと区別される。データが完全に観察される場合は、エンドポイント間の関係がデータ内に明確に反映されるため、分割生存時間モデルと状態推移モデルは同様の結果が生成されることが期待される。観測されたデータ収集期間を超えてパラメータ推定値を推定する必要がある場合、各モデル（分割生存時間モデル又は状態推移モデル）が異なる方法で外挿するため、異なる結果が生成される可能性が高い。特に、分割生存時間モデルを用いて外挿された死亡率は、外挿以前の死亡率の傾向によって決定されるが、状態推移モデルを用いて外挿された死亡率は、死亡率と早期の疾患関連イベント（がんの進行等）の構造的な関係によって決定される。また、疾患のプロセスに対する治療効果の推定方法もアプローチにより異なっている。分割生存時間モデルでは治療効果は生存曲線（PFS, OS 等）に直接影響するが、状態推移モデルでは治療効果は疾患プロセスのさまざまな部分に影響する可能性があり、これらの効果の相互的な影響によって推定生存曲線に対し全体的に影響を与える。分割生存時間モデルでは通常報告されている生存エンドポイント（PFS, OS 等）を使用して直接パラメータ化が可能であるが、状態推移モデルでは個々の推移確率の推定値が必要となる。これらの要約を表 2-6 に示す。

表 2-6 分割生存時間モデルと状態推移モデルの特徴

項目	分割生存時間モデル	状態推移モデル
状態メンバーシップの推定方法	各健康状態に対応する生存曲線から、状態メンバーシップを推定する。	疾患の進行に対応する推移確率を用いて、集団単位又は患者レベルでのシミュレーションを行い、状態メンバーシップを推定する。各サイクル終了時の状態メンバーシップは、前のサイクルの状態メンバーシップに推移確率を適用することで推定する。これを分析期間が終了するまで繰り返す。
データの入力	PFS や OS 等の通常報告される臨床エンドポイントの生存時間データを要約データあるいは IPD から得られる。IPD が得られない場合でも対応できることがメリットである。	個人レベルでの推移がわかる生存時間データが得られる場合、多状態生存時間解析を用いて推移確率を推定できる。検証試験の場合でのみ IPD が得られることが多い。間接比較や長期アウトカムに用いられる外挿データは要約データとしてしか得られないことが多い。IPD が得られない場合は妥当性の確認が難しくなる。
イベントの時間依存性を反映する手法	生存曲線の推定の際に直接反映することができる。	トンネル状態モデル ¹³ 、セミマルコフモデル、患者レベルでのシミュレーション、ペイオフアプローチ ¹⁴ 等を用いる必要がある。
OS の外挿方法	OS の経時的な傾向から推定する。	死亡率と健康状態間の推移確率（スピード）から推定する。

¹³ 罹患の状態に「トンネル状態」を設定し、罹患年数ごとに費用や効用値を設定する方法

¹⁴ 再発後の費用と QALY を、再発後生存期間の曲線化面積、効用値、割引等をもとに一定値に算定、設定し、再発した患者は一律に再発時点の後はその費用と QALY を得られるとみなし、再発時点で「ペイオフ（清算）」する方法。

項目	分割生存時間モデル	状態推移モデル
OS における治療効果の外挿方法	OS における治療効果の経時的な傾向から推定する。各治療での PFS と OS の生存曲線が必要となる。	個々の推移への治療効果の影響から OS としての治療効果が全体的に推定される。そのため個々の推移確率に対する生存時間データ及び治療効果の推定が必要である。
OS の外挿の妥当性に関する懸念点	死亡率と（死亡率に対する）治療効果は外挿部分には適切に反映できない可能性がある。	PPS 等の中間の健康状態の推定は、現段階では、信頼性が担保されない可能性がある。健康状態と長期間でのイベントの傾向の間には明らかな関連があるが、個々の推移確率の傾向や治療効果は外挿部分に反映することが難しい。
意思決定プロセスに使用する際の留意点	外挿の尤もらしさを評価することや、意味を持つ感度分析を行うことが比較的難しい。	外挿の尤もらしさを評価することや、意味を持つ感度分析を行うことが比較的容易である。イベント発現率や治療効果の個々の要因が特定可能であるため、メカニズムが理解しやすいことに起因する。

2.7 分割生存時間モデル適用時の利点と留意点

分割生存時間モデルでは、PFS や OS 等の通常報告される生存エンドポイントから状態メンバーシップの推定値が直接得られることにより、直感的で容易にモデルが構築できる。すなわち、これらのエンドポイントの要約データ又は IPD のいずれかを使用することで、分割生存時間モデルが構築できることを意味する。

分割生存時間モデルのアプローチの制約は、生存エンドポイントが別々に推定されているという基本的な仮定から生じている。一般に特定のエンドポイントの外挿は、そのエンドポイントの試験内のみの傾向を反映している。単純なシナリオを使用すると、ベースラインリスク（プラセボのハザード）と治療効果の試験内の傾向を用いた外挿は、疾患の自然な進行を考慮できずに、適切な外挿ができない可能性がある。いくつかの研究では、より適切な外挿を目指した統計モデルを開発することによって、この制約を克服することを試みている。ただし、提案された方法は意思決定の主観的な側面の割合を増やすだけとなっている。エンドポイント間（PFS 及び OS 等）の関連性の欠如は、分割生存時間モデルによる外挿の生物学的及び臨床的妥当性を感度分析で確認する可能性も減少させている。また、エンドポイント間の独立性の仮定は、意思決定の不確実性を定量化する手段としての分割生存時間モデルの価値を低下させている。

以下に、分割生存時間モデルを適用する際の利点と留意点をまとめた。

表 2-7 分割生存時間モデル適用時の利点と留意点

利点	留意点
• PFS と OS の曲線下面積から状態メンバーシップを直接推定可能。 • IPD が入手できなくても患者情報の要約データがあれば実施可能。	• 分割生存時間モデルでは生存エンドポイントが別々に推定されていることを仮定している。そのため、エンドポイントの推定結果は、その試験での傾向だけが反映されたものになりやすい。 • 疾患の自然な進行を考えずに、ベースラインのリスクや治療効果等のその試験の傾向が強く反映された推定結果は不適切な結果を導きやすい。 • 近年、新たな外挿の統計モデルを構築することで、上記の制約を克服しようとしているものの、この方法は主観が入りやすい。 • 分割生存時間モデルでの外挿の妥当性を確認することが感度分析の目的となる。

代替アプローチとしては、TA311[45]で用いられた奏効率で層別した分割生存時間モデルのようなハイブリッドモデリングがある。この手法は決定樹モデルを用いた後、分割生存時間モデルを用いる手法である。当然、状態推移モデルも候補となるが、データの利用可能性、推移確率の推定値の一般化可能性、時間依存性の推移確率の実装等の重要な問題が存在する。

今後、分割生存時間モデルが適用されている状況に対し状態推移モデルを実際に適用することに焦点を当てた研究が必要となるであろう。その際、モデル化された事象と競合している因子、慎重なモデル選択、適切な統計モデル化等を考慮し推移確率を推定することが重要である。また、選択した効果と独立していない打ち切りによるバイアスも考慮する必要がある。さらに、要約データだけが利用できる場合に個々の推移確率を推定することには多くの課題が残る。分割生存時間モデルと状態推移モデルのそれぞれが適切である条件の理解が必要である。

2.8 分割生存時間モデルの推奨使用方法のまとめ

分割生存時間モデルの推奨使用方法の要約を表 2-8 に示す。TSD19 で列挙された 13 種類の推奨使用方法を次章以降に示す。

表 2-8 分割生存時間モデルの推奨使用方法の要約

- モデルの概念化プロセスの重要性を強調すること
- ステークホルダーが分割生存時間モデルの制約を認識すること
- 分割生存時間モデルと並行して個々の臨床イベントに対応する生存曲線が要求される
- 多状態生存時間解析の使用をサポートするためのガイダンスとさらなる研究が必要
- 分割生存時間モデルと状態推移モデルを並行して利用することを推奨
 - 分割生存時間モデルの外挿の妥当性の検討に役立つ
 - 外挿部分のさらなる不確実性に対処可能
 - IPD が利用可能な検証試験においても対応するべきである

2.8.1 モデリング手法の選択、記述、正当化

推奨 1

モデルの概念化プロセスはルール通りに報告され、選択されたモデリング手法の理論的根拠は、理論的及び実践的考察に基づいて明示的に正当化されるべきである。

モデルの概念化プロセスの概要は、ルール通りに報告する必要がある。選択したモデリング手法と構造をサポートする理論的及び実用的な留意事項を記載しなくてはならない。モデリングのアプローチと構造を選択する理由は、疾患 / 医療技術の主な特徴と整合性が取れている必要がある。これらには、健康関連のイベント、イベントのリスクの有無とその変化の程度、治療効果の大きさ等が含まれる。外挿の必要性、及び費用対効果を判断する際の外挿の重要性を考えると、モデリング手法による外挿の適切性を示さなくてはならない。モデリング手法の選択は、利用可能なエビデンスによって制約される場合がある。この制約を提示し明確にする必要がある。分割生存時間モデルは直感的で実装が容易であり、試験で用いられる生存エンドポイントから予測可能である。ただし、これは費用対効果に関する意思決定のための適切なモデリング手法として分割生存時間モデルが正当化されるものではない。モデル化されたエンドポイント間には構造的な関係がないため、分割生存時間モデルが外挿を正当化する理想的な

モデリング手法ではない可能性があることを認識することが重要である。

推奨 2

アプレイザル評価において、分割生存時間モデルを説明する際に、一貫性のある用語が使用されるべきである（例えば、用語「分割生存時間モデル」の使用）。

分割生存時間モデルは不正確な説明がされてしまうことが多い。分割生存時間モデルの用語とそれに関連する説明は、状態推移モデル（マルコフモデルやセミマルコフモデル等）とより明確に区別して説明する必要がある。特に「分割生存時間モデルはマルコフモデルに似ている」や「マルコフモデルのような」といった説明は避けるべきである。

推奨 3

主な構造上の仮定を要約する際は、NICE 手法ガイド[46]の要求通りに説明され正当化されるべきである。

分割生存時間モデルの場合、以下の構造上の仮定の明確な説明と正当化を含める必要がある。

- (i) OS を含むすべてのエンドポイントは、別々にモデル化され推定される（独立性）。
- (ii) 各エンドポイントの試験期間内のハザードと治療効果（ハザード比）の傾向は、外挿部分においても一般化可能であると仮定される。状態推移モデルでは、この一般化の仮定は、可能な推移すべて、治療間で差異が生じる推移、患者の病歴のベースライン後の推移への影響等にも及ぶ。

2.8.2 外挿に関連する不確実性の表現方法

分割生存時間モデルのエンドポイント間の構造上の独立性とは、エンドポイントの外挿がそのエンドポイントの試験期間内の傾向のみを反映していることを意味している。疾患の自然な進行を考慮せずに試験期間内の傾向からのみ外挿すると、適切な外挿結果が得られない可能性がある。

推奨 4

すべてのステークホルダーは、分割生存時間モデルの外挿の具体的な制約を認識すべきである。

外挿のすべての方法は、分割生存時間モデルの使用、状態推移モデルの使用のいずれの場合でも、不

確実性の影響を受ける。ただし、分割生存時間モデルで中間の健康状態（Progressed 等）の情報を注意深く扱わないと、分割生存時間モデルによる外挿の不確実性の程度が高まる可能性がある。また、分割生存時間モデルで外挿時に疾病の自然な進行の程度を無視することは、その信頼性の評価可能性を制限する。さらに、分割生存時間モデルのフレームワークのいくつかのシナリオを探索する（異なるパラメトリック生存関数の使用や、外挿部分に対し治療グループ間で同じハザードを適用すること等）のは容易ではあるが、治療効果と特定の健康状態との関係等を探索することは比較的困難である。

推奨 5

結果に大きく影響する外挿部分に対するモデルの選択には、関連するすべてのエビデンスを反映すべきである。

外挿部分に対するモデルの仕様を説明する際には、関連するすべてのエビデンスを考察して説明する必要がある。この際、統計的考察だけに基づいて説明してはならない。関連するエビデンスには、検証試験以外のデータと、検証試験のよる詳細なデータが含まれる場合がある（推奨 6 を参照）。外部データを使用し、パラメトリック生存時間モデルにより、外挿部分に対する妥当性を評価することができる。これらのデータから、未治療の患者又はすでに確立された治療がなされた患者の結果に関する情報が得られることもある。比較的新しい治療法の長期的な結果の妥当性を評価するために、これまで得られたデータが利用できる可能性は低く、試験期間を超えた治療効果の推定されたモデルの軌跡（外挿部分）の妥当性の評価は専門家の推論と判断に大きく依存するだろう。

推奨 6

各試験の個々の臨床イベントに対応する生存曲線は、分割生存時間モデルの結果と同時に提供されるべきである。

通常、検証試験の PFS や OS 等の一般的に報告されるエンドポイントの生存曲線が提出される。同時に、経時的な傾向に変化がないかを確認できるハザード関数のプロットが、生存率と治療効果それぞれに関し添付されるべきである。これにより健康状態メンバーシップが変化するにつれて、ハザードと治療効果が変化するかどうかを確認できる。さらに追加情報として、個々の臨床イベントの生存曲線を生成して提供する必要がある。例えば、3 つの健康状態（無再発、再発、死亡）のモデルの場合、死亡を打ち切りとした再発までの時間、死亡以前の再発を打ち切りとした死亡までの時間と PPS の情報を提供する必要がある。これらの情報は、治療ごと時点ごとにリスクセット（観察対象者数）の情報も含んだ Kaplan-Meier 曲線の形式で提供する必要がある。これらのデータを解釈する際に考慮する事項は推奨 7 で説明されている。

推奨 7

費用対効果の情報を適切に提供するために外挿が必要となる場合における、現行の NICE 手法ガイドに従った異なる仮定を用いた検討は、分割生存時間モデルによって容易となる。

この検討には、外挿部分におけるベースラインリスクと治療効果の不確実性の検討が含まれるべきである。外挿部分のシナリオには、検証試験外のデータや試験に含まれる個々の臨床イベントに関するデータを含むすべての関連するデータに基づき妥当と判断される結果の範囲を反映すべきである（推奨 5 及び 6 を参照）。さらに治療効果については、NICE 手法ガイドで提案されている外挿部分の治療ベネフィットに関するシナリオを用い検討する必要がある。すなわち、(1)外挿無し、(2)治療期間と同等と仮定、(3)長期部分は減衰すると仮定、の 3 パターンの検討である。特定の外挿の仮定における費用対効果に関する結論の信頼性は、すべてのステークホルダーによって明確に承認されるべきである。

2.8.3 モデリングにおける代替手段の使用

分割生存時間モデルにおけるエンドポイント間の構造的な関連性の欠如は、不適切な外挿の可能性を高めてしまい、外挿の基本的なメカニズムを理解することや、その臨床的及び生物学的な妥当性を評価することを難しくする可能性がある。さらに分割生存時間モデルのフレームワークでは、イベント発現率と特定の健康状態の関係、及び治療効果と特定の健康状態の関係を明らかにする有益な代替シナリオを検討することは困難である。中間の健康状態が治療効果の変化に関連することが知られており、外挿が必要な状況下では状態推移モデルを用いることがこれらの問題を解決し、分割生存時間モデルよりも理論的及び実用的な利点をもたらす得る。

状態推移モデルはこれまで適切に実装されることがあまりなかった。強い仮定が確認されることなく、観察されたデータを正確に表現できず、信頼性の高い外挿ができなかった。ただし、より頑健な方法も近年提案されつつある。

推奨 8

状態推移モデルでの推移確率は、適切な統計的方法を用い関連するすべてのエビデンスを反映して推定されなければならない。

状態推移モデルでは、仮定の正当性と評価が明確になされた推移確率を、確立された多状態モデルの手法を用いて推定し、外挿部分のパラメータ推定結果の一般化可能性を慎重に考察するべきである。選択バイアスと情報を持つ打ち切りによって生じるバイアスを考慮して、推移確率の推定の際には何らかの調整が発生する場合がある。状態推移モデルによって作成された予測モデルは、試験期間内に観測されたエンドポイントのデータに対して検証されなくてはならない。さらに外部データが個々の推移のモ

デルの仕様をどの程度サポートできるかについて考察する必要がある。状態推移モデルでは、要約データに基づく間接比較の結果を組み込む方法論とその事例が少ないことから、このような解析は慎重に検討されるべきであり、必須の仮定は明確にされるべきである。ERG が状態推移モデルを検証するために、追加の要約データ並びに IPD へのアクセスが要求される場合がある。統計モデリングに加えて、各推移に対応する Kaplan-Meier 曲線も提示する必要がある（推奨 6 を参照）。

推奨 9

多状態生存時間解析を用いた状態推移モデルの適切な仕様の記載をサポートするためには、さらなる研究とガイダンスが必要である。

状態推移モデルに対して複数の状態をイベントとみなした多状態生存時間解析の使用の実現可能性については最近の研究で議論されているが、要求される生存時間モデルの仕様をどのように記載するかについて根本的な論点がいくつか残っている。第一の論点は、打ち切りが存在する場合、どのように投与開始時以降の推移確率と治療効果のロバストな推定値が導出できるかである。投与開始時以降の推移（PPS 等）が OS の外挿の主要な要因となりやすいため、このことは重大な問題である。第二の論点は、適切なあてはまりが期待されるモデルのエンドポイント（OS 等）が複数の生存分布からなる関数で表現される場合、モデルのあてはまりをどのように評価するかである。第三の論点は、信頼できるモデル構造の作成をサポートする、特に予後及び治療効果の変化に対する信頼できる指標となる中間エンドポイントの選択をサポートするような、さらなる研究が必要であるということである。

推奨 10

検証試験外のデータ、特に間接比較に用いたデータを状態推移モデルへ組み込むことをサポートするには、さらなる研究が必要である。

このような研究は、検証試験では IPD が利用できるが、比較対照の試験は主要エンドポイント（PFS、OS 等）の要約された生存曲線の形でのみデータが利用できるような、TA 内で典型的に起こりうる状況に焦点を当てるべきである。

推奨 11

分割生存時間モデルの外挿の妥当性の検証と外挿部分の不確実性への対処のために、状態推移モデルは分割生存時間モデルと同時に使用されなければならない。これは IPD が利用可能な検証試験の妥当性を示す場合においても同じである。

さらなる研究が必要であることから、分割生存時間モデルを状態推移モデルに完全に置き換えることは推奨できない（推奨 9 及び 10 を参照）。この話題に関し今後研究が進んだ場合には、本推奨は再検討

する必要があるだろう。健康状態メンバーシップの変化が明示的にモデル化されるため、外挿に関しては分割生存時間モデルよりもいくつかの利点がある一方で、試験期間中の個々の推移確率の傾向が試験終了以降の傾向にも反映されることが外挿の妥当性と依存関係にあることに、ステークホルダーは注意すべきである。

推奨 12

分割生存時間モデルと状態推移モデルを用いた結果の説明には、生存年（Life Year, LY）と QALY に治療間で差異が生じている状態を提示し、その差異の正当性を説明しなくてはならない。

NICE の STA のエビデンス提出テンプレートでは、これらの情報をまとめることが要求されている。すなわち、観測された差異を生成するモデルのメカニズムの説明と、入手可能なエビデンスに基づいた差異の正当性のいずれの情報も提示されなくてはならない。

推奨 13

分割生存時間モデル及び状態推移モデルの適用に関連して起こりうるバイアスの特定と、そのバイアスが手法の違いによってどの程度変化するかについては、さらなる研究が必要である。

イベントのハザード（リスク）が時間とともにどのように変化するか、それらが中間の状態及び患者の病歴とどの程度関連するか、すなわち治療がイベント発現率に影響しているか、打ち切りの程度等、さまざまな異なるシナリオを検討するシミュレーションが研究には含まれる。また、より柔軟なパラメトリック生存時間モデルを使用した分割生存時間モデルを実装することが、外挿の信頼性に影響を与え得るかについての研究もなされるべきであろう。

3 スイッチングとはなにか？

（治療）スイッチングとは、ランダム化比較試験で対照群に割り付けられた患者が、治療を対照治療から新治療に切り替えることである。クロスオーバーと呼ばれることもある。スイッチングが存在する場合、ITT 解析で推定された対照群の効果に新治療の効果が混在し、バイアスが含まれることになる。費用対効果評価を行うにあたり、対照治療に対する新治療のベネフィットをより正しく示すためスイッチングの調整が必要になる。

スイッチングは、倫理及び実務上の理由から発生する。倫理的には、他に有効な治療が存在しない場合において、新治療の有効性が中間解析により示されているにもかかわらず、対照群の患者が新治療を使用することを許容しないことは難しいということである。一方実務上の理由としては、スイッチングを許容しないデザインの場合患者リクルートが困難になるということがある。

抗がん剤等のランダム化比較試験において、無増悪期間（PFS）を主要エンドポイントとすることが FDA や EMA により認められるため、臨床試験は全生存（OS）ではなく PFS に基づく検出力を担保するための例数で計画され、またランダム化された群を増悪後も維持することは、薬事承認の観点から大きなメリットはない。上市にあたり OS による優越性を示す価値が大きい一方、費用対効果評価の観点からは、生涯にわたる評価を求められることが多く[47][48][49][50]、OS を正しく推定することは費用対効果におけるチャレンジの 1 つである。

当 TSD ではスイッチングに伴い発生する生存時間の推定値へのバイアスの調整法に焦点を当てている。その理由としては、スイッチングは生存時間だけではなく、QOL や費用にも影響を与えるが、QOL は多くの場合治療群ではなく健康状態ごとに対し定義されること、費用は外部データから推定されることが多いことから、これらに対する影響は比較的大きくはないためである。

また、どのようなケースにどのような調整方法を行うかを明確に示すことは困難であるため、当 TSD では指針を示すのみにとどめている。どの調整法がバイアスを最小化するかは、様々な因子（例えばスイッチングの割合、症例数、得られたデータ、標本抽出の方法、治療効果の大きさ）及びそれらの交互作用により変わってくるからである。

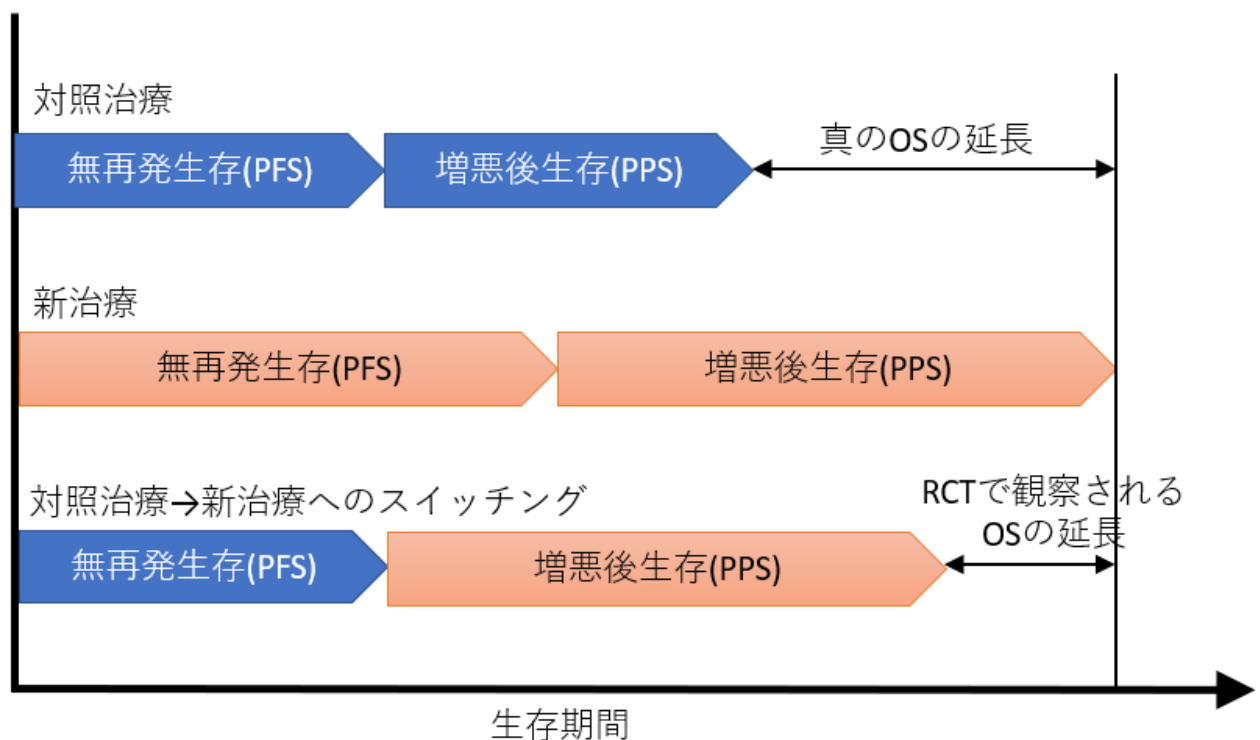
3.1 スイッチングがもたらす問題

医療経済家や意思決定者にとってスイッチングが問題となる理由の本質は、臨床試験で行われた治療が、費用対効果評価が評価する decision problem(検討課題)と乖離してしまうことにある。

当 TSD におけるバイアスとは、推定された治療効果と、スイッチングが存在しなかった場合に観測されたであろう効果の差と定義する。図 3-1 に、スイッチングによるバイアス発生の様子を示す。Control Treatment（理想の対照群）と Intervention（新薬群）は、本来臨床試験において観察することが期待されている PFS、増悪後生存（PPS）であり、PFS、PPS の合計（OS）の差が True OS difference（真の OS の

差)である。一方, Control→Intervention (スイッチの存在する対照群)が, スイッチングの存在のもと観測される対照群を示す。PFS は Control Treatment と同じであるが, PD 後一部の患者でスイッチングが発生する Control→Intervention の PPS は, 新治療の効果が含まれるため Control Treatment のそれより長い。スイッチングの存在下で ITT 解析によって観測される OS の群間差が RCT OS difference で示され, 真の OS difference よりも過小評価されていることがわかる。このバイアスにより, 費用対効果の結果は新治療に不利 (増分費用効果比 (ICER) が大きくなる) な結果をもたらす。

図 3-1 新の OS の延長とスイッチング存在下において観察される OS 延長



NICE の費用対効果評価ガイドライン^{エラー! ブックマークが定義されていません。}にも, スイッチングに関する叙述があり, 以下のように言及されている。

- スイッチングの存在下では ITT 解析が不適切となるため統計的手法による調整が必要になる。
- スイッチングのあった患者を打ち切りや解析から除外するようなシンプルな調整方法は選択バイアスの影響を受けやすいため避けるべきである。
- スイッチングの影響を評価するために, それぞれの手法の特性 (メリット及び限界) と分析するデータの特性を考慮して検討, 正当化すること。
- 前述の「データの特性」とは, 当該の臨床試験におけるスイッチングの発生メカニズム, ベースラ

インデータ及び時間依存性共変量の利用可能性、割り付けられた治療を継続した場合に期待される治療効果を含む。

3.2 スイッチングの調整法(シンプルな方法)[51]

本節ではスイッチングの調整法として使用されうるシンプルな方法について触れるが、使用できる状況は限定的である。それぞれの方法についてどのような問題点があるかを示し、スイッチングの調整に関する理解を深めていただきたい。

3.2.1 ITT 解析

ITT 解析は本来スイッチングの調整を目的としたものではないが、ランダム化比較試験(RCTs)の標準的な解析方法である。スイッチングの存在下においてより有効性の高い新治療のもたらす“真の”生存への効果の推定値は、対照群の患者の一部が新治療に移行してくることにより薄まってしまうので、費用対効果評価には不適である場合が多い。

3.2.2 Per protocol analysis(スイッチングのある患者を除外/打ち切りとする)

スイッチングのある患者を解析対象集団から除外、あるいはスイッチング時点での打ち切りと扱う方法である。このような方法では、情報のある打ち切り/除外による選択バイアスが発生する。

3.2.3 スイッチング後の費用を含める方法

ITT 解析による方法は、対照群からのスイッチングがあった場合、スイッチング後の有効性に関しては新治療の効果を享受する一方で、費用に関してはスイッチングの無かった対照群のものを採用するため、スイッチングのあったケースは新治療の効果を対照治療の費用で享受する形となる。一方でこの方法は有効性に関しては ITT 解析と同様であるが、スイッチング後の費用を解析に含める方法である(費用に関しても ITT 解析の考え方を導入している)。RCT において発生したことを忠実に反映させた費用対効果評価を行っていることになる。一方でこの方法が HTA において有用かどうかは、評価したい decision problem によるが、多くの場合適さないだろう。なぜなら多くの費用対効果評価の目的は、「新治療が使用可能である状態」と「新治療が使用できない状態(比較対照)」の比較であるが、この方法によれば比較対照が「従来の治療を使用しており、増悪後に新治療に切り替える」となるためである。スイッチングを行う患者は診断により選択されると考えられたため、他の患者と異なった集団である。よってスイッチングにより新治療を投与された患者における治療効果は、新治療群に割り付けられた患者で観測される真の治療効果よりも劣るであろう。前者にかかった新治療分のコストは真の治療効果にはつながらないということになるので、これを分析に含めても正しい費用対効果評価であるとは言えない。

3.2.4 PFS のみに基づくモデリング

疾患増悪後にのみスイッチングが許容される研究の場合、PFS のデータはスイッチングによる影響を受けない。よって、PFS のみを用い、疾患増悪後の生存期間に関しては除外して評価する方法が考えられる。しかしながら、この方法は新治療が PFS のみに影響を与えるとみなしていることにはならない。よってこの方法は疾患増悪後の生存期間の延長に関し考慮しておらず、おそらく新治療の費用対効果を過小に見積もるため、企業の立場から保守的な解析となる。しかしながら、必ずしも過小評価になっているとは限らない。なぜならこの方法は疾患増悪後の新治療の効果のみではなく、疾患増悪後にかかる費用も無視するためである。OS に対する新治療の効果は PFS に対するそれより薄まるであろうから、疾患増悪後の効果と費用を共に無視した場合に新治療の費用対効果が過大評価されるか過小評価されるかは一概に言えない。NICE のガイドラインエラー! ブックマークが定義されていません。においても、生存期間に影響を与える治療の場合は生涯にわたる評価を求めているため、この方法は適していないだろう。

3.2.5 疾患増悪後の死亡リスクを両群同一とみなす解析

この方法は前述の PFS のみに基づく方法の拡張である。疾患増悪後の死亡リスクを両群で変わらないとして、OS をモデリングする。

この方法は、新治療の追加的効果が疾患増悪までしか影響を及ぼさないとみなしていることに他ならない。この仮定は、柔軟に変更可能である。新治療は PFS を延長するが OS を延長しないかもしれないし、新治療の効果が疾患増悪後にも維持され、PFS を延長した以上に OS を延長するかもしれない。あるいは治療効果がある時点までのみ維持される、という仮定もあり得る。このような仮定は任意に変更可能であり、臨床データ、専門家の医学的見地、治療や疾患の生物学的特性に基づく必要がある。これらの情報による裏付けをもってもなお、この方法で用いられる仮定は良く議論されるべきである。

3.2.6 両群で同一の OS を設定する方法

前述のとおり、新治療は PFS を延長するが OS を延長しないかもしれない、という仮定も可能であり、この仮定による解析は新治療の“worst-case”な解析であるといえ、スイッチングの調整を行う他の方法と比べた際に ICER が最大の値を示すであろう。この方法による ICER であってもなお意思決定者が受け入れ可能な金額であった場合、新治療の受け入れ可能性は高いと言える。

3.2.7 逐次モデルの使用

治療のアルゴリズムが確立されており、それぞれの治療ステージにおける治療効果データが活用可能である場合、疾患増悪後の治療を逐次的にモデル化することができる。疾患増悪後にはアルゴリズムに従った治療が行われると仮定し、その期間のデータは(新治療を評価する臨床試験ではない)別のデータソースから推定する。スイッチングによるバイアスの影響を受けた期間のデータが他のデータソースから補完され、スイッチングの影響を受けない分析が可能となる。しかしながら、疾患の終末期の治療の有効性データを獲得するのはしばしば困難であるという課題がある。

3.3 スイッチングの調整法(高度な方法)

3.3.1 Inverse Probability of Censoring Weights(IPCW 法)

IPCW 法は、選択的な治療切り替えや脱落のような情報のある打ち切りが存在する中で治療効果を推定するために開発された。IPCW においてはまずスイッチング時点で打ち切りがあったとみなし、共変量からその確率を求めるモデルを作成する。その後打ち切り、すなわちスイッチングがなかった症例を、それぞれの打ち切り確率に応じた重みづけをして解析に用いる方法である。これにより、打ち切り確率の観点から同様の性質を持つ患者の情報を、打ち切りのなかった患者に重みづけることができるので、打ち切りに伴う選択バイアスを取り除くことができる。打ち切りの多い特性を持っているにもかかわらず、打ち切られなかった症例には、より高い重みが付与される。

IPCW 法では、“no unmeasured confounders(未測定の交絡因子がないこと)”という重要な仮定をおく。この仮定は、打ち切り（すなわちスイッチング）を予測するためのすべてのベースライン変数及び時間依存性共変量が測定されており、利用可能である、というものである。この仮定は実際には現実的ではないが、重要な予測因子が欠けておらず、近似的に仮定が成立している限り IPCW 法は実用性に堪える方法である。死亡リスクとスイッチングの発生確率の間に相関があるため発生する選択バイアスは、この仮定のもと調整が可能であり、調整された Kaplan-Meier 推定量、ログランク検定やハザード比を得ることができる。

“no unmeasured confounders”の仮定は IPCW 法の限界と密接に関連している。“no unmeasured confounders”の仮定を、観測されたデータのみから検証することはできない。IPCW 法は、もともと観察研究の解析を主眼に、周辺構造モデル (Marginal Structural Model; MSM) として開発されたが、通常 RCT ではこの方法を適応するのに十分な例数を確保できない (特に対照群でスイッチングがない症例の数が不足することが多い)。このため IPCW 法による推定は不安定になり、信頼区間幅も広がるであろう。更に、スイッチングにつながる重要な予測因子が RCTs では測定されていないことが多い (例えば、患者自身の治療選好等) か、試験のどこかのタイミング (治療中止や疾患増悪時) で予測因子データの収集をやめてしまうため、IPCW 法の使用を難しくする。更に悪いことに、スイッチングを決定的に発生させてしまう共変量があった場合、その共変量によってスイッチング有無の群が完全に分断され、比較できなくなるため IPCW 法は適用ができなくなるという弱点がある。

3.3.2 Rank Preserving Structural Failure Time Model (RPSFTM)

RPSFTM は、RCTs における治療不遵守の課題を解決するために考案された、治療の因果効果を反事実モデルの枠組みのもとで推定する方法である。治療群と独立の反事実生存時間（仮に治療を受けなかった場合の生存時間）が存在すると仮定し、g 推定によって推定する。RPSFTM は操作変数法の一つであり、治療とアウトカムの両方を説明するすべての因子について情報を得られないような場合にしばしば用いられる方法である。スイッチングが予後に依存していると考えられる状況で因果効果を推定するため、割り付けられた治療群を実際の治療を予測する操作変数として利用する。

RPSFTM は「未測定の交絡因子がない」という仮定に依存せず、割付群、観察された生存時間、観察された治療歴のみで治療効果を推定する。標準的な 1 パラメータのモデルでは、すべての患者においていつ治療を受けたかによらず治療効果が等しく（"common treatment effect"の仮定）、ランダム化の意味は単にベースライン時の背景因子がランダムに分布することのみであり、よって無治療での生存時間は割付群と独立であると考えられる。

標準的な 1 パラメータの RPSFTM の限界は、"common treatment effect"の仮定と、ランダム化の仮定による。後者は RCT では基本的には問題がないが、重要なベースライン因子が群間で偏る可能性は留意すべきである(RPSFTM の解析においてベースライン因子を調整することもできる)。一方で"common treatment effect"の仮定はより重大である。もともと新治療群に割付けられていなかった患者が、試験途中で新治療を受けたときの治療効果が、もともと新治療群に割付けられていた患者における治療効果と異なる場合、RPSFTM の治療効果の推定値はバイアスの影響を受ける。スイッチングが、疾患増悪後のみ認められる場合、患者が新治療から得ることのできる治療効果は劣るであろうから、"common treatment effect"の仮定は臨床的に妥当ではないと考えられる。RPSFTM をデータに適用する、"treatment group", "on treatment"解析について後に説明する。この方法を適用することの正当性は、感度分析を行って説明する必要がある。

RPSFTM :

Robins らにより提唱された反事実加速モデルは以下のように定式化される。

$$U_i = \int_0^{T_i} \exp[\psi A_i(t)] dt$$

U_i は患者 i の反事実生存時間であり、生存時間 T_i 、観察された治療 $A_i(t)$ （治療の有無により 1 か 0 を取る。治療は時間 t によって変化するとする）から既知の関数により導かれる。 ψ は未知の治療効果パラメータである。この式は、反事実生存時間は、治療を受けていなかった期間と治療を受けた期間に $\exp(\psi)$ の重みを付けた総和である、と解釈できる。これを g 推定により推定するが、そ

の方法は、取りうる ψ の値のうち、割付群と反事実生存時間が独立になるような治療効果 ψ_0 を探索するものである。スイッチングの問題にこの方法をあてはめると、無治療下での生存時間 U_i は割付群と独立であるので、 U_i の期待値は両群で合致するはずである。新薬群の患者は最後まで新薬を使用する一方、対照群の患者の一部が観察途中で新薬に切り替わるであろうが、いずれの群においても新薬を使用する期間に $\exp(\psi)$ の重みを付け、新薬を使用していない期間との総和を取ることで U_i を求めることができる。この値が一致するよう、 ψ_0 を求めればよい。具体的にはログランク検定、Wilcoxon 順位和検定や Wald 検定の検定統計量(z)が 0 になる ψ が ψ_0 の点推定値である。

White ら[52]は RPSFTM において打ち切りの扱いが問題になると指摘している。正/負の治療効果が、打ち切りまでの時間を延長/短縮する。一方、スイッチングが発生したとき、施される治療はその予後と関連があると考えられる。すなわち、反事実生存時間の打ち切りは、予後因子に依存しており、即ち情報のある打ち切りである。

3.3.3 Iterative Parameter Estimation algorithm(IPE)[53]

RPSFTM は加速モデルに基づくノンパラメトリックな手法であったが、IPE によりこれをパラメトリックな手法に拡張する。調整していない ITT のデータにパラメトリックなモデルを当てはめ、治療効果の推定の反復計算の最初の推定値を求める方法である。”common treatment effect”という基本的な仮定は共通である一方、パラメトリックへの拡張に伴い適切なパラメトリックモデル選択を行う必要がある。g 推定でなくパラメトリックモデルを使用する利点は、収束が早い点にある。

IPE :

以下のように加速モデルを用いる点では RPSFTM と同様である。

$$T_{L,i} \stackrel{d}{=} e^{-\eta X_i} T_{R,i} \quad (1)$$

$T_{L,i}$ は患者 i が対照治療を受けた場合に観測される潜在生存時間、 $T_{R,i}$ は患者 i が割り付けられた治療のもとで実際に観測された生存時間、 $e^{-\eta X_i}$ は治療効果(対照治療の場合 $X_i = 0$ 、新治療の場合 $X_i = 1$)を示し、 $\stackrel{d}{=}$ は分布が等しいことを示す。

この式に(時点 U_i で発生した)スイッチングを考慮すると、以下のように書き換えられる。

$$T_{L,i} \stackrel{d}{=} U_i + e^{-\eta X_i} (T_{R,i} - U_i) \quad (2)$$

スイッチングが発生するのは対照群のみに限定されるという前提であるので、スイッチングが発生するまでの U_i の期間は新治療の効果 $e^{-\eta}$ を享受しないが、その後の $(T_{R,i} - U_i)$ の期間は新治療の効果を受ける、という構図である。なお、新治療については $U_i = 0$ として、試験治療の開始時(=新治療の開始時)から新治療の効果の影響を受けるとモデル化することができる。

IPE では ITT に基づく調整等を加えていない元データの生存時間をもとに(1)式をに分布を当てはめ、パラメトリックにパラメータ($e^{-\eta}$)を推定する($e^{-\hat{\eta}^1}$)。推定されたパラメータを用い、(2)式により対照群の患者のうちスイッチングのあった患者の $T_{L,i}$ を求める($T_{R,i}^*$)。スイッチングのあった患者の $T_{R,i}^*$ と、それ以外の患者の $T_{L,i}$ をもとに再度(1)式により $e^{-\eta}$ を推定する($e^{-\hat{\eta}^2}$)。これを繰り返し、 $e^{-\hat{\eta}^j} - e^{-\hat{\eta}^{j-1}}$ が十分小さく (Branson らは 10^{-5} を基準として提案している)なったところで収束したとして繰り返しを止める。

3.3.4 Alternative “two-stage” method

本節のここまでに説明を加えた方法は”standard”なスイッチングの調整方法であるが、”two-stage”法も紹介する。この方法は、疾患増悪までをランダム化比較試験、その後を観察試験とみなす。まず、スイッチングの発生した患者における治療効果を推定し、それらの患者の生存時間を調整する。その後に新薬群における治療効果を推定する。スイッチングのある患者の治療効果を推定するため、structural nested failure time models(SNM)と g 推定を使用する方法が提案されている。SNM は「未測定の変絡因子がない」という仮定を置くため、同じ仮定に基づく IPCW 法と同じ問題点を持つ。

g 推定を用いない、単純化された two-stage 法を紹介する。この方法はがん領域の RCTs のように、疾患増悪後のみスイッチングが許容されるような場合に適応可能である。この方法は、前述の状況においてスイッチングの時点では疾患の状態はどの患者においても同じである、即ち第二の”baseline”である。疾患増悪時に測定された共変量と、スイッチングの有無を示す指示変数を含む加速モデルにより、対照群でスイッチングのあった患者における疾患増悪後の新治療の効果を、同じく対照群であるがスイッチングのなかった患者との比較により評価する。この結果導かれた加速因子を用い、スイッチングのあった患者で実際に観察された生存時間を”shrink”させることで、仮にスイッチングがなかった場合に観察されたであろう、スイッチングの影響を受けていない反事実データを導く。前述の g 推定を併用する方法では、疾患増悪後の時間依存性変絡因子の調整を行うが、上記の簡易版においては調整を行わない。よってこの方法は疾患増悪とスイッチングの間には時間依存性変絡がないという強い仮定を要する。また、スイッチングのあった患者における治療効果をパラメトリックに推定することから、分布の仮定も必要となる。

この簡略化された two-stage 法は、IPCW 法のように時間依存性変絡を調整する方法と比べると理論的に劣るが、第二のベースライン時点以外で時間依存性変絡因子を測定する必要がないという実務上の利点がある。そのために研究デザインに課される制約は、第二のベースライン、即ち疾患増悪後のみスイッチングを許すこと、第二のベースラインでの予後予測因子を測定しておくことのみである。更に、

スイッチングが第二のベースライン直後に発生する場合、疾患増悪とスイッチングの間に起きる時間依存性交絡は微々たるものである。

RPSFTM や IPE 法と異なり、簡略化された two-stage 法は”common treatment effect”の仮定を必要としない。なぜなら、この方法はまずスイッチングのあった患者のみで治療効果を推定するからである。

3.3.5 Using external data

スイッチングの影響を受けた RCT データではなく、外部データ、例えば対照薬の別の試験や、レジストリーデータにより OS を推定することも有効である。外部データの使用には当然課題もあり、例えば試験の登録基準に起因する患者背景の差異、試験の行われた時代や地域からくる標準治療の差異、診断基準の差異等が、異なる試験のデータの比較可能性を失わせる原因となる。個票データが利用可能であれば、患者背景等の調整により反事実生存時間を導くことも可能であるが、これも使用するデータに調整が必要な予後因子がすべて含まれていることが前提であり、現実的とは言えない。よって外部データの使用は一般的に活用可能であるとは言い難い。

3.4 費用対効果評価への適用

3.4.1 理論的境界

IPCW 法と two-stage 法はスイッチングの発生に”no unmeasured confounders”の仮定、即ち利用可能な交絡因子によってスイッチングの発生確率が説明できると考える。この仮定が正しいか否かを証明することはできないが、類似の研究、専門家の意見等がその妥当性の判断に有用である。試験のなかでスイッチングが何故発生するのか、どのような場合にスイッチングが発生するのかを考察することも重要となる。IPCW 法では、スイッチングのなかった患者のみのデータで対照群の有効性を推定するため、そのような患者のデータが少ない場合には適応が難しい。two-stage 法では第二のベースラインとスイッチングの間に発生する時間依存性交絡の調整が行われないため、より時間依存性交絡の影響を受けやすい。

RPSFTM と IPE 法は”common treatment effect”の仮定に基づく方法であり、その臨床的/生物学的妥当性が重要となる。スイッチングが疾患増悪後に発生する場合、増悪後に投与された新治療の効果は、治療初期から割り付けに基づき投与された場合のそれよりも劣ると考えるのが自然である。この仮定を緩めようとする試み、例えば多変量による RPSFTM も、因果効果の推定が困難であるためうまくいっていない。スイッチングのある患者の治療効果と、新薬群に割り付けられた患者での治療効果の比較も、時間依存性交絡が存在するため妥当とは言えない。ランダム化の時点でさまざまな重症度の患者が混在しているような研究であれば、第二のベースライン時においても重症度によるサブグループ解析を行うことで、対処可能な可能性がある。治療の作用機序や、臨床の専門家の意見を考慮することはここにおいても重要となる。

RPSFTM と IPE 法は、RCT において対照群で使用する薬剤がプラセボではなく、他の薬剤である（新治療ほどではないが生存延長が見込まれる）場合にも、その影響を受ける。これらの方法が用いる反事実生存モデルでは、患者がそれぞれの時点で新治療に関し“on”か“off”のいずれかにある、という状況を想定するためである。もし対象群の患者が、まず（新治療ではない）実薬を投与され、疾患増悪後には緩和治療を受けるような場合、“off”が新治療でない実薬、あるいは緩和治療のいずれかになり、カテゴリが2つに増えてしまう。これに対処するには、因果効果パラメータを追加する必要があるが、このような方法はまだ開発されていない。このような場合でも通常の RPSFTM や IPE 法を適応することが可能であるが、治療方針と、治療効果に関するいくつかの重要な仮定を置く必要がある。これに関連して、標準的な“on treatment” RPSFTM や IPE 反事実生存モデルは、治療効果は患者が“on” treatment であるときのみに現れ、治療中断と同時に消えてしまう、とみなす。この仮定が臨床的に正しいか否かは確認を要する。もし治療効果が治療中断後にも継続する場合、RPSFTM や IPE 法を当てはめるには治療効果にタイムラグを想定するか、そうでなければ“treatment group”をもとに、即ち新薬群の患者は常に“on” treatment であり、仮に新治療の投与が中断されたとしても死亡するまでずっと新治療の効果が発現していると扱う必要がある。“treatment group”に基づくこの方法は、治療中断の時点を考慮せず、実際に新薬を投与されていた事実よりも、新薬群に割り付けられていた事実をもとに評価を行っており、ITT 解析に近い考え方であるといえる。試験において新治療群の患者に見られた治療効果は、試験終了後の治療によるものもすべて新治療に起因し、逆に対照群でスイッチングが発生しなかった患者に見られた治療効果は、試験終了後の治療によるものも含めすべて対照治療に起因するとみなす。この方法は、試験終了後の治療が実医療を反映している場合、費用対効果評価の観点から見ると妥当なものであるといえる。特に、試験終了後の治療のコストを費用に計上するのであれば、実医療の費用と効果の両方を勘案したモデルであるといえる。よって、“treatment group”にもとづくこのようなアプローチは、実薬対照の試験や、治療効果が治療終了後にも保たれるとみなしてよい場合において検討する価値がある。

ランダム化に基づく RPSFTM や IPE 法は一般に、スイッチングがある場合に ITT 解析と同様検出力が損なわれ、また推定値の信頼区間が広くなることは重要なポイントである。一方で観測に基づく IPCW 法や two-stage 法はその影響を受けないが、データがまばらな場合はこれらの方法によっても信頼区間は広くなる。

3.4.2 実務上の限界

これまで議論したようなスイッチング調整方法を実際の費用対効果評価に適用した場合の、実務上の限界についても考慮する必要がある。Latimer は生存時間データの、試験で観察した期間以降への外挿の方法について、推奨される方針を示している。そこではパラメトリックモデルをそれぞれの治療群に独立に当てはめて外挿する方法、1つのパラメトリックモデルに、治療群を変数として含める方法の2つが示されている。それらのどちらの方法においても、スイッチング調整方法と併用した際に問題が発生する。RPSFTM、IPE と two-stage 法はいずれもスイッチングを調整した反事実データを生成するが、治療による正の（あるいは負の）延命効果は患者が打ち切りとなる確率を増加（又は減少）させる。一方でスイッチングの発生原因から、治療の内容は予後因子の影響を受ける。反事実データ上で観測される

打ち切りが予後因子の影響を受けている、これは即ち情報のある打ち切りであることを示す。RPSFTM, IPE や two-stage 法にはこのことによるバイアスが含まれるため前述の外挿法のいずれにも適用することができなくなる。このバイアスに対応するため、治療スイッチの前に生じ得る打ち切り、すなわち再打ち切りの定義を設定する必要がある。再打ち切りが、反事実イベントである治療スイッチよりも早く発生した場合、その患者はイベント発生ではなく打ち切りと扱う。この再打ち切りのタイミングは、治療効果が高いほど早い傾向にある。再打ち切りにより、長期観察の情報が失われることになるが、これは費用対効果評価で生存時間データを外挿する際の障害となる。更に、再打ち切りを用いる方法では長期観察時の治療効果が失われることから、治療効果が時間と共に変わる場合は結果がバイアスを含むことになる。

IPCW 法は反事実データに基づく調整済みハザード比と重みづき Kaplan-Meier (WKM) 曲線で治療効果を示す。しかしながら、IPCW 法はそれぞれの患者に重みづけを行う関係で、反事実データセットにパラメトリックモデルをあてはめることは困難である。Kaplan-Meier 曲線の図から生存時間を復元する方法がWKMを使った反事実データセットの代用として活用されている[54]。"Survivor function"アプローチと呼ばれる、比例ハザード性をもとにした外挿法も代用法として使用されている。この方法は、(スイッチングの影響を受けない) 観測された治療群の生存時間について IPCW 法に基づくハザード比をパラメトリックモデルに当てはめ、そこから導かれたハザード関数を IPCW 法に基づくハザードの逆数で乗じ、対照群のハザード関数を求め、それから生存関数を導く。この方法では、ハザード比をそれとは別で当てはめられたパラメトリックモデルに適用するので、誤差が発生するが、その度合いは小さいと考えられる。

3.4.3 方法とプロセスガイダンス

理論的な仮定に関する知識、スイッチング調整方法に伴う限界、費用対効果評価の中での実用可能性、シミュレーション実験でのパフォーマンスから、スイッチング調整の方法を選択するための実用的なガイドラインを示す。スイッチングの調整法に付随する限界から、どのようなケースも網羅する決定的なガイドラインを示すことはできないが、それでもなお有用性は高いと考える。ガイドラインを、図 3-2 のように解析フレームワークの形で作成した。留意いただきたいのは、このガイドラインが観測データのスイッチングの影響を調整する方法に焦点を当てていることである。つまり、外部データを活用するような方法については本ガイドラインの対象外である。しかしながら、スイッチングの課題に取り組むに当たり反事実生存時間の推定のために外部データを活用することが可能であるか、また現実的であるかは検討する必要がある。スイッチング調整の方法と外部データの活用は解析の両輪として考えるべきである。スイッチング調整方法は試験期間中に観察された事象に限定した議論である一方、経済評価は長期間への外挿が必要となるためである。このことは図 3-2 の Step (5)にも触れているが、TSD14[13]で外的妥当性と臨床的な尤もらしさが生存時間の外挿においてもっとも重要な要素であることに触れているので、これを参照すること。スイッチングの有無にかかわらず、外部データを探すことは有用な方策である。

図 3-2 の説明に入る。Step (1)は、スイッチングの発生メカニズムの検討である。これにより、どの調整方法が適切でありえるかを検討することができる。例えば、スイッチング発生の予測因子が測定されているか(もしない場合、IPCW 法は適切とは言えない)、経済評価の比較対照が RCT の対照薬として含まれているか、等が最たる例である。スイッチングが許容される時点がいつ始まるのかも、重要なポイントで、これにより two-stage 法が適切か否かが判断できる(疾患に関連したある時点以降でのみスイッチングが許される場合に限りこの方法は適切である)。

Step (2)では、スイッチングの発生割合を評価する。もし対象群の 90%を超える患者にスイッチングが発生している場合、対照群の例数が高々500 例程度であれば IPCW 法は大きなバイアスを含むだろう。スイッチングの割合が低かった場合でも、IPCW 法の使用には条件があり、仮にスイッチングが対照群の患者のうち 50%にしか発生しなかったとしても、もしそれらの患者が疾患増悪のあった患者の 90%に相当していたとしたら(故にスイッチングの条件に合致した、というシナリオ)、IPCW 法は同様にバイアスをもたらすだろう。スイッチングの条件を満たした患者のうち、実際にスイッチングが発生した患者の割合が重要となるのである。同様の課題は、スイッチングが限られた患者集団、例えば治療奏功の経験があったり、特定のバイオマーカーが確認された場合においてのみ認められる場合にも発生する。IPCW 法と two-stage 法のようにランダム化に基づく方法は、スイッチングの割合が高かったとしても比較的バイアスの影響を受けにくいため、この観点からは望ましい。一方で、治療効果に高い時間依存性があることがわかっていたり、RCT の対照群が実薬対照であったりする場合は、通常の反事実生存モデルは不適切であるといえる。

Step (3)では、上記ステップをもとに各調整方法が前提とする重要な仮定が成り立つか否かを判断し、いずれの調整方法が妥当であり得るかを評価する。RPSFTM や IPE アルゴリズムであれば、“common treatment effect”の前提が成立するか否かが重要である。スイッチング発生の頻度にもよるが、対数累積ハザードや Q-Q プロットは比例ハザード性や加速因子が時間依存性であるかを確認するために有用なツールである。ランダム化時点で患者の疾患ステージが異なっているような場合、“common treatment effect”の担保のため疾患ステージによるサブグループごとの治療効果を確認する必要がある。一方で、異なる疾患ステージの患者がランダム化されていること自体も、スイッチングに伴うバイアスの原因となる可能性はある。

治験データから“common treatment effect”の仮定を検証することは難しいので、外部データと、臨床及び生物学的観点からの専門家の意見は必ず取り入れるべきである。患者特性や疾患状態が治療効果に与える影響に関し既にわかっていることがあればその情報は活用すべきである。これらの情報により“common treatment effect”の妥当性が示される場合、RPSFTM や IPE 法は有効な方法となる。RPSFTM、IPE 法及び IPCW 法のいずれにおいても、治療効果(ハザード比、加速因子)の大きさを考慮することが重要である。

RPSFTM や IPE 法を使用するに当たり、治療効果があらわれる期間(すなわち、治療中断後も治療効果が継続するか否か)を検討する必要がある。もし治療効果が治療中止後も継続すると見なせる場合、割り付けられた群に基づく方法(もしくは治療効果にずれを許容する方法)の使用が検討されるだろう。実際の治療に基づく方法と割り付けられた群に基づく方法のどちらを使用するかは、費用対効果評価の

目的、臨床の観点からの意見、生物学的妥当性とデータの利用可能性をもとに決定され、説明されるべきである。それぞれの方法に基づく生存時間の推定や費用対効果評価の結果の両方を示すことも、感度分析の観点から望ましいだろう。臨床の専門家の意見をもとに、治療中断の後に治療効果が消失するか、継続するのがあるいは悪化させるのかを示すことも、選択した方法の正当化に重要な知見となる。RCT で用いられた比較対照（実薬か否かにかかわらず）も、方法の選択において考慮材料になる。実薬対照であった場合、RPSFTM や IPE 法は適切ではない一方で、割り付けられた群に基づく方法は治験での観察に基づいて、想定される治療パスを設定することで、正当化されるだろう。

実際の治療に基づく標準的な方法である RPSFTM や IPE 法はすべての患者が即ち完全なコンプライアンスを達成した場合の新治療の効果と、対照群の何も治療を受けていない状態での効果の差を推定することに留意すること。この仮定は経済評価において適切ではない場合がある。なぜなら、治験において観測された投与中止は、実際の臨床現場で発生する中止をある程度反映していると考えられるからである。したがって、無治療群の生存時間を実際の治療に基づく方法で治療中止に伴い薬効がなくなるという想定のもとで推定することは妥当ではあるものの、その生存時間を比較する実薬群の生存時間には実際に治験で観察された生存時間を用いるべきである。

IPCW 法は未測定の変数因子がないという仮定のもと正当化される。よって重要な共変量が測定されていない可能性については、臨床専門家の意見や、同様の疾患領域の別試験等をもとに評価する必要がある。しかしこれだけでは未測定の変数因子がないという想定を保証するには不十分である。治療スイッチの意思決定に影響を与えたと考えられるすべての予後因子情報、即ち治療スイッチが適切かどうかについての医師の意見、患者の状況やスイッチングへの選好を記録することが必要である。しかしながらこれらのような情報は通常 RCTs で測定されることはない。これに加えて、治験のどこかの時点（例えば疾患増悪のタイミング等）でしばしば共変量を測定することを止めてしまうことも、IPCW 法の適用を困難とする。これらの課題は、Step (1) (2)と共に勘案されるべきである。

Two-stage 法の適応を考えるにあたり、それより前にスイッチングが起こりえない第二のベースライン（例えば疾患増悪）が適切に設定できることが必須である。そのような時点が定義可能である場合に限り two-stage 法は適応可能であるが、一方でその時点からスイッチングが発生するまでの時間が長ければ長いほど時間依存性交絡に伴うバイアスが発生する可能性が高まる。

スイッチング調整法の適用後、Step (4)ではそれらの調整法が適切であったかを検討する。RPSFTM、IPE 法や two-stage 法では、これは再打ち切りの発生頻度や、治療中断後の治療効果の継続の有無を確認するために RPSFTM や IPE 法の結果を割り付けられた群に基づく方法から得られた結果と比較することを意味する。更に、g 推定の結果を評価し、RPSFTM が単一の治療効果に収束しているか、RPSFTM や IPE 法で推定した反事実生存時間は群間で等しいはずだが、異なっていないかを検証する。IPCW 法では、それぞれの患者で算出された重みを評価することが特に重要であり、特定の患者に重みが偏っている場合、結果に悪影響を与える可能性がある。Two-stage 法の結果は、他の方法を用いた結果の適切性を評価するために活用可能である。例えば、スイッチングのあった患者での治療効果と新薬群に割り付けられた患者での治療効果の推定が、two-stage 法と RPSFTM や IPE 法で同様であれば、RPSFTM や IPE 法の使用は妥当であり、そうでなければ妥当ではない。

複雑なスイッチング調整法と合わせて、ITT による解析も検討する余地がある。他の方法によるスイッチング調整がうまくいかない場合、時に ITT 法が最もバイアスが少ない。例えば治療効果が小さく (Latimer et al [55] のシミュレーションによると、新薬群で HR が概ね 0.75 から 1.00)、且つスイッチングのあった患者における新治療の治療効果がもともと新薬群に割り付けられた患者における治療効果よりも 15% 少ない場合では、ITT による解析が IPCW や RPSFTM、IPE 法よりも望ましいことがわかっている (ただし、ITT 法にバイアスが含まれていることに変わりはない)。もしスイッチングのあった患者における治療効果の減少がさらに大きく、25% 減少する場合、治療効果が大きい (HR で 0.50 程度) としても ITT 法は IPCW や RPSFTM、IPE 法よりも望ましい。スイッチング調整法にはいずれも課題があるため、ITT による解析は常に示されるべきである。スイッチングの割合が低かったり、治療効果が低かったり、スイッチングの患者における治療効果が大きく減少する場合、ITT はもっともバイアスが少ない。

Step (1) から (4) で理論上や実施上の適切性、パフォーマンスの観点からスイッチング調整法の評価を行った後、Step (5) では、必要に応じて調整法を外挿の手法と組み合わせることになる。RPSFTM、IPE 法と two-stage 法においては、再打ち切りが反事実 Kaplan-Meier 曲線の裾に与える影響を確認し、再打ち切りが生存時間の外挿に与える影響を評価するべきである。再打ち切りのなかった新薬群の生存時間の外挿には、スイッチング調整済みの治療効果を用いる、“生存関数に基づく”方法が適している。しかし、外挿法の選択には NICE DSU の提供している TSD14 も参考にすること [13] [56]。IPCW 法と比例ハザード性に基づいた外挿法を併用する場合には、重み付き Kaplan-Meier 推定量を反映したデータを再現できる適切な方法を選択する必要がある。

最後に、粗解析からは適切なスイッチング調整法が決められなかった場合、調整法の違いによる不確実性を評価するために感度分析を行う必要がある。

図 3-2 スイッチング調整法フロー図



3.5 ディスカッション

スイッチング調整法は費用対効果評価においてあまり使われていなかったり、誤用されたりしてきた。NICE TA のレビューの結果、時に適切と思われる方法が使用されていたものの、多くの場合では単純な調整法が使用されており、それらの方法がバイアスを含むことは前述のとおりである。RPSFTM や IPCW 法のような、複雑ではあるが適切な可能性がある方法が使用されていても、それらに関する深い考察はなされておらず、故に選択した手法の限界については議論がなされてこなかった。このことは、スイッチング調整法が時に ICER の値に大きな影響を与えることを考えると、重大な問題である。3.4.3 章にまとめた分析フレームワークは、不適切であったり一貫性を欠いたりする調整法の使用を避け、適切な手法の選択基準をより厳密に定めるものであった。

RPSFTM/IPE 法と IPCW 法には大きな違いがあり、それらの方法が求める仮定の違いも大きいので、どれか 1 つの方法が常に他の方法より優れているということはない。試験の性質、スイッチングの性質の両方を勘案し、適当と思われる手法を柔軟に選択することが重要となる。IPCW 法はもともと観察研究のために開発された経緯があるので、RCT の環境で使用するためには「未測定の変動因子がない」という前提条件に依存している点に十分注意する必要がある。RPSFTM と IPE 法は割付けられて使用された新薬と、スイッチングされて使用された新薬が「共通の治療効果」を持つという仮定が必要であるという限界があり、この仮定はスイッチングが疾患増悪後にのみ認められる場合には満たされない可能性がある。今まであまり使用されていなかったが、two-stage 法は、特に RPSFTM、IPE 法や IPCW 法のような他の調整法による調整が大きなバイアスを含む場合に使用を検討する価値がある。この方法は第二のベースラインを定義できなければならない、という難点はあるものの、「共通の治療効果」の仮定が必要なく、「未測定の変動因子がない」という仮定も、第二のベースライン時点でのみ成立すればよい。一方で、この方法は、第二のベースラインとスイッチング発生時点の間に時間依存性変動因子がないという強い仮定を求めるという欠点もある。スイッチングが第二のベースラインの後ただちに発生する場合、時間依存性変動因子はさほど問題にならないが、そのようなケースはなかなかないであろう。

おわりに

本報告書では、NICE DSU が公開した TSD 14（1 章：Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data）、16（3 章：Adjusting survival time estimates in the presence of treatment switching）及び 19（2 章：Partitioned Survival Analysis as a decision modelling tool）をもとに、費用対効果評価における生存時間データの扱いについて、その方法論と課題をまとめ、それらの対処についてまとめた。冒頭でも触れた通り、費用対効果評価においては生存時間データに限らず、様々な外挿が行われる。これは費用対効果評価が、実際には発生していない事象（例：「将来的に、新薬が対照薬の代わりに使われたらどうなるか？」）や、データとして得られていない事象（例：特定の疾患状態下での臨床成績）を論じるものである以上避けて通ることはできず、外挿に関する諸課題はある程度受け入れて議論を進めることが必要であろう。しかしながら、費用対効果評価により作成されたモデルが現実の臨床実態をどれだけ正確に映しているかは、本報告書でまとめたような、起こりうる多くのバイアスの検証や、臨床実態に関する知識をどれだけ集約できたかによると考えられる。その努力をもってすれば、モデルは机上の空論ではなく実臨床を映す有用なツールとなるだろう。

医療経済評価の目的は、将来の費用と結果を予測することではない。医療経済評価は、ある時点で必ずなさねばならない医療に関する意思決定をより良くする手段の 1 つとして行われる。すなわち、分析時点で入手可能なエビデンスを統合し、その時点での疾患に関する理解や医療システム等を踏まえて分析が行われる。その後、時が経つにつれエビデンスが集積し、疾患の解明が進むことにより、基本的には予測されたものは全て誤りとなるだろう。ゆえに、医療経済評価の評価は、将来の費用と結果を正確に予測したかどうかではなく、その分析はその時点で有用であったか、それはより良い決定につながったか、という観点からなされるべきである。

引用文献

- [1] 齋藤信也ら.(2013) 英国国立保健医療研究所（NICE）における社会的価値判断—NICE ガイダンス作成のための諸原則（第二版）—. 保健医療科学, 62(6), 667-78
- [2] 医薬品、医療機器及び再生医療等製品の費用対効果評価に関する取扱いについて. (2019). Available: https://c2h.niph.go.jp/tools/system/notification_cost-effectiveness.pdf
- [3] National Institute for Health and Care Excellence. Available: <https://www.nice.org.uk/>
- [4] NICE Decision Support Unit. Available: <http://nicedsu.org.uk/>
- [5] 保健医療経済評価研究センター. Available: <https://c2h.niph.go.jp/>
- [6] 中央社会保険医療協議会における費用対効果評価の分析ガイドライン第 2 版 (2019). Available: https://c2h.niph.go.jp/tools/guideline/guideline_ja.pdf
- [7] 鎌江伊三夫. (2016). 医療技術評価ワークブック. じほう
- [8] 五十嵐中, 佐條麻里. (2014). 「薬剤経済」わかりません！！. 東京図書
- [9] Drummond, M., Sculpher, M. J., Claxton, K., Stoddart, G. L., Torrance, G.W., (監訳) 久繁哲徳, 橋本英樹. (2017). 保健医療の経済評価 第 4 版. 篠原出版新社
- [10] 日本製薬工業協会. (2016). 医薬品の価値の科学的な評価—データサイエンス担当者のための費用対効果評価の現状と手法解説—. Available: http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/2014ds_tf3.html
- [11] 日本製薬工業協会. (2018). 費用対効果評価の実践—NICE 技術評価ガイダンスの事例から分析モデルのバリエーションを学ぶ—. Available: <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/cost-effectiveness.html>
- [12] Agnes B. et al. Survival Modelling in UK Oncology Technology Appraisals Since the Publication of Good Practice Guidelines. The Evidence forum May 2014. 2014. Available: <https://www.evidera.com/wp-content/uploads/2015/04/Survival-Modelling-in-UK-Oncology-Technology-Appraisals.pdf>
- [13] Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. (2011). Available: <http://www.nicedsu.org.uk>
- [14] Gibson E. et al. Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: A systematic approach to data analysis and extrapolation. pharmacoeconomics. 2017 Dec;35(12):1257-1270.

- [15] Tremblay G. et al. A criterion-based approach for the systematic and transparent extrapolation of clinical trial survival data. *JHEOR* 2015;2(2):147-60
- [16] Guyot P. et al. Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves, *BMC Med Res Methodol* 2012;12:9
- [17] 大橋 靖雄ら(1995). 生存時間解析—SAS による生物統計. 東京大学出版会
- [18] 大橋 靖雄ら(2016). 生存時間解析 応用編: SAS による生物統計. 東京大学出版会
- [19] Agnes B. et al. Survival Modelling in UK Oncology Technology Appraisals Since the Publication of Good Practice Guidelines. The Evidence forum May 2014. 2014. Available: <https://www.evidera.com/wp-content/uploads/2015/04/Survival-Modelling-in-UK-Oncology-Technology-Appraisals.pdf>
- [20] Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974 Mar;30(1):89-99.
- [21] Royston P. et al. Flexible proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*. 2002; 21:2175-2197;.
- [22] Jackson C. et al. Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. *Int J Biostat*. 2010;6(1):Article 34.
- [23] Jacobs IJ. et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet*. 2016 Mar 5;387(10022):945-956.
- [24] Harrell, F.E. Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. Springer, New York, 2001.
- [25] Jackson C. et al. Extrapolating survival from randomized trials using external data: A review of methods. *Med Decis Making*. 2017 May;37(4):377-390.
- [26] Wolchok JD. et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *N Engl J Med*. 2017 Oct 5;377(14):1345-1356.
- [27] Lambert P. et al. Further development of flexible parametric models for survival analysis. *Stata J*. 2009;9:265–90.
- [28] NICE TA400 (2016). Nivolumab in combination with ipilimumab for treating advanced melanoma. Available: <https://www.nice.org.uk/guidance/ta400>

[29] Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU technical support document 19: Partitioned survival analysis for decision modelling in health care: a critical review. (2017) Available: <http://nicedsu.org.uk/technical-support-documents/partitioned-survival-analysis-tsd/>

[30] NICE TA531 (2018) Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. Available: <https://www.nice.org.uk/guidance/ta531/>

[31] Reck, M, Rodríguez-Abreu D, Robinson A et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med 2016; 375:1823-1833

[32] キイトルーダ点滴静注 20mg、同点滴静注 100mg 申請資料概要 Available: <http://www.pmda.go.jp/drugs/2016/P20161025002/index.html>

[33] Modeling the economic outcomes of immuno-oncology drugs: alternative model frameworks to capture clinical outcomes Gibson et al. ClinicoEconomics and Outcomes Research 2018;10 139-154

[34] NICE Decision Support Unit. Introduction to evidence synthesis for decision making. Available: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD1-Introduction.final_.08.05.12.pdf

[35] NICE Decision Support Unit. A general linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. Available: <http://nicedsu.org.uk/wp-content/uploads/2017/05/TSD2-General-meta-analysis-corrected-2Sep2016v2.pdf>

[36] NICE Decision Support Unit. Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Available: <http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD3-Heterogeneity.final-report.08.05.12.pdf>

[37] NICE Decision Support Unit. Inconsistency in networks of evidence based on randomised controlled trials. Available: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD4-Inconsistency.final_.15April2014.pdf

[38] NICE Decision Support Unit. Evidence synthesis in the baseline natural history model. Available: <http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD5-Baseline.final-report.08.05.12.pdf>

[39] NICE Decision Support Unit. Embedding evidence synthesis in probabilistic cost effectiveness analysis: software choices. Available: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD6-Software.final_.08.05.12.pdf

[40] NICE Decision Support Unit. Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. Available: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD7-reviewer-checklist.final_.08.05.12.pdf

[41] NICE Decision Support Unit. Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. Available: <http://nicedsu.org.uk/wp-content/uploads/2019/10/TSD-20-mvmeta-final.pdf>

[42] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会. ネットワークメタアナリシスの概要および留意事項 — ISPOR レポート (2011) を参考に — . Available: http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/pdf/nma_20191110.pdf

[43] NICE Decision Support Unit. Cost-effectiveness modelling using patient-level simulation. Available: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD15_Patient-level_simulation.pdf

[44] Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. *Med Decis Making*. 2017 May;37(4):427-439.

[45] NICE TA311 (2014) Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation. Available: <https://www.nice.org.uk/guidance/ta311>

[46] National Institute of Health and Care Excellence, Guide to the methods of technology appraisal. 2013. Available: <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>

[47] National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal. 2013; available from <http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9> (accessed July 2013).

[48] Briggs, A., Claxton, K., Sculpher, M. Decision modelling for health economic evaluation. Oxford University Press Inc., New York; 2006.

[49] Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. Cost-effectiveness in health and medicine. Oxford University Press, USA, 1996.

[50] Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies. 2006. Canada, 3rd edition.

[51] Watkins C., Huang X., Latimer N., Tang Y., Wright E. J. (2013) . Adjusting overall survival for treatment switches: Commonly used methods and practical application. *Pharmaceutical Statistics*, 12(6), 348–357.

[52] White I.R., Babiker A.G., Walker S., Darbyshire J.H. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine* 1999; 18(19):2617-2634.

[53] Branson M., & Whitehead J. (2002) . Estimating a treatment effect in survival studies in which patients switch treatment. *Statistics in Medicine*, 21(17), 2449–2463.

[54] Guyot P., Ades A.E., Ouwens M.J., Welton N.J. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology* 2012; 12(1):9.

[55] Latimer N., Abrams K., Lambert P., Crowther M.J., Wailoo A., Morden J.P. et al. Adjusting for treatment switching in randomised controlled trials – a simulation study. University of Sheffield Health Economics and Decision Science Discussion Paper No.13/06 2013. 2013.

[56] Latimer N.R. Survival Analysis for Economic Evaluations Alongside Clinical Trials: Extrapolation with Patient-Level Data. Inconsistencies, Limitations, and a Practical Guide. Medical Decision Making 2013.

2019 年度データサイエンス部会 KT5

作成担当者

荒西 利彦	日本イーライリリー株式会社
河田 祐一	中外製薬株式会社*
中島 章博	帝人ファーマ株式会社

レビュー担当者

東 美恵	エーザイ株式会社
渥美 淳	東レ株式会社
石渡 量太	ユーシービージャパン株式会社
奥山 ことば	MSD 株式会社*
直井 一郎	大日本住友製薬株式会社
野島 俊秋	興和株式会社
町田 光陽	塩野義製薬株式会社
松下 泰之	第一三共株式会社*
吉田 瑞樹	ファイザーR&D 合同会社

*TF リーダー

担当副部会長

酒井 弘憲	エーザイ株式会社
-------	----------