

柔軟な生存時間のモデリング方法と
外部情報を利用した生存時間の外挿方法

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

2022 年度継続タスクフォース 7

「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

2023 年 6 月

本報告書の中で Technical Support Document 21 (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、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

1. はじめに	3
2. 柔軟な生存時間のモデリング方法	6
2.1. 柔軟なパラメトリックモデル	9
2.2. 混合分布モデル	17
2.3. ランドマークモデル	19
2.4. 区分モデル	27
2.5. 治癒モデル	32
2.6. 超過死亡モデル/相対生存時間モデル	36
2.7. ベイズ流アプローチの適用	39
3. 生存関数の外挿と外部情報の取り込み	40
3.1. 外部データの組み込み	41
3.2. 複雑な生存時間モデルに対する外挿	46
3.2.1. 柔軟なパラメトリックモデル	46
3.2.2. 混合分布モデル	46
3.2.3. ランドマークモデル	47
3.2.4. 区分モデル	47
3.2.5. 治癒モデル	47
3.2.6. 超過死亡/原因別死亡（相対生存）モデル	47
3.3. 外部データを利用する際の考慮すべき事項	48
4. TSD 21 のシミュレーション研究と推奨事項	50
4.1. TSD21 のシミュレーション研究	50
4.2. TSD21 の推奨事項	52
5. おわりに	55

1. はじめに

本邦では2019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[1]。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[2]。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22のTSDが発行されている[3]。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。次ページに成果物の一覧を示す。

本報告書は、費用対効果評価における柔軟な生存時間のモデリング方法に着目した。背景として、近年、がん免疫療法の登場などにより、複雑なハザードを持つ生存関数が想定されるケースが増加しており、このような場合には柔軟な生存時間のモデリングが必要となる。また、本邦の分析ガイドライン[4]に示されている通り、費用対効果評価の分析期間は「評価対象技術の費用や効果におよぼす影響を評価するのに十分に長い分析期間を用いる」とされている。そのため、長期における費用と効果を推定するためには、臨床試験の追跡期間を超えた期間に対する外挿が必要であり、その際にも生存時間のモデリングが必要となる。そこで、実務担当者を対象に柔軟な生存時間のモデリング方法及び外部情報を利用した外挿方法を紹介することを目的とした報告書を作成した。

生存時間のモデリングは、TSD14「Survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data」[5] (日本製薬工業協会の報告書でも紹介[6]) でも触れられていたが、主に標準的な生存時間のモデリング方法を紹介し、柔軟な生存時間のモデリング方法については十分に紹介されていなかった。そのため、2020年(2022年に更新)にTSD21「Flexible methods for survival analysis」[7]が公開された。本報告書では、TSD21の内容を参考に以下の内容を事例と共に紹介する。

- 近年の複雑なハザードの生存関数に対する柔軟な生存時間のモデリング方法 (2章)
- 臨床試験の追跡期間を超えた期間に対する外挿及び外部情報の取り込み方法 (3章)
- TSD 21 のシミュレーション研究と推奨事項 (4章)

生存時間のモデリングに関しては、国内外で企業分析とEvidence Review Group (ERG) や日本の公的分析班でモデル選択や外挿方法において異なるアプローチを採用した事例が散見され、その一部は本報告書でも紹介している。例えば、2.1項で紹介する事例では企業分析とERGで標準的なモデルを用いるべきか柔軟なモデルを用いるべきか議論となった。また、2.5項で紹介する事例では日本の公的分析班が企業分析の生存時間の外挿方法の妥当性に関して言及してい

る。そのため、事例を通して論点を把握し、標準的な生存時間モデリングだけでなく柔軟な生存時間モデリングや外部データを利用した外挿方法を理解し、実装できるように準備しておくことは有用と考える。将来、生存時間をエンドポイントとした臨床試験を用いて承認申請を行い、さらに当該データを用いて費用対効果評価を行う予定のある担当者の方は是非ご一読いただけたら幸いである。なお、本報告書は費用対効果評価や生存時間解析の基本的な知識を有することを前提にしている。そのため、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）のホームページ[8]、中医協ガイドライン第3版[4]、日本製薬工業協会の過去の報告書[9]、成書[10][11][12]等を、生存時間解析については成書[13][14]をそれぞれ参照されたい。また、生存時間をモデリングする際には、必要に応じて、治療スイッチング、Partitioned Survival Model の適用等を検討する必要がある、これらも NICE TSD[15][16]及び日本製薬工業協会の過去の報告書[6]を参照されたい。

【参考】医薬品評価委員会データサイエンス部会の費用対効果評価に関する主な成果物

- 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（2016年4月）
https://www.jpma.or.jp/information/evaluation/results/allotment/2014ds_tf3.html
- 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（2018年9月）
<https://www.jpma.or.jp/information/evaluation/results/allotment/cost-effectiveness.html>
- 患者レベルのモデルを用いた費用効果分析（2019年12月）
https://www.jpma.or.jp/information/evaluation/results/allotment/cost_effectiveness_modelling.html
- 費用対効果評価における生存時間解析の諸問題（2020年2月）
https://www.jpma.or.jp/information/evaluation/results/allotment/c_e_a_problems.html
- 非ランダム化比較試験データを用いて治療効果を推定するための統計的手法（2020年2月）
<https://www.jpma.or.jp/information/evaluation/results/allotment/non-randomization.html>
- NICE TSD Evidence Synthesis シリーズの紹介：ネットワークメタアナリシスを中心として（2020年12月）
https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd_evidence_synthesis.html
- NICE TSD の紹介・QOL 値 QOL 値の測定と費用効果分析への QOL 値の利用（2021年10月）
https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd.html
- NICE TSD の紹介・モデル作成 モデル作成のためのエビデンス収集とレビュー（2021年11月）
https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd_model.html
- 間接比較において母集団調整法は有用か？ Matching Adjusted Indirect Comparison 及び Simulated Treatment Comparison（2022年1月）
https://www.jpma.or.jp/information/evaluation/results/allotment/indirect_comparison.html

- 事例から学ぶ費用対効果評価の留意事項（2023年3月）
https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202303_cerpitfall.html
- 医療技術評価における多変量メタアナリシスと代替エンドポイント（2023年4月）
https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202305_TSD_20.html

2. 柔軟な生存時間のモデリング方法

従来、費用対効果評価における生存時間のモデリングは、生存時間の分布に指数分布やワイブル分布といった標準的な分布を用いたパラメトリックモデル（標準的なパラメトリックモデル）で十分であった（図 2-1）。しかしながら、近年ではがん免疫療法などの登場により、治療に対する反応の遅延、長期生存者の存在により生存時間曲線は複雑な形状を示す事例が増えてきた。このような複雑な形状を表現するためには、従来の標準的なパラメトリックモデルでは不十分であり、より柔軟な生存時間のモデルが必要となり、実際の費用対効果評価でも利用されるケースが増えてきている。

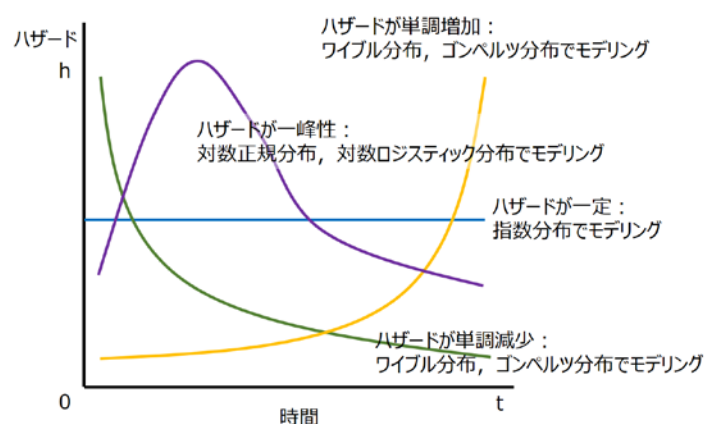


図 2-1 標準的なパラメトリックモデル[5]

本章では、柔軟な生存時間のモデリング方法として、表 2-1 の生存時間モデルについて、モデルの概要、限界、課題、適用事例をそれぞれ紹介する。

表 2-1 柔軟な生存時間のモデリング方法

モデル	特徴	前提となる考え方
柔軟なパラメトリックモデル（制限付き3次スプライン関数を用いたモデル）	ノットを設定し、ノット間のハザードをスプライン関数で表現することで、複雑な形状のハザード関数を柔軟にモデル化する。	時間とともにハザードの形状が複雑に変化する。
混合分布モデル	混合分布の成分ごとに異なる生存関数を設定し、混合割合で重みづけて全体を推定する。	成分間でハザードが異なる。 ※患者集団の違いに着目
ランドマークモデル	介入後の治療への反応の有無（ランドマーク；例えば、がん領域であればRECIST基準[17]）に基づき、その後の生存曲線をモデリングする。	予後を予測できるランドマークが存在する。 ※反応後のデータ（例えば、RECIST基準に基づく奏効の有無）に着目

モデル	特徴	前提となる考え方
区分モデル	区分（期間）ごとに標準的なパラメトリックモデルをあてはめる。柔軟なパラメトリックモデルとは標準的なパラメトリックモデルを区分ごとにあてはめる点で異なる。	区分ごとにハザードが変わる。 ※区分間のハザード変化に着目
治癒モデル	治癒するグループと治癒しないグループを想定し、生存曲線をモデリングする。	一定の割合の患者が注目するイベントを発現しない。 ※治癒の存在に着目
超過死亡モデル 相対生存時間モデル	対象疾患による死亡（超過死亡）率をモデル化し、他の原因による死亡率と統合することで、複雑な形状のハザード関数を推定する。	死亡の原因を特定できる場合を想定している。 ※原因別の死亡に着目

また、これらとは別にベイズ流のアプローチも提案されており、その概要を紹介する。これらの方法を適用する際には、R パッケージ（柔軟なパラメトリックモデル：flexsurv[18], flexsurvPlus[19], 混合分布モデル：mixtools[20], 治癒モデル：flexsurvcure[21], smcure[22], mixcure[23]), SAS マクロ（柔軟なパラメトリックモデル）[24]および4章で紹介するシミュレーション研究で用いられた STATA コード[7]が役立つと思われる。なお、標準的なパラメトリックモデルについては TSD14[5]または製薬協成果物[6]を参照されたい。本報告書の数式表記は TSD21 に沿った記載としている。

様々なモデルを紹介する理由は4章で示すように、あらゆる状況において普遍的に利用可能な生存時間のモデリング方法は存在しないためである。そのため、TSD14[5]で紹介されたアルゴリズム(図 2-2)を参考にモデルを選択する必要がある。その際、図中の赤枠部分において、ハザードの形状や臨床的な観点を考慮し、さらに従来の標準的なパラメトリックモデルだけでなく、表 2-1 に示した柔軟なパラメトリックモデルの特徴や前提となる考え方を踏まえた上で状況に応じてモデルを使い分けていく必要がある。さらに、いくつかのモデリング方法を用いて感度分析を行い、モデルの選択が評価結果に与える影響を考察することも重要である。

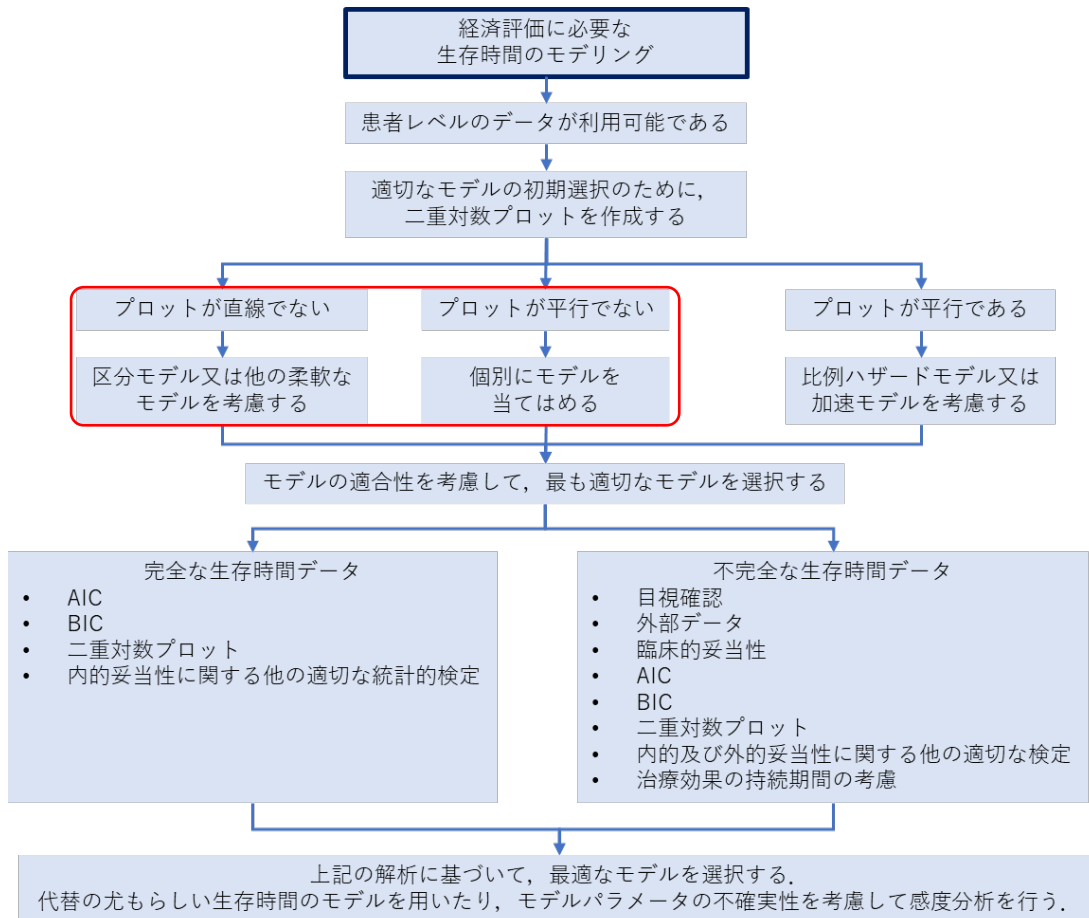


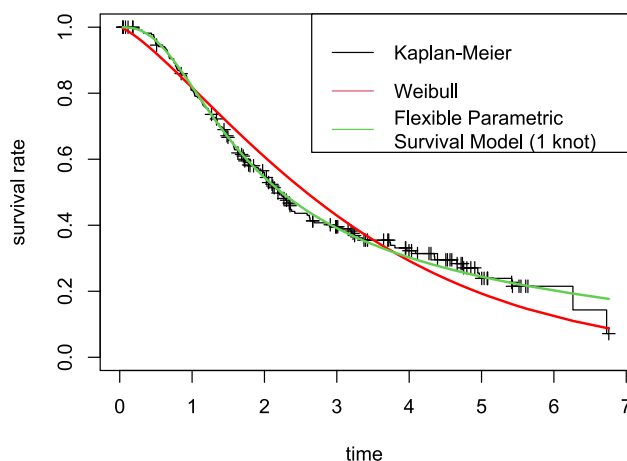
図 2-2 生存時間モデル選択のアルゴリズム
(NICE TSD14 Figure 3 を和訳し、一部改変) [6]

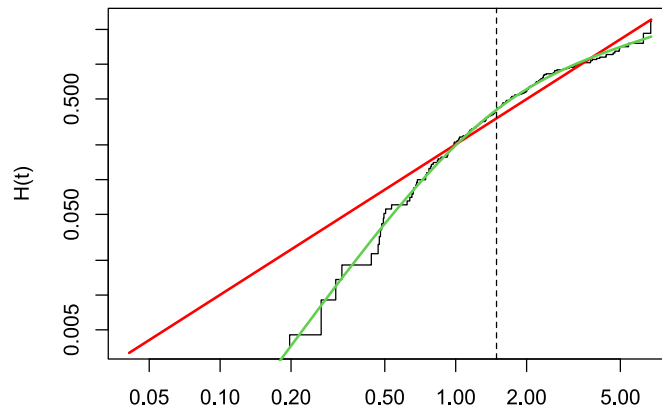
2.1. 柔軟なパラメトリックモデル

時間とともにハザードの形状が複雑に変化し、標準的なパラメトリックモデルでは表現しきれない場合、柔軟なパラメトリックモデル (Flexible Parametric Survival Models : FPMs) の適用が考えられる。

概要

- 制限付き 3 次スプライン関数を用いたモデル
 - 対数累積ハザードと時間の対数 (対数時間) との関係に、制限付き 3 次スプライン (Restricted Cubic Splines ; 最高次が 3 次の区分多項式に両端の外側が一次関数であると制限したもの、自然 3 次スプラインとも呼ばれる) を用いたモデルが Royston and Parmar(2002) [25]によって提案され、TSD21 の本文中ではこれを「柔軟なパラメトリックモデル」と呼んでいる。名称が紛らわしいため、実際にモデルの説明をする際は、制限付き 3 次スプラインを用いたモデルなどと表記することを推奨する。
 - ◇ 制限付き 3 次スプラインは製薬協の成果物[6]でも紹介されているため、適宜参照されたい。
 - 図 2-3 は柔軟なパラメトリックモデルとして制限付き 3 次スプライン関数を用いたモデリングの例である。Kaplan-Meier 曲線に対して柔軟なモデルの方がワイブル分布を用いたモデルよりもあてはまっていることが確認できる。





二重対数プロット

図 2-3 ワイブル分布を用いた標準的なパラメトリックモデルと柔軟なパラメトリックモデル

- ハザードの形状は区間と区間の結び目（ノット:knots）の数と位置に依存する。

- ✧ ノットの数を増やすことで、複雑な形状のハザードを表現することができる。

- 例えば、ワイブル分布を用いた生存関数を対数累積ハザードの尺度で表すと以下となる。

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

λ : ワイブル分布の尺度パラメータ

γ : 形状パラメータ

- 柔軟なパラメトリックモデルでは、線形関数である $\log(\lambda) + \gamma \log(t)$ を対数時間の制限付き 3 次スプライン関数である $s(\log(t))$ に置き換える。

- ✧ さらに共変量 x とそれに関連するパラメータ β も追加でき、以下の式で表される。

$$\log[H(t)] = \log[-\log[S(t)]] = s(\log(t)) + x\beta$$

- ✧ β は共変量 x に対応する対数ハザード比として解釈できる

- 柔軟なパラメトリックモデルでは、上記のように二重対数変換が用いられるが、ロジット変換やプロビット変換を用いることもできる。

- 制限付き 3 次スプラインのノットの数と位置について

- ノットの位置は、対数イベント時間（非打ち切り例の対数時間）に対し

	<p>て、ノットの数に応じて一様に配置する方法が典型的である。例えば、両端を除くノットの数がある場合は 25, 50, 75%点となる。図 2-3 は R パッケージ flexsurv[18]を用いて作成したものであるが、本パッケージの標準設定でも同じ方法が採用されていた (version 2.2 で確認)。</p> <ul style="list-style-type: none"> ➤ 追跡期間の範囲内で予測された生存関数は、基本的な形状を捉えるのに十分なノット数であれば、ノットの数と位置には敏感ではないことが示されている[26][27][28]。ノットが必要以上にあったとしてもほとんど影響はないものの、全体的なイベントが少ない場合には、スプライン関数のパラメータを多く持つことは推定精度が十分ではなくなるため、実用的ではないかもしれない。 ➤ このモデルを用いる目的が生存関数の長期間への外挿である場合、最後のノット以降で外挿されるため、ノットの数を変更することは、重要な影響を与える可能性がある。 ➤ ノットの数を決める際、AIC (赤池情報量規準) や BIC (ベイズ情報量規準) [6]といったモデル選択の指標がよく使われているが、あくまで目安である。異なるノットの数にした場合の影響(影響のなさも含めて)を把握したい場合は、感度分析が推奨される。 <p>● 時間依存性効果(Time-dependent Effects)について</p> <ul style="list-style-type: none"> ➤ 治療群を共変量としてモデリングする場合、前述のモデルでは比例的な治療効果を想定している。比例ハザード性のような比例的な治療効果の仮定は、生存時間モデルや外挿の適合が悪くなる可能性があるが、時間依存性の効果をモデルに取り入れることで、この仮定を緩和することができる。時間依存性の効果は、追加のスプライン項を作成し、これらと目的の共変量(治療群など)との間の交互作用を作成することで表現できる。 ◇ 例えば、ひとつの共変量 x を用いた場合、前述の対数累積ハザードの尺度で表したモデルは次式となる。 $\log[H(t)] = \log[-\log[S(t)]] = s(\log(t)) + x\beta + x \times s_t(\log(t))$ <p>1つ目のスプライン関数 $s(\log(t))$ はベースラインハザード関数であり、2つ目のスプライン関数 $s_t(\log(t))$ により、治療効果も時間の関数となる。</p> ➤ ベースラインと時間依存性効果では、異なる数のノットがありえることに注意する。 ➤ 時間依存性の効果を検討する場合、上述の方法の代わりに、治療群ごとに別々のモデルを当てはめることもできる。この場合、追跡期間を超え
--	---

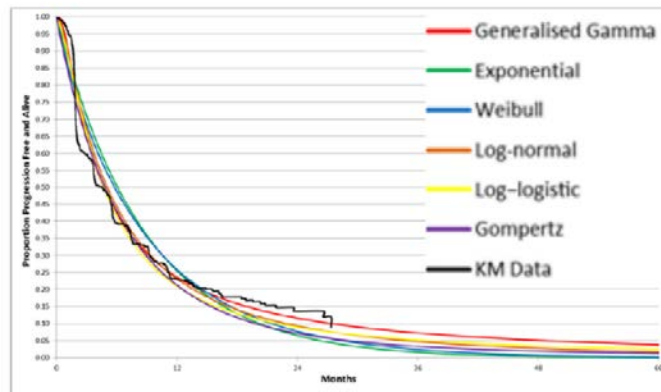
	<p>た長期の外挿箇所では非現実的な治療効果が予測される可能性があるため、外挿されたハザード関数をハザード比とともに時間に対してプロットすることで、相対的な治療効果を把握しやすくなる。</p> <ul style="list-style-type: none"> ● ポリハザードモデル(Poly-hazard Model) <ul style="list-style-type: none"> ➢ その他の柔軟なパラメトリックモデルとして、ポリハザードモデルが提案されている[29]。このモデルは、K個のハザードの合計である全体的なハザード関数を想定している。これは原因別の競合リスクモデル、つまり、各構成要素は全体的なハザードに加法的に寄与すると考えるものである。Demiris et.al [29]は、次のように定義されるポリハザードモデルを提案している。 $h(t) = \sum_{k=1}^K h_k(t)$ <p>バスタブ型のハザード関数を捉えたいという動機からすべてのハザード成分にワイブルモデル $h_k(t) = \lambda_k \gamma_k t^{\gamma_k - 1}$ を使用することが提案されている。</p> ➢ 複数のパラメトリック分布を選択できるため、複数の競合する死因が存在して、それらに関する情報が得られる場合に適している。
限界	<ul style="list-style-type: none"> ● 長期間の外挿が目的の場合、AIC/BIC をモデル選択の唯一の基準として使用すべきではない。 ● 十分な数のノットが使用されていれば、観測されたデータの範囲内では良好な適合性を示すが、長期間の外挿箇所が信頼できることを意味するものではない。 ● 外部データが組み込まれていない場合、予測が過度にモデルに依存し、現実から乖離する可能性がある。長期間の外挿は追跡期間終了時の傾向に基づくが、ノットの数が多い場合、限られたイベント数に基づいている可能性がある。外部データの使用に関しては3章を参照されたい。 ● 複数の時間依存性効果がある場合、ある共変量のハザード比は他の時間依存性共変量の値に依存する。そのため、時間依存性効果を持つ多くの共変量をハザード比で要約しようとするときにどのように示すのか等が問題となる。 <ul style="list-style-type: none"> ➢ 共変量間に交互作用がない場合でも、1つの時間依存性効果のハザード比が、2つ目の時間依存性効果の値に依存するためである。 ➢ 生存関数の予測および経済的意思決定モデルで使用するための平均生存期間または境界内平均生存時間(Restricted Mean Survival Time: RMST)の推定に関心がある場合は重要な限界ではない。

課題	<ul style="list-style-type: none"> ● 追跡終了後に共変量の効果について仮定する場合は、対数累積ハザードの尺度ではなく対数ハザードの尺度で仮定の妥当性を考える方が簡単かもしれない。 <ul style="list-style-type: none"> ➤ 例えば、ある時点以降、2つの治療群に関連するハザードが同じであると仮定する場合である。 ● 罰則付きスプライン関数を用いた研究も行われている[30]。これらのモデルではより多くのノットを選択し、罰則関数を尤度に組み込むことで、関数を強制的に滑らかにする。しかし、ユーザーは初期ノットの数と組み込む罰則関数を定義する必要がある。このようなアプローチをとった場合、長期間の外挿にどのような影響があるかは現在のところわかっていない。
事例 ⁱ	<p>前治療歴のある進行性腎細胞がんを対象としたニボルマブの技術評価ガイダンス (NICE TA417 Published: 23 November 2016) [31]の事例を以下に示す。</p> <ul style="list-style-type: none"> ● CheckMate 025 試験 (ニボルマブとエベロリムスのランダム化非盲検第 III 相臨床試験; n=821; 最短観察期間 14 カ月) に対して、無増悪生存期間 (Progression Free Survival Time : PFS), 投与中止までの期間 (Time to Discontinuation : TTD) のそれぞれにスプラインを用いた柔軟なパラメトリックモデルを検討していた。ここではニボルマブ群の生存時間のモデリングについて紹介する。 <p><u>企業分析</u></p> <ul style="list-style-type: none"> ● PFS について、6 個の標準的なパラメトリックモデルと 6 個のスプラインモデルを検討した (図 2-4)。AIC と BIC が最も良く (表 2-2), 図の目視確認によるあてはまりの程度もよいモデルとして、2-knots spline odds modelⁱⁱを採用した。

ⁱ 英国では日本の専門組織に当たる組織として NICE における Appraisal committee がある。Appraisal committee は、企業分析や ERG による分析を踏まえて当該品目の費用対効果評価を取りまとめ、その結果に基づき意思決定に係るガイダンスを出している。事例の紹介として、企業分析、ERG、評価委員会のそれぞれの立場からの見解を記載した。

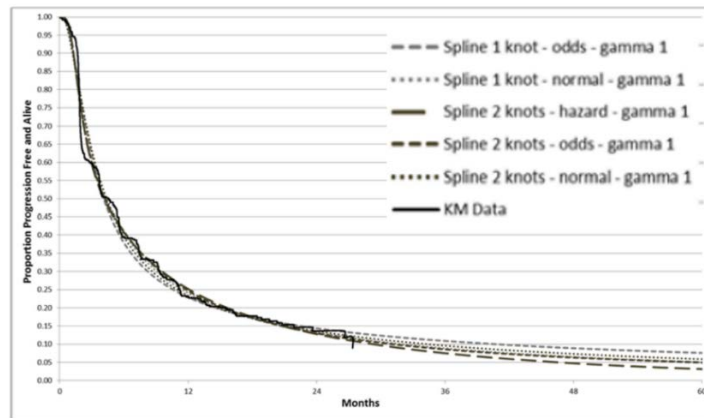
ⁱⁱ おそらく対数累積オッズについてのモデルを採用したと考える。

Figure 33: Parametric model fits to stratified PFS data from CheckMate 025, nivolumab arm



Key: KM, Kaplan-Meier; PFS, progression-free survival.

Figure 35: Spline model fits to stratified PFS data from CheckMate 025, nivolumab arm



Key: KM, Kaplan-Meier; PFS, progression-free survival.

Spline model の凡例の odds, normal, hazard はそれぞれ

log-logistic, log-normal, weibull 分布のモデルの拡張に相当する。

図 2-4 CheckMate 025 試験のニボルマブ群の PFS に適用したモデルのあてはまり (TA417 committee-papers-4 Figure 33, 35 を一部改変)

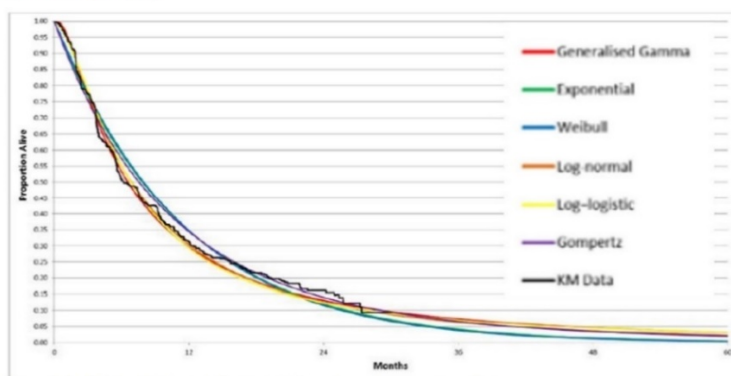
表 2-2 CheckMate 025 試験のニボルマブ群の PFS に適用したモデルの AIC, BIC (TA417 committee-papers-4 Table 33, 35 を一部改変)

Model	AIC	BIC
Generalised gamma	1932.912	1944.961
Log-normal	1944.538	1952.570
Log-logistic	1951.954	1959.986
Gompertz	2006.797	2014.829
Weibull	2018.543	2026.575
Exponential	2020.251	2024.267
Spline 2 knot(s) - odds	1897.302	1897.302
Spline 2 knot(s) - hazard	1897.665	1897.665

Spline 1 knot(s) - odds	1909.947	1909.947
Spline 1 knot(s) - hazard	1915.430	1915.430
Spline 1 knot(s) - normal	1921.659	1921.659
Spline 2 knot(s) - normal	1923.369	1923.369

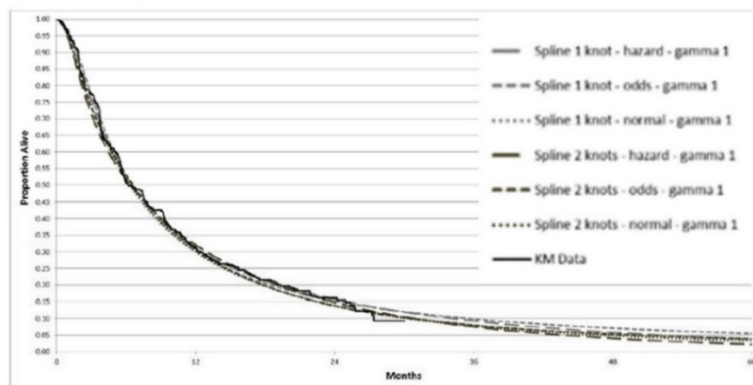
- TTD についても同様のモデルを検討した結果、2-knot spline hazard model を採用した (図 2-5, 表 2-3).

Figure 41: Parametric model fits to un-stratified TTD data from CheckMate 025, nivolumab arm



Key: KM, Kaplan-Meier; TTD, time to treatment discontinuation.

Figure 43: Spline model fits to un-stratified TTD data from CheckMate 025, nivolumab arm



Key: Kaplan-Meier; TTD, time to treatment discontinuation.

図 2-5 CheckMate 025 試験の TTD に適用したモデルのニボルマブ群へのあてはまり (TA417 committee-papers-4 Figure 41, 43 を一部改変)

表 2-3 CheckMate 025 試験の TTD に適用したモデルの AIC, BIC (TA417 committee-papers-4 Figure 37, 38 を一部改変)

Model	AIC	BIC
Generalised gamma	4429.807	4448.561
Log-normal	4432.481	4446.546
Log-logistic	4449.747	4463.812
Gompertz	4525.747	4539.812

Exponential	4532.696	4542.073
Weibull	4533.875	4547.940
Spline 2 knot(s) - hazard	4415.016	4438.458
Spline 2 knot(s) - odds	4418.138	4441.580
Spline 1 knot(s) - normal	4424.155	4442.909
Spline 2 knot(s) - normal	4425.559	4449.001
Spline 1 knot(s) - hazard	4425.759	4444.513
Spline 1 knot(s) - odds	4429.522	4448.275

Evidence Review Group (ERG)による分析

- PFS への柔軟なパラメトリックモデルの適用については企業分析の方法を支持した。特に RECIST 基準を用いた免疫療法の進行度評価の問題、すなわち、RECIST 基準では一次的な進行であってもイベントになるため、免疫療法の効果を過小評価する問題について言及し、時間によって相対効果が異なる仮定は合理的と考えた。一方で、TTD への柔軟なパラメトリックモデルの適用については、複雑なモデルの使用は正当化されていないため、単純なモデル(対数正規もしくは一般化ガンマ)でデータに適合すると言及したⁱⁱⁱ。

評価委員会の判断

- TTD を予測するために、企業分析が複雑なスプラインモデルを適用したことに注目した。また、ERG が用いた単純なモデルは試験の開始時点ではよりデータに適合しているように見えたが、終了時点ではあまり適合していないと考えた。企業または ERG が提示したどの曲線も TTD の Kaplan-Meier 曲線全体にうまく適合しているとは思えなかった。最終的には対数正規モデルをベースケースとし、感度分析として一般化ガンマモデルが用いられた。

※その後、CheckMate 025 試験は 2020 年に 5 年間の追跡調査の結果が公表されている [32]。ニボルマブ群の 36, 48, 60 ヶ月時点の PFS 率はそれぞれ 9% (95% CI : 6-12%), 6% (95% CI : 4-9%), 5% (95% CI : 3-8%)であり、上述の 2-knots spline odds model の外挿箇所と同様の結果であった。

ⁱⁱⁱ 企業分析ではノットの位置を報告しておらず、解析に使用したコードも提出していなかった。そのため、ERG では R パッケージ flexsurv のデフォルト設定の位置と仮定したとのことである。

2.2. 混合分布モデル

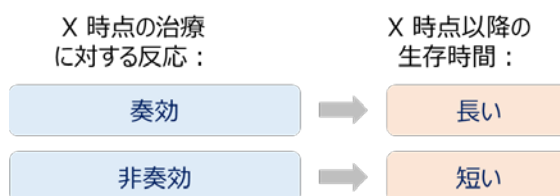
混合分布モデル (Mixture Models) は、評価対象集団の中で異なるクラスターまたはサブグループが異なるハザードおよび生存プロファイルを持っていると考えられるような場合に有用な方法である。例えば、治療に反応した人は、反応しなかった人とは異なる生存プロファイルを持っているかもしれない。

概要	<ul style="list-style-type: none"> ● 混合分布モデルの定義 $f(t) = \sum_{k=1}^K p_k f_k(t), \quad \text{where } \sum_{k=1}^K p_k = 1$ <p>$f(t)$: k 番目の混合成分の加法的な分布関数 $f_k(t)$ からなる全体の分布関数 p_k : k 番目の混合成分が全体の分布関数に寄与する割合, 混合確率</p> <ul style="list-style-type: none"> ➤ 各混合成分には、それぞれ異なる分布を選択可能である。 ➤ 各混合成分は、混合確率を含むすべてのパラメータにおいて共変量効果で調整できるため、治療に対する反応が異なる可能性のある集団を特定するために使用することができる。 ➤ $\sum_{k=1}^K p_k = 1$ という制約のため、混合確率は多項分布でモデル化する [33]. ● 混合分布モデルの仮定 <ul style="list-style-type: none"> ➤ ハザード関数は、標準的なパラメトリックモデルを混合することで適切にモデル化することができる。 ➤ 混合成分それぞれのデータは、ロバストな生存モデリングに十分であり、それぞれの混合成分の分布の選択が妥当である必要がある。 ➤ 本モデルはサブグループ解析をしているわけではないため、潜在的なサブグループを表現するために混合分布モデルを使用した場合、患者を特定のサブグループに割り当ててのではなく、各個人が各グループに属する確率を推定することになる。混合分布モデルを用いた教師なし学習によるクラスタリングとみなすと理解しやすいかもしれない。 <ul style="list-style-type: none"> ◇ 各患者は各分布に属する確率が割り当てられ、患者が個別のグループに決定的に分離されるわけではない。 ➤ 混合分布モデルからの長期間の外挿は、各混合成分が受け取る重み (p_k) と、それぞれのハザード比の相対的な大きさの両方の組み合わせに影響を受けるため、事前の知識または他の情報源からのより長期の生存情報を用いて、長期間の外挿に対する妥当性を評価する必要がある。
----	--

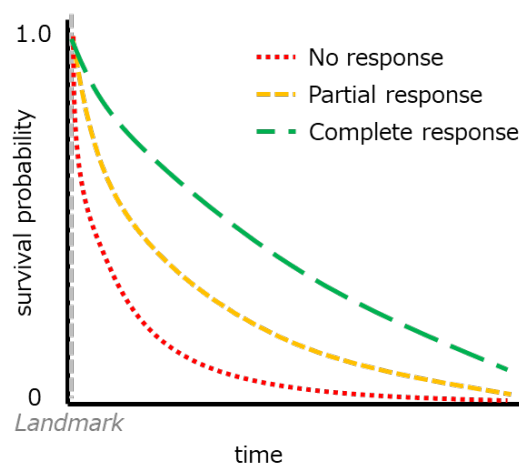
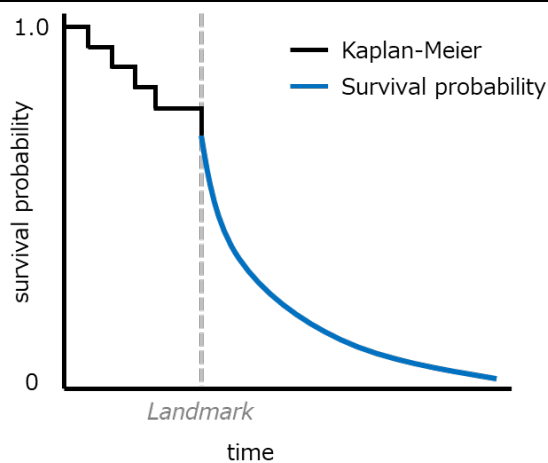
<p>限界</p>	<ul style="list-style-type: none"> ● 混合成分の数，混合する各分布の選択には限界がある。 ● 複雑なハザード関数を捉えることが目的であれば，スプライン関数を用いた柔軟なパラメトリックモデルの方が混合分布モデルよりもはるかにモデリングが簡単である。 ● 過剰な解釈をする危険性がある。 <ul style="list-style-type: none"> ➤ 例えば，n 個の混合成分を想定してモデリングをした際に「患者の標本は異なるリスクプロファイルを持つ n 個の混合から構成されている」と解釈してしまう可能性がある。そのような想定を行ったのは解析を行った者であるため，注意が必要である。 ● モデル収束に課題があり，推定が難しいことが知られている [34]。 <ul style="list-style-type: none"> ➤ 例えば，$p_1 + p_2 = 1$ の 2 成分モデルでは，同じパラメータの組み合わせで，同一の関数を与えるものが 2 つある。つまり，$p_1 = 0.2$ と $p_2 = 0.8$ の場合，推定された成分分布パラメータが与えられれば，成分分布パラメータを入れ替えて，$p_1 = 0.8$ と $p_2 = 0.2$ の場合も同じモデルが得られることになる。
<p>課題</p>	<ul style="list-style-type: none"> ● 前述の通り，混合分布モデルの解釈には注意が必要である。 ● 混合分布モデルの各パラメータを共変量で調整して推定することはもちろんのこと，各治療群に対して全く異なる混合分布モデルを当てはめることも可能である。あるいは，治療効果を共通として推定することも可能であり，その場合の混合成分は，より柔軟なベースライン関数をモデル化するために使用される。各患者に割り当てられた混合確率や，共変量が含まれている場合にはその共変量の影響を考慮し，サブグループの性質や存在について言及する場合には，慎重に検討する必要がある。また，生存プロファイルの異なるサブグループの存在は生物学的な妥当性に基づくべきである。
<p>事例</p>	<ul style="list-style-type: none"> ● 混合分布モデルは，特定の疾病が治癒したグループと治癒していないグループが存在すると考える「混合治癒モデル」として使われる場合が多い。混合治癒モデルの事例は 2.5 項を参照されたい。

2.3. ランドマークモデル

ランドマークモデル (Landmark Models) は、各患者が治療に反応するか否かでその後の生存時間が異なる可能性があると考える方法であり、治療への応答の有無が生存時間の重要なサロゲートであると仮定する。ランドマークモデルでは、治療への応答の有無が観測される「ランドマーク」の時点で、全ての患者が治療への応答カテゴリに従ったグループに分けられる。例えば、抗がん剤領域の応答カテゴリとして、RECIST 基準による効果判定に基づいた「完全奏効」、「部分奏効」、「安定」、「進行」、またはこれらを「奏効」、「非奏効」に統合したものが考えられる。ランドマークモデルは、このように「ランドマーク」が存在する場合に有用な方法である。



概要	<ul style="list-style-type: none"> ● ランドマークモデルの定義 <ul style="list-style-type: none"> ➤ ランドマークの時間を越えた生存時間関数 $S(t)$ を以下に示す。 $S(t) = S(l) \times \left(\sum_{i=1}^k S_i(t T > l) \times \frac{n_i}{\sum_{i=1}^k n_i} \right)$ <p> $S(t)$: 時間 t での生存率 l: ランドマークの時点 $S_i(t T > l)$: i 番目の応答カテゴリの患者に対する時点 l での生存が所与の下での時点 t での生存率 k: 応答カテゴリの数 n_i: 応答カテゴリ i の患者数 </p> ➤ $S(t T > l)$ はランドマークの時点 l 以降で、各応答カテゴリに使用される生存時間モデルである。 ➤ ランドマークモデルのイメージを図 2-6 に示す。この例ではランドマーク時点までの生存率は Kaplan-Meier 曲線に基づいており、ランドマーク時点以降は、3つの応答カテゴリにあてはめられたモデルを各カテゴリの患者割合に基づいて併合し、推定される。
----	--



時点(t)での生存確率

=ランドマーク時点での生存確率×

$$\left(\begin{array}{l} \text{no response の患者割合} \times \text{no response に対する時刻(t)での生存確率} + \\ \text{partial response の患者割合} \times \text{partial response に対する時刻(t)での生存確率} + \\ \text{complete response の患者割合} \times \text{complete response に対する時刻(t)での生存確率} \end{array} \right)$$

図 2-6 ランドマークモデルの概要 (TSD21 Figure 5 を一部改変)

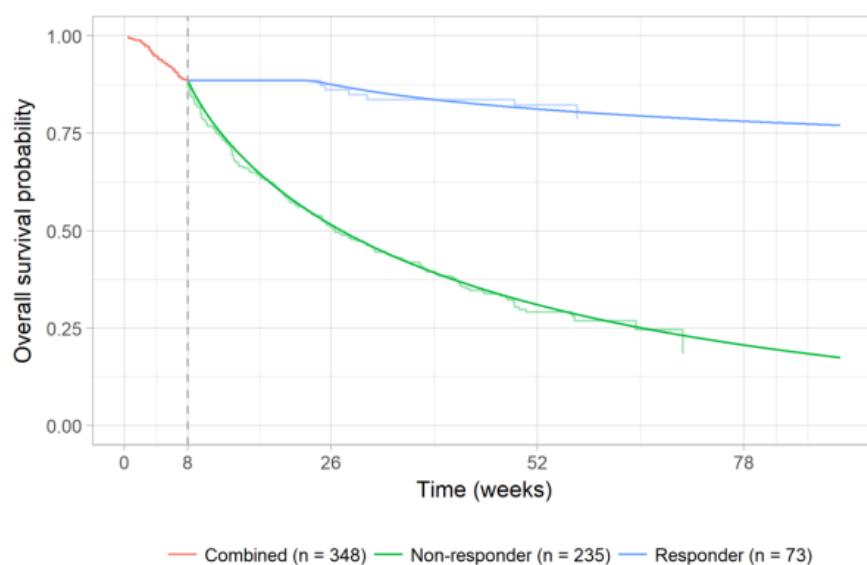
- ランドマークモデルの仮定
 - (応答を評価するための) ランドマークの時点が適切である.
 - ランドマークの時点及び応答カテゴリは臨床的に意味があり, 生存時間のサロゲートとなっている.
 - 頑健な生存時間モデルを構築するために各応答カテゴリには十分なイベント数があり, 各応答カテゴリの生存時間モデルは長期の生存関数を適切に捉えている.

限界

- ランドマークの時点は任意に設定できるが, 分析結果に大きな影響を与える可能性がある. 例えば, ランドマークの時点が早いと遅い時点の反応を見落

	<p>としてしまう可能性があり、ランドマークの時点が遅いとランドマークの前に一部の患者が死亡しグループ化の意義が低下する可能性がある。</p> <ul style="list-style-type: none"> ● ランドマークの時点での応答カテゴリがその後の生存時間の適切なサロゲートと見なされない場合、カテゴリ化された各グループの生存時間を個別に推定することは正当化されない可能性がある。 ● 患者を応答カテゴリにグループ化すると、各グループの患者数またイベント数が少なくなる可能性があるため、生存時間モデルのあてはめにおいて大きな標準誤差と不確実性が生じる。 <ul style="list-style-type: none"> ➤ 例えば、ランドマークの時点が比較的遅い場合、「非奏効」グループでは患者の多くがランドマークの時点より前に死亡する可能性があり、「奏効」グループでは死亡例が非常に少なく、頑健な生存時間モデルのあてはめが難しくなる可能性がある。 ● 全ての患者で応答が測定されるとは限らない。また、応答が測定された場合でも測定したデータにエラーが発生する可能性がある。 ● 応答カテゴリに使用される生存時間モデルが適切にモデリングされていない場合、全体の予測結果が不適切な結果となる。
課題	<ul style="list-style-type: none"> ● 通常、ランドマークモデルでは各応答カテゴリの生存時間をモデル化するために標準的なパラメトリックモデルが用いられる。ただし、生存関数の外挿部分でハザードの追加の変化が予想される場合に十分に表現できない可能性がある。
事例①	<p>プラチナ製剤を含む化学療法後の局所進行性切除不能又は転移性尿路上皮癌に対するニボルマブの技術評価ガイダンス（NICE TA530 Published: 04 July）の事例[35]を以下に示す。</p> <p>ニボルマブのOSのモデリングにCheckMate 275試験とCheckMate 032 試験の2試験が用いられた。</p> <p>CheckMate 275 試験の概要</p> <p>転移段階におけるプラチナ製剤による治療後もしくはプラチナ製剤による術前または術後補助療法から1年以内に、進行または再発した転移性または切除不能な尿路上皮がん患者を対象にした第II相単群非盲検試験（n = 270）</p> <p>CheckMate 032 試験の概要</p> <p>進行期または転移性固形がんを対象としたニボルマブまたはニボルマブとイピリムマブの併用療法の第 I / II相非盲検試験（プラチナ製剤治療歴有の転移性尿路上皮がんの群ではn = 78）</p> <p><u>企業分析</u></p>

- 持続的かつ長期的な奏効のモデル化には、標準的なパラメトリックモデルは適さないと考え、奏効の有無に基づくランドマークモデルによるモデリングを試みた。
- ランドマーク時点の奏効の有無（レスポナー/ノンレスポナー）別にそれぞれの集団に一般化ガンマ分布を用いた。ランドマーク時点は奏効までの時間の中央値とした。



Source: Response to clarification letter, Figure 34⁷

^a The ERG requested corrected figures because the number of responder was incorrect in the original CS

図 2-7 OSに対するランドマークモデルのあてはまり
(TA530 committee-papers Figure 5.4より引用)

ERGによる分析

- ランドマークモデルを用いる理由として企業の主張する「標準的なパラメトリックモデルでは異なるレスポンスは正確に記述できない」ことを支持する数学的な正当性が示されていない。標準的なアプローチの一つである一般化ガンマ分布のモデルは、ランドマークモデルに必要な仮定が不要であり、異なる反応をモデル化するのに十分な柔軟性があると判断した。

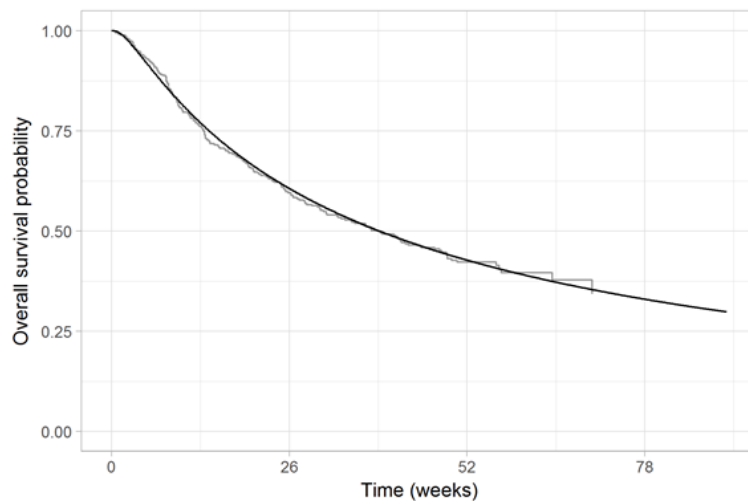


図 2-8 OSに対する一般化ガンマ分布のモデルのあてはまり
(TA530 committee-papers Figure 5.2より引用)

評価委員会

- 企業分析の奏効に基づくランドマークモデルのアプローチを検討することには同意したが、モデル化に不必要な複雑さをもたらすというERGの意見にも同意した。モデルの適切性をサポートするためにはより多くの証拠が必要と述べた。ERGによる5年生存率の推定値は約13%、企業モデルだと約20%であり、これについて臨床専門家の意見や他の免疫療法の評価の際の推定値を考慮し、ERGによる推定値が臨床的に妥当と判断した。

事例②

内分泌療法後の進行性乳がんを対象としたエベロリムス（エキセメスタン併用）の技術評価ガイダンス（NICE TA421 Published: 21 December 2016）の事例[36]を以下に示す。本事例は上記で説明した治療への反応に基づくランドマークモデルをそのまま適用したものではないが、TSD21でランドマーク解析の適用事例として紹介されており、生存曲線に関する各治療群の特徴や部分的な類似性に注目してランドマーク解析を適用し、モデル化した結果について議論されたものである。ランドマークの概念を用いた生存時間分布のモデル化の方法の一つとして紹介する。

- BOLERO-2試験（内分泌療法後のHER2陰性ホルモン受容体陽性乳がん患者を対象とした、エベロリムス+エキセメスタン群（以下、エベロリムス群）、又はプラセボ+エキセメスタン群（以下、プラセボ群）の二重盲検ランダム化比較試験；n=724）に対して、ERGは全生存期間に探索的解析に基づくランドマーク解析を適用した。ランドマーク解析の手順はERGによる分析の項に示す。

企業分析

- 各群の治療開始後の全生存期間を対数ロジスティック関数によりモデル化した

ERG による分析

- エベロリムス群は疾患進行後に直接的な効果はなく、プラセボ群と同様の効果であり、生存期間のベネフィットは疾患進行前のフェーズに限られ、進行後の死亡率は2群間で同一と仮定した。
- 目視確認によりエベロリムス群への最適なあてはまりが得られるまで、プラセボ群の生存プロットを段階的にシフトしながら探索的解析を実行した(図 2-9)。

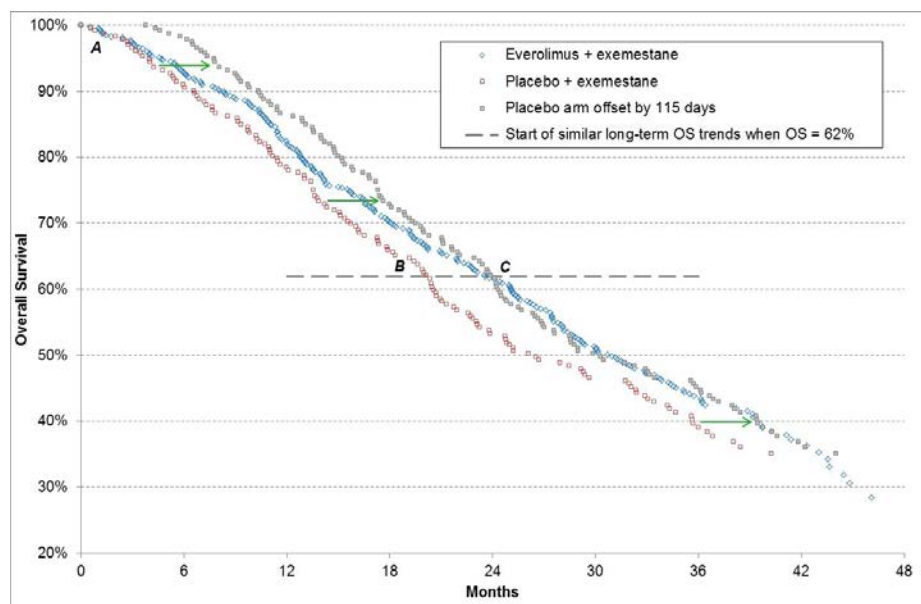


図 2-9 長期生存者のパターン比較のためプラセボ群の生存時間を 115 日右にシフトさせた全生存期間 (TA421 committee-papers Figure 2 より引用)

- 図 2-9 は推定生存率が 62%になるポイント (C) 以降の長期生存の傾向が2群で同様になっていることを示している。
 - つまり、エベロリムス群(図中の青線)は推定生存率が62%を下回った後はプラセボ群を115日シフトさせた線(図中の灰色線)と同様の推移を示している。
- この点は生存率が62%の時点のランドマークに対応するリスク集団での Kaplan-Meier ランドマーク解析で確認した(図 2-10)。

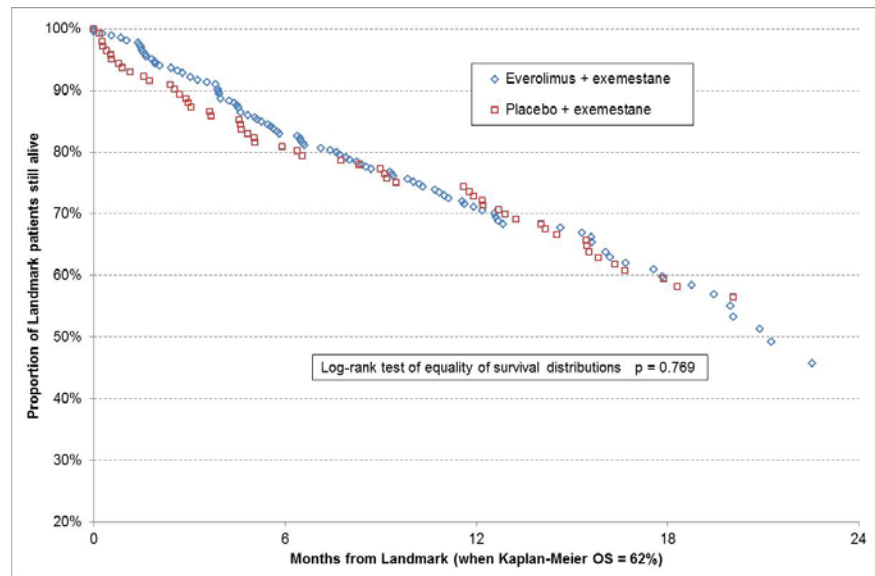


図 2-10 各群で生存率 62%の時点以降の長期生存患者に対するランドマーク解析 (TA421 committee-papers Figure 3より引用)

- 図 2-10において、2つの群の分布は視覚的に類似しており、全生存時間の分布の異質性に関するログランク検定によるp値は 0.769 であった。2群間の生存時間が異なることの統計的根拠は示されなかったため、ランドマークのポイントを超えた全ての患者に、共通の生存時間モデルが適用された。
 - エベロリムスに起因する生存時間の違いについてはパラメトリックな生存時間のモデリングに頼らなくても、試験結果から直接推定できる。具体的には図 2-9のB点で生存しているエベロリムス群の62%の長期生存率が、C点でのプラセボ群の62%の長期生存率と同一であると見なすことができるため、生存時間の群間差はエベロリムス群のA点からC点までの生存曲線下の面積とプラセボ群のA点からB点までの生存曲線下の面積の差で計算できる。
- 全生存期間をモデル化するための外挿の開始点を2群で共通にすることで、開始点の違いによりバイアスが混在するリスクを排除できる。そのため、全生存時間のモデル化の開始点を共通の62%とすることが望ましい。図 2-11 に2群を併合した長期生存者データの Kaplan-Meier ランドマーク分析の結果を示す。
 - 決定係数R²が0.993であることから、指数パラメトリックモデルのあてはまりはよいと考えられる。

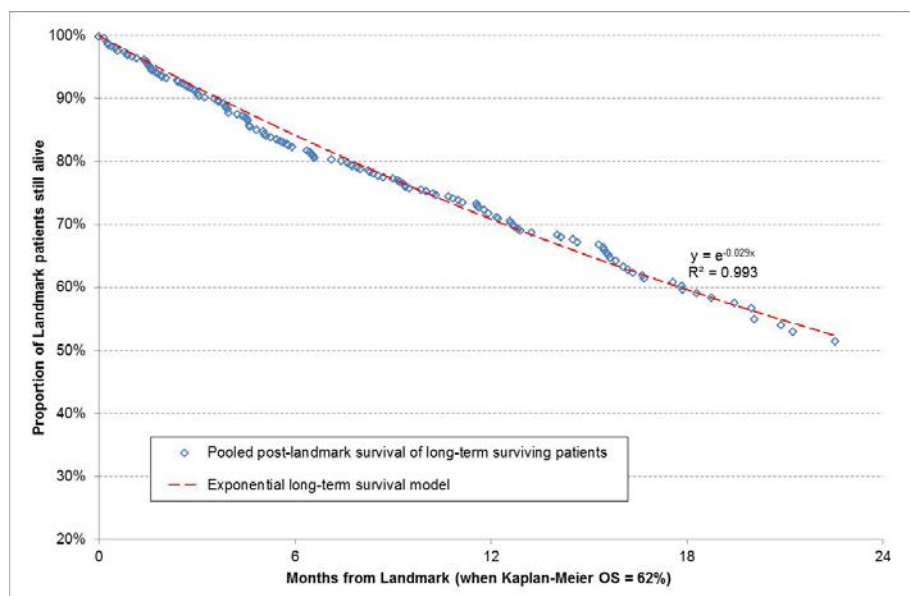


図 2-11 2群を併合した（ランドマークを超えた）長期生存者データにあてはめた指数パラメトリックモデル（TA421 committee-papers Figure 4より引用）

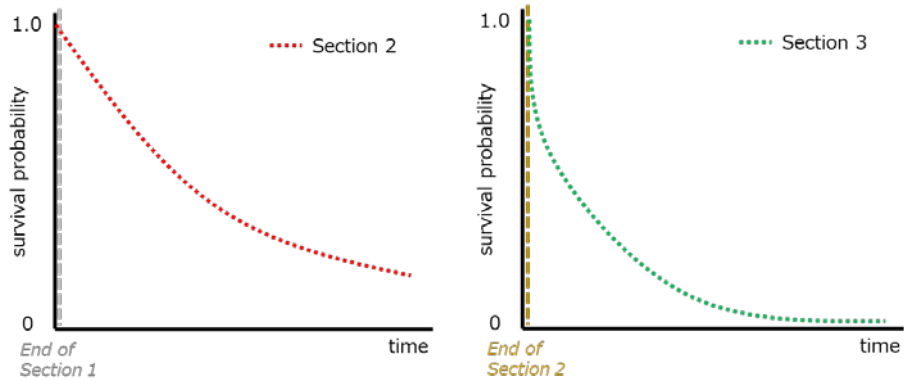
評価委員会の判断

- 企業分析が統計的に最適であると考えて適用した対数ロジスティック関数を用いたモデリングについて理解したものの、ERG が疾患の進行後および治療の中止後のエベロリムスからのベネフィット（平均余命の延長）はないという仮定に基づくランドマーク解析を使用して、全生存期間を異なる方法でモデル化したことも理解した。その上で、評価委員会は ERG のアプローチが合理的であり、ランドマークのアプローチが BOLERO-2 試験の結果を反映した生存延長の合理的な近似を提供する可能性が高いことに同意した。全生存期間をモデル化するための企業分析と ERG の両方の方法は ICER への影響は小さいが、企業分析の方法よりも ERG の方法で全生存期間をモデル化の方が好ましいと結論付けた。

2.4. 区分モデル

区分モデル (Piecewise Models) は、標準的なパラメトリックモデルがあてはまらない状況や異なる期間に複数のデータソースが存在する状況で適用される。ハザード関数を経時的に調べ、ハザード関数に変化が生じているか否か、発生している場合はその発生箇所を特定し、ハザード関数ごとに区分に分け、区分ごとに別の生存時間モデルをあてはめる。

<p>概要</p>	<ul style="list-style-type: none"> ● 区分モデルの定義 <ul style="list-style-type: none"> ➤ 時点 $0 = t_0 < t_1 < \dots < t_j = \max(t)$ により時間を J 区間に分割する。 ➤ j 番目の区間 $[t_{j-1}, t_j)$ の t に対するハザード成分 $\lambda_j(t)$ を指数関数や他のパラメトリックモデルで定義する。 ➤ 全体的なベースラインハザード関数を $\lambda_0(t)$、j 番目の区間のハザード関数を $\lambda_j(t)$ と定義する。 ➤ 全体的な生存関数は、累積ハザード関数 $\Lambda_0(t)$ に関連している。 生存関数: $S(t) = \exp(-\Lambda_0(t))$ ● 図 2-12の区分モデルのように、J 個の区分毎に様々なパラメトリックモデルで定義された条件付き生存関数を掛け合わせることで説明できる。 <ul style="list-style-type: none"> ➤ 生存関数の最初の区分 (セクション1) はKaplan-Meier曲線の生存関数を適用しており、セクション2, 3に様々な生存時間モデルを適用し、全体の生存関数が構成される。 <div data-bbox="606 1321 1133 1680" style="text-align: center;"> </div>
-----------	--



時間(t)での生存確率

= セクション3の時間(t)での生存確率 ×

(セクション1の最後の生存率 × セクション2の最後の生存率)

図 2-12 区分モデルの概要 (TSD21 Figure 6 を一部改変)

- 区分モデルの仮定
 - 別のモデルをあてはめる時点, すなわち区分の設定が適切である.
 - 各セクションで使用される生存時間モデルが適切である.
 - 頑健な生存時間のモデル化のため, 各セクションのデータ量が十分である.

限界

- 区分を任意に設定できるため, 設定の仕方が分析結果に大きな影響を与える可能性がある.
- ハザード関数が区分間で不連続である場合, 区分モデルは臨床的に不適切かもしれない. 例えば, 図 2-12に示されている生存時間モデルのハザード関数をプロットした場合, 生存時間モデルの結合点(図のセクション2とセクション3の境目)でハザードにかなりの「ジャンプ」が見られる.
 - 変化点で区分ハザード関数の連続性を可能にする方法を適用することも可能であり, スプラインの使用などハザード関数を柔軟に平滑化する他の方法と類似している(2.1節参照).
- 単一のデータセットに区分モデルをあてはめる場合, 前半のセクションと比べ後半のセクションで例数が減少する. これはリスク集団の患者数が少なく, 観察されるイベント数が少ないセクションでの課題であり, 生存時間モデルのあてはめでは大きな標準誤差と不確実性が生じる. さらに重要な点は外挿に使用されるモデルは最後のセクションにあてはめられたモデルであり, イベント数が少ないことが想定される.
- 各セクションで使用される生存時間モデルが適切でない場合, 予測全体に悪影響が及ぶ. これは最後のセクションで特に重要である.

<p>課題</p>	<ul style="list-style-type: none"> ● 通常、区分モデルでは、複雑なハザード関数をより正確に表すため、生存時間データの区分毎に標準的なパラメトリックモデルを適用する。データが利用できない期間にハザードの変化が予想される場合は、より柔軟なモデルの適用や、外部データを使用も可能である。ただし、外部データを使用する場合は、時点毎のリスク集団の共変量プロファイルを考慮する必要がある。外部データの使用に関しては3章を参照されたい。
<p>事例</p>	<p>化学療法適応前の転移性ホルモン再発前立腺がん治療のためのアビラテロンの技術評価ガイダンス（NICE TA387 Published: 27 April 2016 Last updated: 27 July 2016）の事例[37]を以下に示す。</p> <ul style="list-style-type: none"> ● COU-AA-302試験（転移性ホルモン再発前立腺がんの患者を対象とした、アビラテロンと経口プレドニゾン/プレトニゾロン(以下、abiraterone acetate plus Prednisolone：AAP)とプラセボと経口プレドニゾン/プレトニゾロン(以下、best supportive care：BSC)のランダム化比較試験；n=1088）に対して、治療期間に区分モデルを検討した。 <p><u>企業分析</u></p> <ul style="list-style-type: none"> ● 治療期間についてのKaplan-Meier曲線、及び対数ロジスティックモデルによる予測曲線を図 2-13に示す。 <div data-bbox="405 1196 1350 1778" data-label="Figure"> </div> <p>図 2-13 治療期間：Kaplan-Meier曲線及び対数ロジスティックモデルによる予測（TA387 committee-papers Figure 3より引用）</p>

- 図 2-13より、AAP群では対数ロジスティックモデルにより予測された生存率がKaplan-Meier曲線に適切にあてはまっているのに対して、BSC群ではKaplan-Meier曲線が少し過大推定されている（図中の青線に対する黄緑の線）。
 - つまり、群間差を小さく見積もる保守的なアプローチになっている。
- その後、対数ロジスティック分布が裾の長い分布であることに関する評価委員会の懸念や、ERGがワイブル分布による外挿を推奨していることを考慮し、2つの部分からなるモデル（対数ロジスティック分布 + ワイブル分布）を適用した。
 - 具体的には、BSC群の治療期間の範囲（おおよそ2.5年）の外挿点まで最もあてはまりのよい対数ロジスティック関数に基づいて推定し、その時点以降、AAP群はワイブル関数によって推定し、BSC群の患者は2.5年の追跡以降は治療中止と仮定した（図 2-14）。

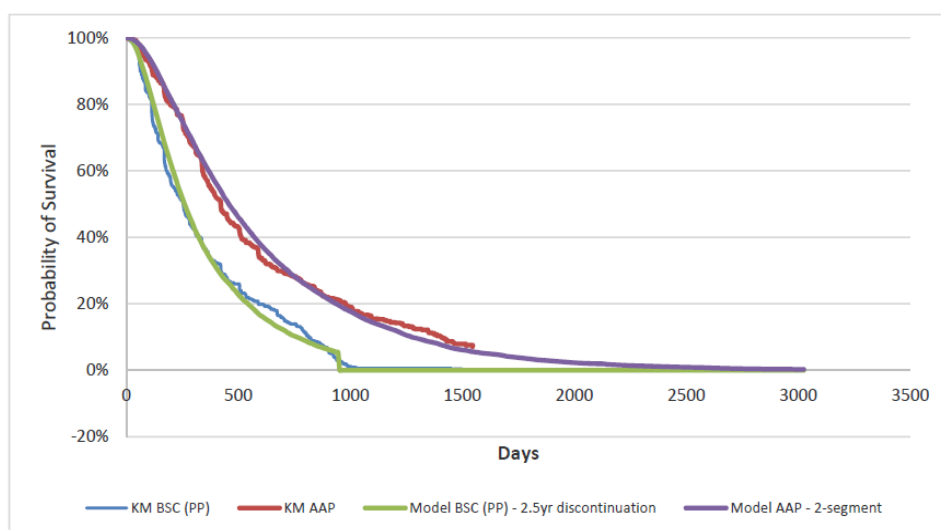


図 2-14 治療期間：Kaplan-Meier曲線及びモデル予測
 (対数ロジスティックモデル分布 + ワイブル分布による区分モデル)
 (TA387 committee-papers Figure 4より引用)

- 図 2-14に示すように、2.5年まで対数ロジスティック関数により予測し、残りの期間をワイブル関数により予測したAAP群の治療期間、2.5年まで対数ロジスティック関数により予測し、その後全患者が2.5年で中止するとしたBSC群の治療期間は、最終的なKaplan-Meierデータとほぼ一致した。

評価委員会の判断

- 評価委員会は、対数ロジスティックモデルを適用すべきか、区分モデルを適用すべきかの明確な回答は避け、モデルの結果に対する様々なパラメトリック分布の使用による ICER への影響を調査することが適切であると結論付けた。

2.5. 治癒モデル

治癒モデル (Cure Models) は、特定疾患による死亡などのイベントに関心があり、一定の割合の患者がそのイベントを発現しないと想定される状況で適用される。ある時点を過ぎると、そのイベントを発現する可能性のある患者が存在せずハザードはゼロとなり、生存関数はゼロ以外の値で横ばいになる (図 2-15)。

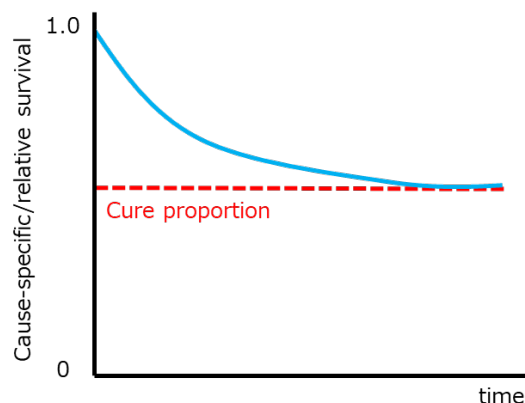


図 2-15 治癒モデルの概要 (TSD21 Figure 7 を一部改変)

がん治療の領域では、治癒モデルにより治癒割合を推定し、未治癒の患者の生存率を推定できる。生存関数がゼロに到達する全原因による死亡は適用対象外となる。

概要	<ul style="list-style-type: none"> ● 治癒は個々の患者レベルではなく集団レベルで定義される。 <ul style="list-style-type: none"> ➢ 集団全体の関心のある疾患特有のハザードはゼロに減少する。 ● 治癒モデルは、未治癒の集団のハザード関数と治癒の集団のハザード関数を組み合わせることで、複雑な形状を持つ全体的なハザード関数を推定できる。 ● 混合治癒モデルの定義 <ul style="list-style-type: none"> ➢ もっとも一般的な治癒モデルは、特定の疾病が治癒したグループと治癒していないグループが存在すると考える。 $S(t) = \pi + (1 - \pi) S_u(t) \quad (1)$ <p style="margin-left: 40px;">π: 治癒した患者の割合</p> <p style="margin-left: 40px;">$S_u(t)$: 未治癒の患者の生存関数</p> ➢ 式(1)にさらに他の原因による死亡率 (期待される生存率) を考慮する必要がある。 <ul style="list-style-type: none"> ◇ 期待される生存率をモデルに組み込み、相対的な生存率モデルを適用する；治癒の定義は「死亡率が一般の人々の死亡率と同じに
----	---

	<p>なる」ことになる。</p> $S(t) = S^*(t) (\pi + (1-\pi) S_u(t)) \quad (2)$ <p>$S^*(t)$: 期待される生存関数</p> <ul style="list-style-type: none"> ● 非混合治癒モデルの定義 <ul style="list-style-type: none"> ➤ 生存関数を以下のように定義する。 $S(t) = \pi^{F_z(t)} = \exp [\ln(\pi) F_z(t)]$ <ul style="list-style-type: none"> ◇ $F_z(t)$は累積分布関数で、通常、ワイブル分布や対数正規分布などの単純なパラメトリック分布を利用する。 ◇ 予想される生存情報を取り込んだ混合治癒モデルに類似な方法として拡張可能である。Edson Z et.al [38]では、分布関数に一般化修正ワイブル分布などを利用した場合について、非混合治癒モデルによる生存時間の推定結果は、他の原因による死亡率を考慮しない混合治癒モデルと同様であったことが紹介されている。 ● 治癒モデルの仮定 <ul style="list-style-type: none"> ➤ 治癒割合の推定に十分な利用可能なデータを有する。 ➤ ある一定の治癒割合が存在する。 <ul style="list-style-type: none"> ◇ 治癒とは特定時点での合理的な仮定に基づくもので、集団での疾患特有のハザード 0 を意味しており、このような仮定が妥当である必要がある。 ➤ 未治癒の部分のモデル化が適切である。
限界	<ul style="list-style-type: none"> ● 治癒割合を確実に推定するためには、分布の裾のリスク集団に十分な患者数が必要である。さらに特定の最終フォローアップの後に原因特有のイベントはないと仮定されている。 <ul style="list-style-type: none"> ➤ 最終フォローアップでのリスク集団が小さい場合、治癒等の強い仮定の妥当性を主張するのは難しい。現実的な仮定でないときに治癒を仮定すると、不正確な外挿を導く可能性がある。 ● 治癒していない患者のモデル選択にAICやBICなどの尺度がよく使用されるが、AICとBICはイベントが多く発生するフォローアップの開始時でのモデルのあてはまりをより重視している。治癒モデルはリスク集団やイベントが少ない分布の裾に大きな関心があるため、注意が必要である。 <ul style="list-style-type: none"> ➤ そのため、生物学的妥当性に基づいて、治癒していない患者について適切なモデルを選択することが重要である。 ● 原因特有の生存関数がプラトーに近づく場合、ハザードの形状を捉えるため適切なアプローチにより原因特有のハザードはゼロに近い値で推定される。

	<ul style="list-style-type: none"> ➤ 治癒モデルでは、試験に基づく疾患特有の死亡率がもはや影響を及ぼさないと想定される時点特定するのに役立つ可能性がある。また、この時点は明確に正当化される必要がある。
課題	<ul style="list-style-type: none"> ● 治癒モデルは生存関数がプラトーになることを前提としている。 <ul style="list-style-type: none"> ➤ このプラトーは、すべての原因による生存に対する長期的なものではないため、治癒モデルからの結果の解釈には注意が必要である。 ● 追跡期間では患者数が限られており、ハザードがゼロでないものの低いことから、治癒のように見える場合がある。 <ul style="list-style-type: none"> ➤ これは特にランダム化比較試験下で考慮する必要がある。治癒モデルは一部の臨床環境では魅力的に見える場合があるが、長期生存を推定する際には患者集団の特性やその他の原因による死亡率を組み込むように注意する必要がある。
事例	<p>25 歳以下の再発又は難治性の B 細胞性急性リンパ芽球性白血病の治療薬 Tisagenlecleucel-T (チサゲンレクルユーセル、以下、tis-T) の技術評価ガイダンス (NICE TA554, Published: 21 December 2018) の事例[39]を以下に示す。</p> <ul style="list-style-type: none"> ● tis-T を評価した 3 試験を併合したデータ (N=193) の全生存期間に混合治癒モデルを適用した例を紹介する。 <ul style="list-style-type: none"> ➤ EFS (Event free survival) のモデルおよび、比較対照技術であるブリナツモマブ、サルベージ化学療法のモデルについての説明は割愛する。 <p><u>企業分析</u></p> <ul style="list-style-type: none"> ● 投与後 32 か月後から生存曲線がプラトーになったことなどを理由に、治癒効果が示唆されたとして、全生存期間に指数分布の混合治癒モデルを用いた。指数分布を選択した理由は AIC と BIC が最も低く、推定した治癒割合が併合データの治癒割合に近かったためである。投与後に治癒した患者は、MacArthur et al. [40]の標準化死亡比 (standardized mortality ratio: SMR) を用いて、年齢と性別を一致させた英国一般集団の死亡率に戻した。治癒しなかった患者は、投与時からパラメトリックな生存曲線に従うとした。 <p><u>Evidence Review Group (ERG)による分析</u></p> <ul style="list-style-type: none"> ● tis-T の全生存期間に対数ロジスティック分布による混合治癒モデルを用いた。企業分析と異なる理由は治癒割合の推定値が臨床試験の EFS の無イベント率を超えないようにしたためである。 ● さらに、企業分析では、患者の長期死亡率が一般集団よりも低く、急性リンパ芽球性白血病の長期生存者の疫学研究で観察された値よりも低いことを

意味するモデルになっていたことを指摘した。

評価委員会

- ERG が最も低い治癒割合を推定する対数正規分布を用いたシナリオ分析を行ったところ、ベースケースである対数ロジスティック分布を用いた場合に比べて ICER がブリナツモマブとの比較で 16,567 ポンド、サルベージ化学療法との比較で 28,925 ポンド増加したことに評価委員会では言及し、推定されたそれぞれの治癒割合のばらつきは長期生存データの不足を反映していることを指摘した。
- また、tis-T が全生存期間に対して治癒効果を有するという前提を支持する頑健なエビデンスはないため、tis-T の全生存期間の外挿は、経済モデルにおける重要な不確実性の要因であると結論づけている。
- 最終的に、ERG のモデルで予測された長期死亡曲線がより妥当であると考え、ERG による企業モデルへの補正は適切であると結論づけた。

※日本の費用対評価効果報告書[41]によると、日本の公的分析班からは「チサゲンレクルユーセルによって長期生存が達成されるという点については、許容できるが、しかしそれは「(5年目以降に)再発が起こらないこと」によって達成されるものであり、cure point とされる5年目以降のEFS曲線を横一直線に外挿することや、5年目になった瞬間にOS曲線の外挿方法を切り替えることを正当化するものではない」と言及があり、比較対照技術であるブリナツモマブ群のOSは5年目から外挿するのではなく、時点0から推計する再分析をしていた。

2.6. 超過死亡モデル/相対生存時間モデル

超過死亡モデル (Excess Mortality Models) では、全原因死亡率を対象疾患による死亡率と他の原因による死亡率に分割する。超過死亡とは、関心のある対象疾患が原因の死亡のことを指す。臨床試験においても同様のアプローチが適用可能であり、また、管理された環境下のよう、適切に記録された死因情報の利用が可能な場合には、原因別モデルを構築することもできる。いずれの場合も、分離した超過死亡にパラメトリックモデルを適用することができる。このアプローチは一般に関心のある死因に関連する共変量の効果を得るために用いられるが、競合リスクの解析にも利用できる。対象疾患による死亡と他の原因による死亡の時間の経過に伴う推移は大きく異なる可能性があるため、このアプローチは、長期にわたる外挿を行う際に特に有用と考えられる。つまり、これらの死亡率を別々にモデル化した後に統合することで、長期にわたる全原因死亡率を推定することができる。特に、対象疾患による死亡率をモデル化し、他の原因による死亡率と統合することで、複雑な形状のハザード関数を推定することができる。これは、ポリハザードモデルにおけるアプローチ (2.1項参照) と同じであるが、ここでは、ある集団における固定された死亡率を用いており、他の原因による死亡率に対する不確実性を有さない。

また、超過死亡モデルは、例えば、(臨床試験で除外されている高齢者など) 死因に関する情報が不足しているか信頼性が低いと考えられる、がん患者を対象としたレジストリデータに適用される。

概要	<ul style="list-style-type: none"> ● 超過死亡モデルの定義 <ul style="list-style-type: none"> ➤ 全原因死亡率 (ハザード) は以下のとおり 2つの要素で表すことができる。 $h_i(t) = h_i^*(t) + \lambda_i(t)$ <ul style="list-style-type: none"> ◇ なお、$h_i(t)$は全原因死亡率 (ハザード)、$h_i^*(t)$は年齢、性別、暦年 (およびその他の一般的な決定要因) で層別化された集団死亡率から得られた背景死亡率、$\lambda_i(t)$は超過死亡率とする。 ➤ 相対生存関数 (相対生存時間モデル) は以下のとおりに表すことができる。 $R_i(t) = S_i(t) / S_i^*(t)$ <ul style="list-style-type: none"> ◇ なお、$R_i(t)$は相対生存関数であり、全原因生存関数$S_i(t)$と背景集団 (背景死亡率の算出に用いる集団) における期待生存関数$S_i^*(t)$の比で表すことができる。 ➤ ここで、例えば、超過死亡率に対して、柔軟なパラメトリックモデルを当てはめることができる。また、死因情報が利用可能な場合、標準的なパラメトリックモデルを原因別データに当てはめることができる。ただし、全原因死亡によるハザード関数を適切に分離するためには、正確な死因情報が必要となる。
----	--

	<ul style="list-style-type: none"> ● 超過死亡モデルの仮定 <ul style="list-style-type: none"> ➤ 長期にわたる外挿には、外挿により得られた超過死亡率と、他の原因による死亡を考慮した背景死亡率が必要となる。RCT では一般に試験期間が短いため、死因の多くが対象疾患に関連しており、死亡率は一般集団よりも高くなるため、超過死亡率は全原因死亡率と同様になることが多い。超過死亡率を分離してモデル化する際には、2つの競合するハザード（対象疾患によるハザードとその他の原因によるハザード）に対して、以下の仮定を置く必要がある。 <ul style="list-style-type: none"> ◇ 他の死因によるハザードは、生命表を用いて、その関心のある集団に対して近似することができる。なお、それらは一般に、不確実性を有しておらず、固定された既知の死亡率とする。 ◇ 超過死亡率を分離することで、短期的なハザードの形状を捉えることができる。その後、外部のデータソースを用いて、他の原因による死亡率を再び導入することができる。 ◇ 対象疾患による死亡率とその他の原因による死亡率をそれぞれ適切に外挿することができる。
限界	<ul style="list-style-type: none"> ● 対象疾患による死亡率を分離するには、関連する外部集団を特定する必要がある。つまり、他の原因による死亡率を得るために、対象疾患の患者から成るコホートであり、現在検討している試験集団と交換可能な外部集団の特定が必要である。対象疾患によっては、喫煙やその他のライフスタイルの特性など、さらに多くの要因によるマッチングが必要な場合がある。 ● このアプローチは、大規模な集団ベース研究でよく用いられるもので、年齢や（競合する原因別ハザードに対する）その他の重要な決定要因の効果を考慮する必要がある。臨床試験では一般に症例数が限られているため、本アプローチの適用が困難な場合がある。 ● 試験期間中の超過死亡率の形状を把握するためには、複雑性を伴うモデリングアプローチを選択する必要がある。 ● 対象疾患による死亡率とその他の原因による死亡率を長期にわたり外挿する際には仮定が必要となる。つまり、対象疾患を含めて、原因別ハザード関数を長期にわたり適切に規定する必要がある。
課題	<ul style="list-style-type: none"> ● 対象疾患に関連する死因を正確に特定することに細心の注意が払われている臨床試験では、原因別モデリングアプローチの適用がより適切かどうかを検討する必要がある。一方、超過死亡アプローチは、例えば、最初から試験の対象から除外されている高齢者など、死因（死亡が対象疾患によるものか否か）に関する情報の信頼性が低いような集団ベースのデータを用いる場合に一般に好まれる。

	<ul style="list-style-type: none"> ● 臨床試験において死因情報を収集しているかどうかにかかわらず、経済モデルを構築するための外挿目的のため、対象疾患による死亡とその他の原因による死亡を別々にモデル化するという考え自体は、更なる検討を行う価値があるかもしれない。 ● 全原因死亡アプローチではなく、原因別アプローチを取ることで、イベント数が減少し、不確実性が増す可能性があることには注意を払う必要がある。ほとんどの試験では、対象となる疾患に関連する死亡が短期的な死亡の大半を占めると考えられることから、短期におけるモデルの当てはまりに大きな影響を及ぼすことはないと考える。
事例	<ul style="list-style-type: none"> ● TSD21 の筆者らが知る限り、本アプローチは医療技術評価ではまだ使われていない。なお、本アプローチとは、超過死亡を単に考慮するだけのアプローチではなく、上述のとおり、レジストリ等の外部情報を用いて、正式に超過死亡モデルを当てはめるアプローチを指す。

2.7. ベイズ流アプローチの適用

生存期間データのモデル化（あるいはHTA（Health Technology Assessment）の他の場面）でベイズ流アプローチを適用することの潜在的な利点は、様々なデータソースから得られたエビデンスを柔軟にモデル化できること、専門家や臨床家の主観的な事前の信念を正式に取り込むことができること、パラメータ推定やモデル構築に伴う様々な不確実性を考慮することができることにあり、そして、これらは最終的に関心のあるアウトカム（例えば、平均生存率や純（金銭）便益）に影響を及ぼすことになる[42]。主観的な事前の信念を正式に取り込むことは昔に比べて容易になりつつあり、例えば、HTAにおいて生存期間データに対するベイズ流アプローチの利用（外挿を含む）の提唱[43]や、ユーザーフレンドリーなソフトウェアの開発が行われている[44]。公表文献やNICEのTAにおいて、主観的な情報のある事前分布を用いた完全ベイズ法を用いた事例は比較的少ないものの、いくつかのHTAでは、介入法（薬剤）のクラス効果（class effect）を用いた「強さの借用」を含め、相対治療効果に対する主観的な事前の信念や外部情報の使用が検討されている[45][46][47]。また、外部情報の利用を含め、ベイズ法を用いて生存期間データを外挿するアプローチがいくつか提案されている。第一に、ベイズ法の柔軟性を利用して、モデルのパラメータには漠然事前分布（無情報事前分布）を仮定し、（対象とする試験から得られたデータを含め）異なる情報源から得られたデータを（尤度関数を通じて）モデル化する方法[48]、第二に、対象とする試験や対象集団と比較したときの外部データの尤もらしさに関する主観的な信念に基づくハイパーパラメータにより、外部データを用いてpower priorを仮定する方法[49]、第三に、外部情報や主観的な信念に基づき、モデルパラメータに対する事前分布を仮定したり、ベイズモデル平均化の枠組みの中で事前モデル確率の事前分布を仮定する方法がある[50][51]。主観的な信念を用いる場合は、専門家や臨床家が容易に彼らの信念を表現できる量に関する情報、例えば、特定の時点における生存割合に関する情報を引き出し、この情報を用いてモデルパラメータに対する暗黙の事前分布を仮定することで、本手法を適用できる[52]。生存メカニズム（つまりは、データ生成のメカニズム）に関しては、HTAではなく生態学に基づく設定となるが、競合する死因の尤もらしさに関する主観的な信念を含めることも検討されている[53]。

HTAにおける経済評価は、疾病集団に対する資源配分の決定に必要な情報を提供するために実施されるが、費用効果分析での生存時間解析自体は通常、試験データに対するモデルの当てはまりに焦点が置かれる。従って、ベイズ流アプローチの枠組みでモデルを当てはめることは、試験指向の分析から脱却するための一つの選択肢となりうる。

このように、ベイズ流アプローチには潜在的な利点がある一方で、今に至るまで適用事例や評価が限られているため、この分野では更なる研究を行うことが重要である。

3. 生存関数の外挿と外部情報の取り込み

一般的に臨床試験では追跡調査を際限なく実施することは難しく、多くの疾患ではフォローアップ終了の時点において、まだ生存している患者がかなりの割合を占めることから、ある介入の生涯にわたる影響を正しく評価するためには、フォローアップ終了後のデータに対し外挿が必要となる。これまでの章では、試験データの範囲内で観察されるハザード関数に対し、より複雑な生存時間モデルを適合させる様々な方法を説明したが、これらの手法の多くは、超過死亡モデルや治癒モデル等のように外部データと組み合わせることで、より適切となる。また、単に試験データの範囲内で組み合わせているのではなく、範囲外への外挿も行われている。

図 3-1 に外部データを用いた外挿のイメージ図を示す。主たる目的である対照群 (A) と介入群 (B) の生存率の差 (期待生存率の増分) の推定は、試験データの範囲内 ($t < t_1$) では、A と B の観察データに適合したパラメトリックモデルによって推定される。試験データの範囲外 ($t > t_1$) では、3.1 章に示すような仮定の下で外部データ (C) を組み合わせることで、適切な推定値の算出を試みる。

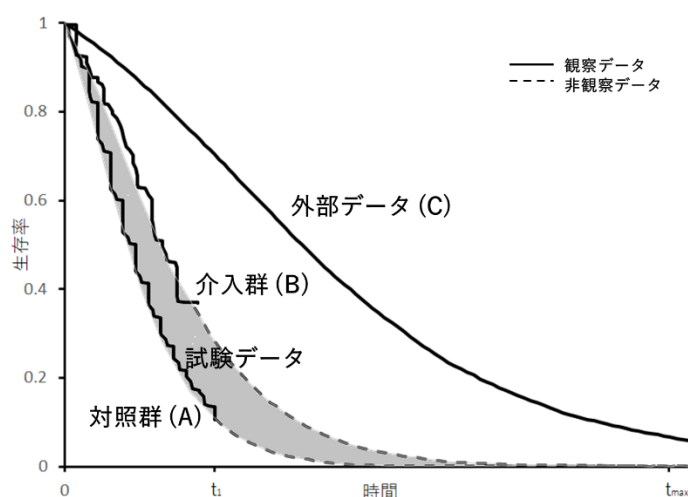


図 3-1 外部データを用いた外挿のイメージ図 (A : 対照群, B : 介入群, C : 外部データ)

統計モデルを用いた外挿プロセスで、ハザードや生存関数の形状を把握するには、以下のような検討が必要である。まずデータが利用可能なフォローアップデータの範囲内でのハザードの形状を捉えるために、可能な限り精緻なモデル化を検討すべきである。これによってデータの範囲内での絶対的なリスク (ハザード) の大きさを正確に予測し、外挿を適切な時点から始めることが可能となる。次に長期的なハザードの形状について考える必要があるが、通常、臨床試験では短期的な疾患による死亡率を捉えることを目的とするため、フォローアップの範囲内のデータに適合したモデルからそのまま外挿したデータには従わないことが多いと考えられる。フォローアップ終了時まで生存している患者は時間の経過とともに加齢しており、他の原因による

死亡の競合リスクの影響を受けている可能性が高い。また、調査対象の疾患による死亡リスクが介入により低下している可能性もある。たとえば、柔軟なパラメトリックモデルのような手法を用いると、TSD14[5]で示したような標準的で単純なパラメトリック手法よりも非現実的な長期的なハザードの形状を推定してしまうこともある。そのため、しばしば外部データの利用や加齢の考慮が必要となる。

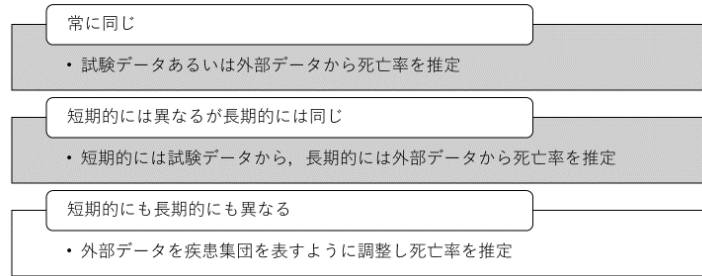
これまで医療技術評価のプロセスにおいては、臨床試験データに適合した生存時間モデルを用い、外部データを考慮せず、比較的長い期間を外挿してしまうことが一般的であった。外部情報が実際にモデル適合のプロセスに組み込まれた事例はほとんどなかったが、近年 Guyot らはモデルへの外部情報の組み込みを試みている[48]。生存時間モデルへの外部情報の組み込みに関しては3.1項で概略を示すが、長期間の外挿に使用する生存時間モデルを議論する際には、この点を十分配慮する必要がある。以下では、2章で示した生存時間モデルに対し、外挿可能性と外部データ組み込みの観点で考慮すべき点を示す。

3.1. 外部データの組み込み

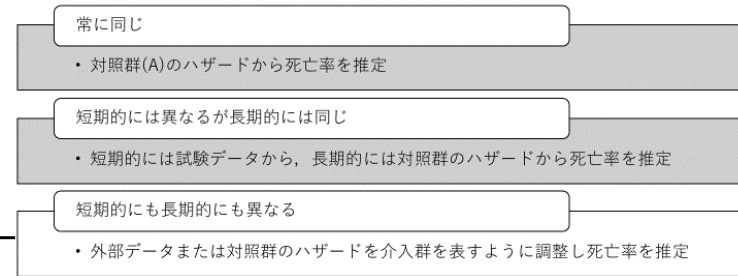
本章では、外部データを用いた生存時間の外挿のフレームワークを示す。外挿の手法の詳細については、Jackson らを参照されたい[54]。図 3-2 に生存時間の外挿に外部データを使用する際に取りうる選択肢を示す。対照群 (A)、介入群 (B) の試験データの範囲外の死亡率は、外部データ (C) と A, B 間のハザードの関係性に関する仮定の下で推定される。

まず、外部データ (C) が対照群 (A, 左上のパネル) と、対照群 (A) が介入群 (B, 右上のパネル) と、すべての期間、少なくとも長期的には、同じ死亡率であるかどうかを確認する必要がある。同じである場合は外部データをそのまま調整なしに A, B の長期死亡率の推定に利用することができる。そうでない場合は、A と C (あるいは B と C) の長期死亡率は異なるが、系統的には類似していると仮定し、外部データ (C) を用いて対象となる疾患集団 (A あるいは B) の長期死亡率を推定する。詳細は Jackson et al. [54] を参照されたい。

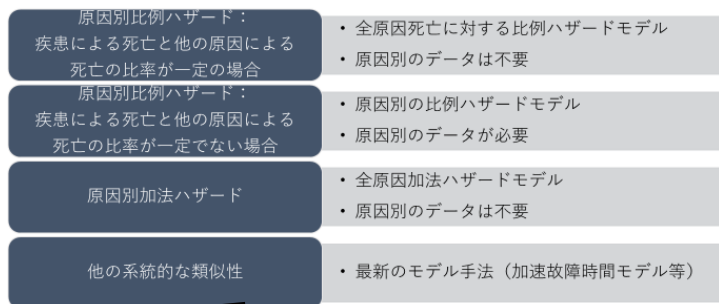
対照群(A)と外部データ(C)間の死亡率の関連性に基づく
対照群(A)の死亡率の推定



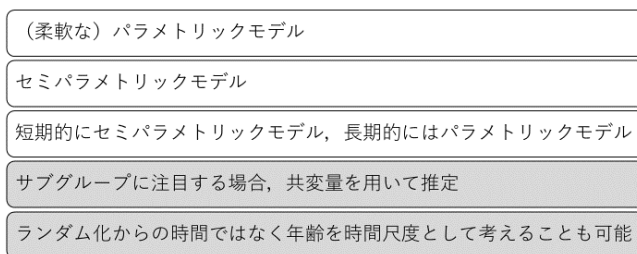
対照群(A)と介入群(B)間の死亡率の関連性に基づく
介入群(B)の死亡率の推定



試験データ(A, B)と外部データ(C)間の
長期的な死亡率の系統的な差に対する選択肢



ハザードの関数形式の選択



対照群(A)と介入群(B)の長期的な生存関数の推定

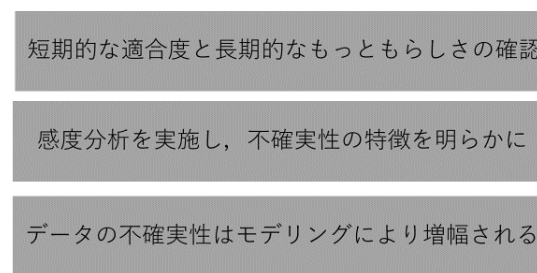


図 3-2 外部データを用いた生存時間データの外挿のモデル選択のフレームワーク

- 外部データの特定

外部データを特定するためには、関連する疾患レジストリを使用することが望ましい。このような疾患レジストリは同様の疾患と診断された患者を代表しており、臨床試験の対象集団を代表している可能性が高い。しかしながら、このような疾患レジストリには新規の介入治療を受けている患者は含まれていないため、臨床試験の対照群の長期死亡率のモデル化には有用かもしれないが、臨床試験の介入群の長期死亡率のモデル化には、その治療法の効果はどのくらい持続するのか、それは、その持続する時点までの対照群のハザード関数に比例して作用するものであるのか等、その有用性を慎重に検討する必要がある。

年齢、性別、暦年などの代表的な要因で層別されている人口動態統計の死亡率は容易に入手することができる。通常、臨床試験終了時に生存している患者のその後の死亡率が、人口動態統計から得られる一般の死亡率と同等であると見なすことは出来ないが、少なくとも外挿が一般の死亡率を下回ることがないことを示すために用いることができる。

- 疾患集団と外部データの死亡率の違い

対象疾患や介入が死亡率に影響を与えないと想定される場合（例えば、QOLのみに影響を与える場合）、患者の長期死亡率は年齢と性別が類似した国民の長期死亡率と同じであると仮定し、人口動態統計から直接取得することができる。疾患または介入が死亡率に影響を与えていたとしても、外部データが同じ疾患および/または介入を受けた患者の疾患レジストリまたはコホートから得られており、試験データの対照群の死亡率が外部データの死亡率と同じである場合には、患者の長期死亡率に外部データの長期死亡率を適用できる。一般の集団よりも疾患集団の初期の死亡率が高いが、その死亡率は徐々に減少し、ある時点で一般集団の死亡率に収束する場合も人口動態統計を同様に適用できる。

- リスクの種類

外挿の問題は競合リスクの問題と捉えることができる。リスクには、対象疾患に関連する死亡率と他の原因に関連する死亡率があり、前者は短期的に（試験の範囲内で）影響を及ぼす可能性が高く、後者は生存患者に対し長期的に影響を及ぼす可能性が高い。全原因死亡を対象疾患に関連する死亡と他の原因による死亡に分割し、他の原因による死亡率として人口動態統計の死亡率を用い、それを上回る超過死亡を対象疾患に関連付けることで、全原因生存時間モデルから疾患関連の死亡を分離することができる（超過死亡／相対生存時間モデルについては2.6項参照）。これは人口動態統計の死亡率が、その疾患で死亡しなかった患者の死亡率と一致しているという仮定に基づくが、併存疾患の影響が強く人口動態統計では死亡率が低すぎるような疾患では、この仮定は当てはまらない。死因に関する情報が得られるならば、標準的なパラメトリックモデルを用いて原因別死亡率を求めることも選択肢となりうる。全原因生存率の適切な周辺推定値（原因別死亡率）を得るためには、共変量のモデル化も必要である。

- 疾患集団と外部データの短期的および長期的な死亡率の違い

試験のフォローアップ期間において、患者集団の死亡率が外部データで表される集団の死亡率と異なる場合、対象となる集団を反映するように外部データを調整することで外挿が可能になる場合がある。このためには、長期的には集団間での死亡率の違いが系統的に説明できる（例えば、共変量等を考慮することで調整可能）という仮定が必要である。例えば、疾患集団と外部データの間で全原因死亡率や疾患関連の死亡率に比例ハザードや加法ハザードを仮定できる場合がある。

- 治療集団と対照集団の死亡率の差

介入群（図 3-2 の B）と対照群の間（図 3-2 の A）の死亡率の差についても同様の判断が必要である。介入が死亡率に影響しないと予想される場合は、B の生存関数はすべての時間において A の生存関数と等しいと仮定できる。もし、相対的な介入効果が試験のフォローアップ期間終了後すぐに消失すると予想される場合は、B のハザードは長期的には A のハザードに等しいと仮定できる。B の生存関数は、臨床試験から得られた相対的な治療効果と、外挿された A の生存関数を組み合わせることで推定することができる。このために必要な仮定は、疾患集団と外部集団との差を外挿するために必要な仮定と類似している。通常、全原因死亡または原因別死亡に対する治療群間のハザード比は一定であると仮定する。また、介入群と対照群の個々のデータが入手可能な場合、外部データと A の生存関数を推定した際と同様の方法を用いて、B の生存関数を A とは独立に作成することもできる。外部データがなくても A と B の生存関数は通常のパラメトリックモデルを用いて独立に推定できるが、治療群間の短期的な差が長期的な差を反映していることを仮定している。Bagust et al. [55]は、治療の作用機序に関する知識が、長期的な推定にどのように使われるかを論じている。例えば、ある薬剤の効果は治療開始後効果が表れるまでに時間がかかり、治療を中止すると徐々に消失する場合が想定される。相対的な治療効果が、試験のフォローアップ終了時以降の時間の経過に伴ってどのように変化するかという仮定は、どの介入が好ましいかを定める重要な要因となる。したがって、この仮定に関する不確実性を考慮することは重要である。NICE [56]では、短期的には有効な治療法の長期的な効果について、悲観的、楽観的、折衷案的な仮定に対応する 3 つのシナリオで検討することを推奨している。例えば、効果をハザード比 $h_B(t) / h_A(t)$ で表現すると、フォローアップ終了時以降 ($t > t_{\max}$, ここで t_{\max} はフォローアップ終了時点) の効果は以下の 3 つの仮定を置くことができる。

1. 悲観的：効果はなくなり、 $t > t_{\max}$ では $h_B(t) / h_A(t) = 1$
2. 楽観的： $h_B(t) / h_A(t) = \exp(\beta)$ となり、すべての t に対して一定
3. 折衷案：長期的に徐々に減少し、 $h_B(t) / h_A(t)$ は $\exp(\beta)$ から 1 に向かって増加する

- 対象となる母集団を表すための外部データ調整

対象疾患の患者集団（いずれかの介入を受けている）と外部集団の長期死亡率が異なる場合、

長期の外部データを調整して A の生存関数を推定するために、試験データと外部データの全原因死亡率に対し、以下のいずれかの仮定を用いることができる。また、B の生存関数の推定にも同様の方法を用いることができる。

1. 比例ハザードの仮定

いくつかの論文 [57][58][59] では、人口動態統計の生命表から推定した死亡率に文献または文献と専門家の見解から得た一定のハザード比を乗じて、疾患関連の死亡率を求めている。これらの研究では、比例ハザード性を仮定している。すなわち、疾患集団と一般的な集団の間のハザード比は、時間の経過とともに一定であると仮定している。これは、ハザードの代わりに、短期間(例えば、状態遷移モデルのサイクルの長さ)の死亡確率が比例していると仮定することで、近似的に実装できる。多くの場合、グループ間の死亡率の比較はハザード比として発表されるため、比例ハザードの仮定は都合が良いが、全原因死亡率は比例ハザード性が成り立たない場合がある。全原因死亡率の比例ハザード性の仮定は、疾患関連の死亡率が人口動態統計の全原因死亡率に占める割合が時間の経過によらず一定である場合にのみ可能である。Benaglia et al. [60]は、この仮定が誤って適用された場合の様々な状況におけるバイアスの程度を推定した。原因別に比例ハザードモデルを実施する際には、人口動態統計から原因別死亡率の推定値を得られる場合が多い。全原因死亡率のハザード比と同様に、外部集団に対する疾患集団の疾患関連のハザード比は、文献から入手するか、試験データと外部データを短期的に比較して推定することができる。介入群の疾患関連ハザードも一定の治療によるハザード比を乗じることで推定でき、疾患関連死亡率に対する介入の効果を表す。

2. 加法ハザードの仮定

一部の論文では、試験データと外部データの間ハザード比を考えるのではなく、疾患集団が一般集団と比較して一定の加法ハザード性を持つと仮定した[61][62][63]。Demiris et al. [64]は、ベイズの枠組みで加法ハザードモデルを研究した。加法ハザードモデルの利点は、原因別のモデル化が不要となることが挙げられる。疾患関連ハザードに加法性がある場合、全原因加法ハザードモデルも成立し、疾患関連リスク差は全原因リスク差に等しいことになる。加法ハザードモデルの下ではリスク差は時間に依存しないため解釈は容易である。一方、比例ハザードモデルではリスク差はベースラインのリスクに依存する。

比例ハザードの仮定と加法ハザードの仮定のどちらが短期的に適合するかは、二重対数プロットのような診断プロット[55][65]によってデータから確認することができる。Breslow et al. [66]が議論したように、特別な場合として両方をモデルに組み込むこともできる。しかし、長期的にどちらかが適合するかは、データからは検証できない点に留意が必要である。

3.2. 複雑な生存時間モデルに対する外挿

上記で示した通り、原因別のハザードを別々に扱うアプローチは、他の原因による死亡率が長期的に影響を及ぼす可能性が高い場合には、外部データを用いた外挿が自然なアプローチである。しかしながら、他の原因による死亡率の推定値を得るために、疾患レジストリや人口動態統計の死亡率データと照合するには、患者レベルの臨床試験データが利用可能であることが望ましい。以下では、2章で説明した各手法を外挿の観点で検討する。外挿のために外部データを組み込むのは、区分モデルの最終区分に限定されるか、他の原因によるハザードを推定するために治癒モデルと組み合わせることが一般的である。しかしながら、外部データを他の標準的な生存時間モデルに組み込むことはできるため、外部データを推定に使用しない場合でも、外挿に使用される周辺ハザード／生存関数を正当化するには、他の原因による死亡のハザード関数を常に考慮すべきである。

3.2.1. 柔軟なパラメトリックモデル

柔軟なパラメトリックモデルは、試験データの範囲内で複雑なハザード形状を捉えるアプローチである。原理的には、本方法を用いると生涯にわたる分析期間で外挿することができるが、長期的なハザードが試験データの範囲内の最終死亡率によって決まり、非現実的な外挿となってしまうことを防ぐ手段はない。年齢の影響がハザードの増加に強く影響する場合は特に問題となる。柔軟なパラメトリックモデルの枠組みを用いて、超過死亡のモデリング手法（外部データの利用）を組み込むことが可能である（2.6項参照）。さらに、標準的なパラメトリックモデルと同様に、試験データと外部データの比較において、比例ハザード、加法ハザードを仮定することができる。

3.2.2. 混合分布モデル

試験データの範囲内では、混合分布モデルのアプローチも、より複雑なハザード関数を捉えるのに適しており、観測されたハザードの形状を良く近似することができる。しかしながら、混合分布モデルを用いた外挿も、長期的な死亡率に影響を与える外部データを考慮しなければ非現実的なものになってしまう可能性が高い。臨床的に意味のある患者サブグループに関する情報は、長期的な外挿のための情報となり得る。一般に、複雑なモデルを用いた外挿は、外挿された関数が安定しない可能性があるためリスクが高いと考えられる。混合分布モデルの構成要素は観察されたデータに最も適合するように定義されるが、より長期的なフォローアップ時間に対する外挿の妥当性を保証できないため、妥当な外挿が得られるかどうかの判断には外部データを使用すべきである。

3.2.3. ランドマークモデル

ランドマークモデルは、ランドマークの時点で生存している患者を、生存期間が類似しているサブグループに分けることを目的としている。外挿する際に、通常は標準的なパラメトリックモデルをランドマーク時点以降のサブグループごとに適合させる。そのため、標準的な生存時間モデルで外挿する場合と同様の制約がある。実際、予後の良いグループでは、残った患者が加齢するにもかかわらず非常に低いハザードが推定されてしまう可能性がある。外部データを利用して長期間を外挿する場合には、合理的に外挿されるかどうかを適切に判断する必要がある。

3.2.4. 区分モデル

区分モデルでは、生存期間をいくつかの区分に分割し、それぞれの区分に別のモデルを当てはめ、最終区分のモデルを用いて外挿する。この場合も、試験データの範囲を超えたハザードの変化は、長期的には加齢の影響を受ける可能性があるため適切に説明することはできない。さらに、イベント数が少ない区分に生存時間モデルを当てはめると、不確実性（ランダムな変動）の影響が大きくなってしまう。試験データの最終の区分に適合したパラメトリックモデルにそれ以前の区分のデータを組み込み、治療群全体の情報を用いて外挿することもできるが、明確にあり得ない外挿を避けるためには、外部データをモデル適合プロセスに組み込むことが望ましい。

3.2.5. 治癒モデル

治癒モデルには外部データが不可欠である。対象疾患以外の他の原因による死亡率が組み込まれていなければ、一定数の患者は死なないと予測されてしまうからである。治癒モデルは原因別データを適合させることができるが、他の原因による死亡率の情報なしに外挿することは無意味であり、全原因死亡率の推定値を得るためには他の原因による死亡率の情報を含めなくてはならない。原因別死亡率を用いた治癒モデルは、外部データで長期的な死亡率を説明する時点を設定するのに有効な手法である。

3.2.6. 超過死亡／原因別死亡（相対生存）モデル

競合リスクの枠組みを用いると、試験データの範囲であっても、原因別死亡率を検討することができる。しかしながら、外挿するためには、まず、他の原因による死亡率を再併合し全原因生存関数を推定しなくてはならない。これは人口動態統計の死亡率から予測可能である。人口動態統計に基づくがん死亡率を用い、全原因生存関数を外挿するために利用されている事例もある [67]。原因別モデルでは、治癒すると仮定するか、長期的な原因別死亡率がある時点で頭打ちとなり他の原因によるハザードで完全に説明できるようになると仮定する。このような外挿方法

により、長期の外部データが加齢に伴う死亡率の増加を捉えることができるという意味で、複雑な形状のハザード関数を考えることができる。この方法では、外部データはモデル化の段階から組み込まれているため、外挿にそのまま利用可能である。通常、長期的な他の原因別の死亡率を仮定する必要があるが、これは年齢の影響を考慮してモデル化することや、直近の死亡率をこれ以降の死亡率に反映することで対応可能である。

なお、日本のチサゲンレクルユーセルの費用対評価の報告書[41]では、以下の点が指摘されていた。

「製造販売業者の提出した分析によれば、チサゲンレクルユーセル群、ブリナツモマブ+/-同種 HSCT 群(以下、ブリナツモマブ群)およびイノツズマブ オゾガマイシン+/-同種 HSCT 群(以下、イノツズマブ オゾガマイシン群)それぞれにおいて、一定期間後の死亡率として標準化死亡比を用いて、OS(全生存)関数を外挿している。しかし、これは病状の進行した PD 患者における超過死亡が考慮されておらず、生存期間が過剰推計されており、OS 関数の適切な外挿方法とは言えない。例えば 15 歳以上 25 歳以下の集団では 3 年時点でキムリア治療後に再発している人の平均余命はおよそ 10 年以上となっている。」

利用する外部データが評価対象集団の患者特性を代表しているかは留意が必要である。

3.3. 外部データを利用する際の考慮すべき事項

利用する外部データの適切性

すべての方法で、外部データを用いることで、対象疾患患者の長期の死亡率を完全に把握できると仮定する必要がある。外部データを使用する際に考慮すべき点は以下の通りである。

- すべての患者において十分に長期間のデータがあり、外部データがどれだけ生存情報を反映するか
- フォローアップ終了時点に残っている患者の背景情報
- すべての患者が死亡する時点までに必要な時間の長さ
- 外挿が必要な患者の割合

臨床試験のサンプルサイズ

臨床試験において考慮すべき要因の一つは、サンプルサイズの大きさであり、特にフォローアップ終了時点に残っている患者数が少なくなることは、生存データに適合した曲線に長期的な生存関数を組み込む(例えば区分モデル)等、異なる生存時間モデルを外挿に用いることを検討する際に、少ない患者で推定することによる不確実性(ランダムな変動)による影響が大きくなり、長期的な生存関数の推定に影響が出る可能性がある。

試験データの患者特性の観点での代表性

すべての試験データで考慮すべき事項であるが、生存関数を外挿する際に特に注目すべき点

は、試験データが潜在的に母集団を反映しているかどうかである。相対的な効果は母集団を反映していなくても同程度であるかもしれないが、費用効果分析で重要となる絶対的な生存関数と平均生存率は異なる場合がある。このことは、生存時間モデルを構築するために試験データをどのように使用するか、また外挿する際にどの外部データを使用するかの両方を考慮すべきことを示唆している。

外挿のための臨床試験の患者レベルのデータの入手可能性

外部データを組み込む際には、少なくとも年齢や性別などの患者特性に関するマッチングが必要となるため、対象疾患の臨床試験の患者レベルのデータを利用することが望ましい。対象疾患の患者レベルの外部データが入手できない場合、臨床試験に参加した患者と外部データの患者を1対1でマッチングさせることができない。代替案としては、臨床試験に参加した患者と類似した特性（年齢や性別等）を持つ外部データの患者の周辺ハザードデータを入手することが考えられる。この外部データは、外挿のための死亡率の情報として、臨床試験のフォローアップ終了時点以降も使用することができる。

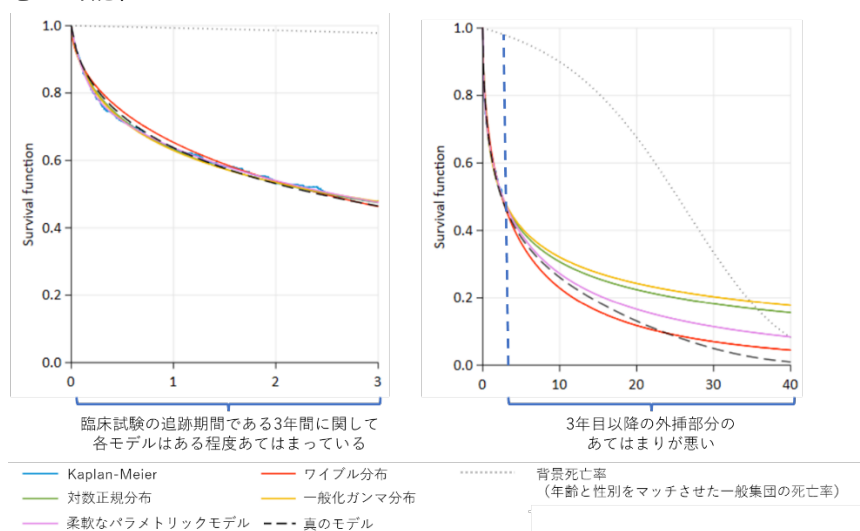
4. TSD 21 のシミュレーション研究と推奨事項

4.1. TSD21 のシミュレーション研究

TSD21[7]では、これまでの技術評価ガイダンスを参考に生物学的に妥当と考えられるシナリオを設定し、シミュレーション研究を行っている。シミュレーション研究では、各シナリオに対して様々な生存時間モデルをあてはめ、試験の追跡期間終了時の境界内平均生存時間と外挿に基づく全平均生存時間を評価した。生存時間に影響を与える治療の経済的評価を行う際、可能な限り正確に外挿することが極めて重要であるため、シミュレーションは治療効果の推定よりも単一の治療群での外挿の妥当性を評価することを重視した。シミュレーション研究の概要を以下に示す。なお、シミュレーション研究の詳細はTSD21を参照されたい。

どのようなシミュレーション研究も、得られた知見は設定したシナリオに限定される点に留意が必要である。TSD21では32通りの異なるデータ生成メカニズムを設定し、背景死亡率にはゴンペルツ分布、疾患特異的死亡率には2成分ワイブル混合分布と2成分ワイブル混合分布では説明しきれない未知の影響を組み込んだモデルを用いてシミュレーションを行った。その結果、評価した生存時間モデルの中で、全ての状況において普遍的に適用可能なモデルはなかった。主な結果は以下の通りである。

- 臨床試験から得られたデータの範囲内であてはまりがよい場合でも、外挿部分ではあてはまりが悪い可能性がある。



- 背景死亡率を考慮する (外部情報を考慮する) ことで外挿を改善する可能性がある。
 - ただし、試験の追跡期間中に疾患特異的死亡率に関する十分な情報が得られていない場合には深刻なバイアスが残る可能性がある (詳細は3.3項参照)。例えば、区分モ

デルは特徴的なハザードの形状に関して柔軟な対応が可能だが、外挿が始まる最後の区分においてサンプルサイズが問題となる可能性が高い。

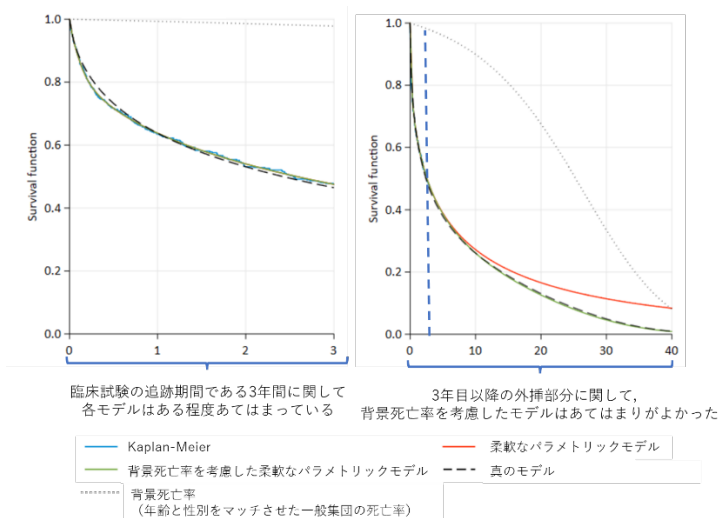


図 4-2 背景死亡率を考慮した柔軟なパラメトリックモデル(TSD21 Figure 22,23 を一部改変)

- 治癒を想定したシナリオにおいて、治癒モデルは平均生存時間の妥当な推定値が得られたが、治癒を想定しないシナリオでは他の方法が比較的良好に機能するシナリオにおいても、治癒モデルは大きなバイアスをもたらす可能性がある。
- 疾患特異的なハザードの変曲点を臨床試験の追跡期間内で捉えられていない場合、外挿部分で大きなバイアスが入る可能性がある。

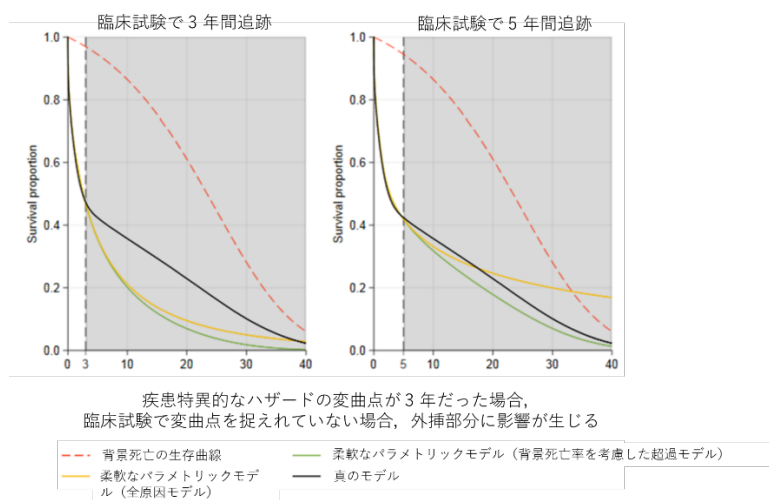


図 4-3 ハザードの変曲点に関する検討 (TSD21 Figure 26 を一部改変)

4.2. TSD21 の推奨事項

TSD21 では、シミュレーション研究を通して、生存時間の外挿のモデリングに関して、以下の推奨事項を提示している。

- 予測した生存関数とハザード関数をプロットする。
 - 試験データが存在する期間と外挿した期間に関して、生存関数とハザード関数を常に視覚化する。特に外挿した期間に関して、モデルの仮定とハザードの変化を把握することは重要である。さらに、予測したハザード関数及び生存関数が妥当と考える理由も示すべきである。
- 一般集団に対する生存関数とハザード関数をプロットする。
 - 予測した生存関数とハザード関数を、一般集団に対する生存関数とハザード関数と比較することで、予測した結果が信頼できるか否かを判断することを助ける。
- 背景死亡率を考慮する。
 - 非現実的な予測を回避するために生存時間モデルに背景死亡率を組み込むことが推奨される。これは標準的なパラメトリックモデル、柔軟なパラメトリックモデル、混合分布モデル、ランドマークモデル、区分モデルに当てはまり、治癒モデルでは必須である。
 - 背景死亡率は、外挿する際にモデルに組み込むか、外挿した周辺生存関数、特に周辺ハザード関数をプロットする際の確認として用いることができる。留意点として、費用対効果評価の分析対象集団が背景死亡率の集団よりも多くの併存疾患を有する場合、期待される生存率は過小評価されるであろう。ただ、用いた背景死亡率が不適切であったとしても、この点を考慮することは背景死亡率を全く考慮しないモデルよりも外挿は良好である可能性は高い。
 - 超過死亡率を 0 と仮定する場合、つまり予測したハザード率と背景死亡率が同じと仮定する場合はその開始時点を記載する。
- その他の外部情報を考慮する。
 - レジストリデータのような外部情報を生存時間モデルに組み込むことは有用かもしれない。しかしながら、この部分の研究は現在進行中であり、確固たる推奨を行うことはできない。
 - もし関連しそうなレジストリデータが特定できた場合には、臨床試験の対象集団、費用対効果評価の分析対象集団、レジストリデータの対象集団を比較し、もし可能であれば、費用対効果評価の対象集団に近い患者のレジストリデータを用いるべきである。利用可能な外部情報のデータソースは早い段階で検討し、この情報を組み込んだ解析を解析計画書で事前規定する必要がある。
- 外挿時の治療効果の仮定を明示する。

- TSD21 では、単一の群に対する外挿に焦点を当てていた。2群について外挿する際、両群のハザード関数が意味する治療効果をプロットで表示し、比例的な治療効果を想定したのか、各群にそれぞれ生存時間モデルをあてはめたのか等を明確に説明すべきである。レジストリデータのような外部情報は、対照群の外挿に最も用いやすいだろう。試験治療群の外挿は背景死亡率の情報を含むべきだが、長期の治療効果に関しては3.1項に示した通り、悲観的、楽観的、折衷案の3つの観点からの感度分析も考慮すべきである。

また、個別のモデルについても以下の推奨事項を提案している。

モデル	推奨事項
標準的なパラメトリックモデル	長期的なハザードが単純な形に従うことが期待される場合には妥当な外挿を与える。しかし、これがあてはまらなケースは多いだろう。そのような場合には背景死亡率又は他の関連する外部情報を組み込むことを常に考慮する必要がある。
柔軟なパラメトリックモデル	観察された生存時間データによくあてはまる可能性が非常に高いが、適切に外挿できない可能性がある。そのため、背景死亡率又は他の関連する外部情報を組み込むことを常に考慮する。
混合分布モデル	直観的に理解しやすいかもしれないが、しばしばモデルが収束せずに悩み、誤った解釈を引き起こす可能性がある。そのため、混合分布モデルは細心の注意を払って使用すべきである。また、背景死亡率又は他の関連する外部情報を組み込むことを常に考慮する。
ランドマークモデル	ランドマークのカテゴリの分類を正当化する必要がある。ランドマーク時点の選択に伴う結果の頑健性を考慮すべきである。ランドマーク後にモデル化したハザードと生存時間を信頼区間と共に示し、各カテゴリに当てはめたモデルに関する不確実性を明確に示すべきである。 さらに、各カテゴリにあてはめるモデルそれぞれについて、背景死亡率又は他の関連する外部情報を組み込むことを考慮することを常に考慮する。
区分モデル	区分に用いた時点を正当化する必要がある。そのためには、選択した時点の選択に伴う結果の頑健性を考慮する必要がある。区分ごとにモデル化したハザードと生存時間を信頼区間と共に示し、各区分で当てはめたモデルに関する不確実性を明確に示すべきである。

	各区分にあてはめるモデルそれぞれについて、背景死亡率又は他の関連する外部情報を組み込むべきであり、特に外挿に利用される最後の区分にあてはめたモデルについては注意が必要である。
治癒モデル	治癒の想定が妥当な場合に有用である。重要なことは背景死亡率を組み込み、治癒後の死亡を考慮することである。治癒の存在が不確実な場合、治癒モデルを用いるよりも背景死亡率を考慮した柔軟なパラメトリックモデルを用いる方が適切な可能性が高い。 治癒率は特にサンプルサイズが少ない場合に不確実性が高くなりやすい。治癒モデルを用いる場合、推定した治癒率を正当化するためのエビデンスを提示すべきである。また、治癒した集団と治癒しなかった集団についてそれぞれ予測したハザード関数と生存関数を示す必要がある。

本領域は現在も研究中の事項が多く、TSD21では今後以下の研究が望まれるとしている。

- TSD21では様々な現実的な状況を想定し、シミュレーションにより生存時間モデルの性能評価を行ったが、想定した状況は網羅的ではないため、他のシナリオを調査することは有用だろう。
- 外部のデータソースを生存時間解析や外挿に組み込むことは様々な方法があるが、特定の状況に最も適した方法を探求するためにさらなる研究が必要である。
 - ▶ ベイズ流の方法はモデルの不確実性と共に、専門家の意見と外部データソースの両方を検討し、医療技術評価に統合する手段を提供すると思われる。
- TSD21で紹介した方法は、主に周辺ハザード関数と生存関数の変化を引き起こすメカニズムによって詳細にモデリングを行うわけではなく、周辺ハザード関数と生存関数をそのままモデリングを行い、外挿しようとした方法である。一方で、外部データの利用などいくつかのアプローチは、全体の周辺ハザード関数に影響し、競合するメカニズムを明らかにするだろう。競合する要因を直接モデル化し、全原因生存関数に影響を与えるメカニズムと共変量の効果を合わせて調整する方法は今後の研究で探求され、モデルの想定に対する感度分析が可能になるだろう。
- 試験治療の生涯のベネフィットに強く影響を及ぼす要因は、長期的な治療効果だろう。長期の相対的な治療効果の外挿は、類似の薬剤クラスからの情報や他の長期の臨床試験の追跡調査から情報を借りること、又は専門家の意見を引き出すことが含まれる。このアプローチの実行可能性を探求するにはさらなる研究と評価が必要である。
- 臨床試験の対象集団は、私たちが意思決定したい対象集団である費用対効果評価の分析対象集団を十分に反映していない可能性がある。また、外挿を実施する際にも同様の懸念があり、対象集団に適合するように再重み付けする (reweight to a target population) 方法 [68]は更なる研究、評価に値する。

5. おわりに

本報告書では、NICE DSU が公開した TSD21 「Flexible methods for survival analysis」 [7] をもとに、費用対効果評価における柔軟な生存時間のモデリング方法及び外挿方法について、概要、限界、課題、適用事例を紹介した。

TSD21 のシミュレーションが示している通り、あらゆる状況に対して普遍的に利用可能なモデリング方法は存在せず、各モデルの特徴と前提となる考え方を理解した上で使い分ける必要がある。また、外挿に関しては背景死亡率を考慮する等の工夫が必要だろう。本邦の費用対効果評価の本格導入においては、基本分析の結果に基づいて薬価調整が行われるが、基本分析の妥当性を説明するために生存時間に関しても様々な感度分析、シナリオ分析を行い、基本分析の結果の頑健性を確認することが重要と考える。生存時間をモデリングする際に本報告書がお役に立てば幸いである。

参考文献

- [1] 国立保健医療科学院ウェブサイト. 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. 2022.
https://c2h.niph.go.jp/tools/system/notification_cost-effectiveness.pdf
- [2] National Institute for Health and Care Excellence. <https://www.nice.org.uk/>
- [3] NICE Decision Support Unit. <http://nicedsu.org.uk/>
- [4] 福田敬. 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 3 版. 2022.
https://c2h.niph.go.jp/tools/guideline/guideline_ja.pdf
- [5] Latimer N. NICE DSU Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. 2013.
<https://www.sheffield.ac.uk/nice-dsu/tsds/survival-analysis>
- [6] 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. 費用対効果評価における生存時間解析の諸問題. 2020.
https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000a1dj-att/c_e_a_problems_01.pdf
- [7] Rutherford M, et al. NICE DSU Technical Support Document 21: Flexible methods for survival analysis. 2020. (updated March 2022)
<https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis>
- [8] 国立保健医療科学院ウェブサイト. <https://c2h.niph.go.jp/>
- [9] 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. 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 –. 2016.
https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc0000007rux-att/2014ds_tf3.pdf
- [10] 鎌江伊三夫. 医療技術評価ワークブック. じほう. 2016.
- [11] 五十嵐中ら. 「薬剤経済」わかりません!! . 東京図書. 2014.
- [12] Drummond, M, et al., (監訳) 久繁哲徳ら. 保健医療の経済評価第 4 版. 篠原出版新社. 2017.
- [13] 大橋靖雄ら. 生存時間解析—SAS による生物統計. 東京大学出版会. 1995.
- [14] 大橋靖雄ら. 生存時間解析 応用編: SAS による生物統計. 東京大学出版会. 2016.
- [15] Nicholas R, et al. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
<https://www.sheffield.ac.uk/nice-dsu/tsds/treatment-switching>
- [16] Woods B, et al. NICE DSU Technical Support Document 19: Partitioned survival analysis for decision modelling in health care: a critical review. 2017.
<https://www.sheffield.ac.uk/nice-dsu/tsds/partitioned-survival-analysis>
- [17] 日本臨床腫瘍研究グループウェブサイト. 固形がんの治療効果判定のための新ガイドライ

-
- ン (RECIST ガイドライン) 改訂版 version 1.1—日本語訳 JCOG 版. 2009.
<http://www.jcog.jp/>
- [18] Jackson C. flexsurv: A platform for parametric survival modeling in R. J Stat Softw. 2016 May 12;70:i08.
- [19] Bennett I, et al. flexsurvPlus: Provides functions to perform survival analysis for economic models. R package version 1.06. 2022.
<https://roche.github.io/flexsurvPlus/index.html>
- [20] Benaglia T, et al. Mixtools: An R package for analyzing mixture models. J. Stat. Soft. 2009, 32, 1-29.
- [21] Jordan A. flexsurvcure: flexible parametric cure models, 2019.
<https://cran.r-project.org/web/packages/flexsurvcure/index.html>
- [22] Chao C, et al. smcure: Fit semiparametric mixture cure models. 2022.
<https://cran.r-project.org/web/packages/smcure/index.html>
- [23] Yingwei P. mixdure: mixture cure models. 2020.
<https://cran.r-project.org/web/packages/mixdure/index.html>
- [24] Dewar R, et al. A new SAS macro for flexible parametric survival modeling: applications to clinical trials and surveillance data. Clin Invest. 2015, 5(12), 855–866.
- [25] Royston, P, et al. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med, 2002. 21(15). 2175-2197.
- [26] Bower H. et al. Capturing simple and complex time-dependent effects using flexible parametric survival models: A simulation study, Communications in Statistics - Simulation and Computation, 2021. 50(11). 3777-3793.
- [27] Rutherford M, et al. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study, Journal of Statistical Computation and Simulation, 2015. 85(4), 777-793.
- [28] Syriopoulou E, et al. Robustness of individual and marginal model-based estimates: A sensitivity analysis of flexible parametric models. Cancer epidemiology, 2019. 58. 17-24.
- [29] Demiris, N, et al., Survival extrapolation using the poly-Weibull model. Stat Methods Med Res. 2015. 24(2). 287-301.
- [30] Liu, X.R, et al. Parametric and penalized generalized survival models. Stat Methods Med Res. 2018. 27(5):1531-1546.
- [31] NICE. TA417. Nivolumab for previously treated advanced renal cell carcinoma. 2016.
<https://www.nice.org.uk/guidance/ta417>
- [32] Motzer RJ, et al. Nivolumab versus everolimus in patients with advanced renal cell carcinoma: Updated results with long-term follow-up of the randomized, open-label, phase 3 CheckMate 025 trial. Cancer. 2020. 126(18). 4156-4167.
- [33] McLachlan GJ, et al. Finite mixture models. Annual Review of Statistics and Its Application, 2019. 6(1). 355-378.

-
- [34] Farewell, VT. Mixture models in survival analysis: Are they worth the risk?. *The Canadian Journal of Statistics*. 1986. 14(3). 257-262.
- [35] NICE. TA530. Nivolumab for treating locally advanced unresectable or metastatic urothelial cancer after platinum-containing chemotherapy.
<https://www.nice.org.uk/guidance/ta530>
- [36] NICE. TA421. Everolimus with exemestane for treating advanced breast cancer after endocrine therapy. 2016.
<https://www.nice.org.uk/guidance/ta421>
- [37] NICE. TA387. Abiraterone for treating metastatic hormone-relapsed prostate cancer before chemotherapy is indicated. 2016.
<https://www.nice.org.uk/guidance/ta387>
- [38] Edson Z, et al. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Programs Biomed*. 2013. 112(3). 343-55.
- [39] NICE. TA554. Tisagenlecleucel for treating relapsed or refractory B-cell acute lymphoblastic leukaemia in people aged up to 25 years. 2018.
<https://www.nice.org.uk/guidance/ta554>
- [40] MacArthur AC, et al. Mortality among 5-year survivors of cancer diagnosed during childhood or adolescence in British Columbia, Canada. *Pediatr Blood Cancer*. 2007. 48(4). 460-467.
- [41] 国立保健医療科学院ウェブサイト. [C2H1902] チサゲンレクルユーセル(キムリア). 2021.
<https://c2h.niph.go.jp/results/C2H1902.html>
- [42] Gong Q, et al. Asymptotic properties of mean survival estimate based on the Kaplan–Meier curve with an extrapolated tail. *Pharm Stat*. 2012. 11(2). 135-140.
- [43] Spiegelhalter DJ, et al. *Bayesian approaches to clinical trials and health - care evaluation*. John Wiley & Sons. 2004.
- [44] Baio G, et al. Survival analysis for health economic evaluation and cost-effectiveness modelling . *J Stat Soft*. 2020. 95(14). 1-47.
- [45] Soares, MO, et al. Experiences of structured elicitation for model-based cost-effectiveness analyses. *Value Health*. 2018. 21(6). 715-723.
- [46] Brard C, et al. Bayesian survival analysis in clinical trials: What methods are used in practice? *Clin Trials*. 2017. 14(1). 78-87.
- [47] Grigore B, et al. Methods to elicit probability distributions from experts: A systematic review of reported practice in health technology assessment. *Pharmacoeconomics*. 2013. 31(11). 991-1003.
- [48] Guyot P, et al. Extrapolation of survival curves from cancer trials using external information. *Med Decis Making*, 2017. 37(4). 353-366.
- [49] Ibrahim J, et al. *Bayesian Survival Analysis*. Springer. 2001.

-
- [50] Soikkeli F, et al. Extrapolating survival data using historical trial-based a priori distributions. *Value Health*, 2019. 22(9). 1012-1017.
- [51] Abrams K, et al. Propensity weighting and extrapolation in non-small-cell lung cancer. Work Package 1, IMI GetReal. 2016.
- [52] Cope, S, et al. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC Med Res Methodol*. 2019. 9(1). 182.
- [53] Walsh DP, et al. Using expert knowledge to incorporate uncertainty in cause-of-death assignments for modeling of cause-specific mortality. *Ecol Evol*. 2017. 8(1). 509-520.
- [54] Jackson, C, et al. Extrapolating survival from randomized trials using external data: A review of methods. *Med Decis Making*. 2017. 37(4). 377-390.
- [55] Bagust, A, Beale, S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Making*. 2014. 34(3). 343–51.
- [56] NICE. Guide to the methods of technology appraisal. 2013.
<https://www.nice.org.uk/article/pmg9>
- [57] Barton P, et al. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technol Assess*. 2004. 8(11). 1–91.
- [58] Rodgers M, et al. Colour vision testing for diabetic retinopathy: A systematic review of diagnostic accuracy and economic evaluation. *Health Technol Assess*. 2009. 13(60). 1–160.
- [59] Chen YF, et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technol Assess*. 2006.10(42). 1–229.
- [60] Benaglia T, et al. Survival extrapolation in the presence of cause specific hazards. *Stat Med*. 2015. 34(5). 796–811.
- [61] Fang CT, et al. Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *QJM*. 2007. 100(2). 97–105.
- [62] Chu PC, et al. Estimation of life expectancy and the expected years of life lost in patients with major cancers: extrapolation of survival curves under high-censored rates. *Value Health*. 2008. 11(7). 1102-1109.
- [63] Hwang JS, et al. Monte Carlo estimation of extrapolation of quality-adjusted survival for follow-up studies. *Stat Med*. 1999. 18(13). 1627–1640.
- [64] Demiris N, et al. Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. *Stat Med*. 2006. 25(11). 1960–1975.
- [65] Latimer NR. Survival analysis for economic evaluations alongside clinical trials: extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013. 33(6). 743–754.
- [66] Breslow NE, et al. Statistical methods in cancer research. Volume II: the design and analysis

of cohort studies. IARC Sci Publ. 1987. 82. 1–406.

[67] NICE. TA463. Cabozantinib for previously treated advanced renal cell carcinoma. 2017.

<https://www.nice.org.uk/guidance/ta463/>

[68] Rutherford M, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016.

<https://www.sheffield.ac.uk/nice-dsu/tsds/population-adjusted>

2022 年度データサイエンス部会 KT7

作成担当者

渥美 淳	東レ株式会社
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*
町田 光陽	塩野義製薬株式会社
吉田 瑞樹	ファイザーR&D 合同会社*

レビュー担当者

東 美恵	エーザイ株式会社
荒西 利彦	日本イーライリリー株式会社
奥山 ことば	MSD 株式会社*
直井 一郎	住友ファーマ株式会社
野島 俊秋	興和株式会社

*タスクフォースリーダー

監事

酒井 弘憲	エーザイ株式会社
-------	----------