

一般化ペアワイズ比較による生存時間解析

福田 武蔵*, 坂巻 顕太郎†, 大庭 幸治‡

Generalized Pairwise Comparisons for Survival Analysis

Musashi Fukuda*, Kentaro Sakamaki† and Koji Oba‡

生存時間を評価項目とするがん領域や循環器領域の医学系研究で、治療効果の指標として Net Benefit や Win Ratio が利用されている。これらは一般化ペアワイズ比較に基づく指標として整理できる。一般化ペアワイズ比較では、新治療と標準治療を比較するランダム化比較試験の場合、各群から 1 人ずつをランダムに選んでできるペアに対する勝ち負けの確率を評価する。このとき、評価項目の大小関係に基づく様々な勝ち負けのルールを考えることができる。Net Benefit は勝ち負けの確率の差、Win Ratio は比として定義される指標である。本稿では、勝ち負けの確率、Net Benefit と Win Ratio の推定方法を一般化ペアワイズ比較の観点から整理し、 U 統計量理論から導かれる漸近分散を説明する。また、打ち切りを含む生存時間における推定の問題と対処方法を説明する。最後に、R パッケージによる実装方法を例示する。

The Net Benefit and the Win Ratio are the effect measures used in clinical trials with survival endpoints such as oncology and cardiology. These measures are based on generalized pairwise comparisons. In a randomized trial, the statistical methods estimate the ‘win’ probability that a subject randomly selected from the new treatment group has a better outcome than a subject from the control group and the ‘loss’ probability in the opposite situation. The Net Benefit is the difference between these probabilities, and the Win Ratio is the ratio between them. In this study, we review several estimators of the win/loss probability in terms of generalized pairwise comparisons, and the variance of those estimators based on U -statistic theory. Especially, we explain how to deal with the censored data to estimate those probabilities. Finally, we illustrate using the actual data to implement those statistical methods by R-packages.

キーワード：一般化ペアワイズ比較, Net Benefit, Win Ratio, 生存時間解析, 打ち切り, U 統計量

1. はじめに

がん領域や循環器領域における治療の効果を評価する医学系研究（臨床試験）で、群間での評価項目（結果変数）の大小関係に注目する Net Benefit (Buyse (2010), Péron *et al.*

* アステラス製薬株式会社：〒 103-0023 東京都中央区日本橋本町 2-5-1 (E-mail: musashi.fukuda@astellas.com).

† 横浜市立大学大学院データサイエンス研究科：〒 236-0027 神奈川県横浜市金沢区瀬戸 22-2 (E-mail: sakamaki@yokohama-cu.ac.jp).

‡ 東京大学大学院情報学環：〒 113-0033 東京都文京区本郷 7-3-1 (E-mail: oba@epistat.m.u-tokyo.jp).

(2018)) や Win Ratio (Pocock *et al.* (2012)) などの効果指標の利用が増えている。これらはいずれも一般化ペアワイズ比較 (Generalized Pairwise Comparisons) という統計手法と関連する指標である (Buyse (2010), Verbeeck *et al.* (2019), Buyse and Péron (2020))。一般化ペアワイズ比較は、評価項目の大小関係に関する順序構造に基づいた勝ち負けのルールを用いて、あるペアにおける評価項目を比較する方法の総称である。例えば、新治療と標準治療を生存時間により比較するランダム化比較試験の場合、各群から 1 人ずつをランダムに選んでできるペアにおいて、新治療群の生存時間が標準治療群の生存時間よりも長い場合に「勝ち」、逆に新治療群の方が短い場合に「負け」とするようなルールである。このような手順を用いる統計手法は Mann-Whitney 検定 (Mann and Whitney (1947)) やその効果量の推定 (Efron (1967), Dobler and Pauly (2018)) と関係する。Buyse は、Mann-Whitney 検定における比較の手順を評価項目の型や数によらず適用できるように拡張し、これを一般化ペアワイズ比較として提案した (Buyse (2010))。一般化ペアワイズ比較に関する指標として、新治療群が標準治療群に対して勝っている確率 (Mann-Whitney 統計量 (Efron (1967), Dobler and Pauly (2018)) あるいは Probabilistic Index (Acion *et al.* (2006)) と呼ばれる) および新治療群が負けている確率、さらにこれらの確率の差として Net Benefit, 比として Win Ratio などがある (Verbeeck *et al.* (2019), Buyse and Péron (2020))。本稿では、このような一般化ペアワイズ比較に基づく指標やそれに関する統計量のことも含めて「一般化ペアワイズ比較」という表現を用いることもある。以降では、一般化ペアワイズ比較に関する方法の中でも、特に Net Benefit と Win Ratio に焦点を当てる。

Net Benefit は、proportion in favor of treatment として Buyse により提案され (Buyse (2010)), その後、生存時間や有害事象などを評価項目に用いるがん領域の臨床試験への応用を中心に発展してきた (Buyse (2010), Péron *et al.* (2015, 2016a, b, 2019a, b), Saad *et al.* (2018), Buyse *et al.* (2020))。Net Benefit が生存時間に対する効果指標として注目されている背景の一つに、比例ハザード性に基づく生存時間解析が不適切である臨床試験が散見されることがある。例えば、がん免疫療法の臨床試験では、群ごとの生存関数が初期では重なり、後期に差が開くような状況が散見され、比例ハザード性が疑われることがある (Saad *et al.* (2018), Buyse *et al.* (2020))。Net Benefit の推定には比例ハザード性の仮定が不要であり、結果を直観的に解釈できると指摘されている (Buyse (2010), Péron *et al.* (2015, 2016a, b, 2019a, b), Saad *et al.* (2018), Buyse *et al.* (2020))。Net Benefit における勝ち負けのルールに関しては、生存時間の群間での差がある閾値 τ 以上長いという条件を考慮する (Péron *et al.* (2016b, 2019b)), 生存時間と有害事象などの複数の評価項目の優先順位を考慮して勝ち負けを定義する (Péron *et al.* (2015, 2016a, 2019a)), などの治療のベネフィット・リスクを評価することも可能である。近年、臨床試験におけるベネフィット・リスク評価に関するニーズも高まっており (European Medicines Agency (2011), Food

and Drug Administration (2021)), 後ろ向きの解析ではあるが, 実際に行われた膵臓がんの臨床試験において, Net Benefit による評価を行った事例も報告されている (Buyse and Péron (2020), Péron *et al.* (2015, 2016a, 2019a)). Net Benefit の推定に関しては, 打ち切りへの対処 (Péron *et al.* (2018, 2021), Chamseddine *et al.* (2021)), 推定量の漸近分散の導出 (Ozenne *et al.* (2021)), 相関のある複数の評価項目から勝ち負けを決める場合の推定量の性能評価 (Buyse *et al.* (2021), Giai *et al.* (2021)) などが報告されている.

Win Ratio は, 循環器領域において生存時間を用いて治療効果を評価する際の複合評価項目に対する解析方法として Pocock らにより提案された (Pocock *et al.* (2012)). 例えば, 心不全の臨床試験では, 「理由を問わない全死亡 (any-cause death)」「心不全による入院 (hospitalization)」という二つのイベントに対する生存時間が考えられるが, いずれかのイベントが最初に発現するまでの時間 (time to first event) という複合評価項目 (composite endpoint) を用い, 一つの生存時間に対する生存時間解析がしばしば行われてきた. しかし, 複合評価項目はイベントの医学的な重要度 (優先順位) を考慮した定義ではないため, 解釈が困難になる場合がある (Pocock *et al.* (2012), Ferreira *et al.* (2020), Redfors *et al.* (2020)). 例えば, ある対象者が入院と死亡のどちらともを起こした場合, 時間的に先行する入院イベントまでの時間により複合評価項目が定義されることとなる. この場合, 医学的な重要度がより高い死亡の情報は治療効果の評価において無視される. Win Ratio は, 複数の評価項目を階層的に比較することで, より重要度が高いと考えられる評価項目を優先的に用いた勝ち負けの評価が可能となる (Pocock *et al.* (2012), Ferreira *et al.* (2020), Redfors *et al.* (2020)). 先の例では, 優先順位を 1) 全死亡, 2) 心不全による入院, という階層 (順番) で勝ち負けを評価するルールを用いると, 両方のイベントを起こした場合でも, 全死亡までの時間の情報を優先して勝ち負けに寄与させることができる (Pocock *et al.* (2012), Ferreira *et al.* (2020), Redfors *et al.* (2020)). Win Ratio の推定に関しては, 推定量に対する漸近分散の提案 (Luo *et al.* (2015), Bebu and Lachin (2016), Dong *et al.* (2016)), 打ち切りにより勝ち負けが確定できない場合の扱い (Dong *et al.* (2020a)), 背景因子を区別した推定 (層別解析) (Dong *et al.* (2018)), 打ち切りがある場合の対処法 (Oakes (2016), Finkelstein and Schoenfeld (2019), Dong *et al.* (2020b, c, 2021)) などの研究が報告されている. また, 循環器領域の臨床試験を中心として, Win Ratio による解析を適用した事例も報告されている (Pocock *et al.* (2012), Fergusson *et al.* (2018), Kotalik *et al.* (2019), Ferreira *et al.* (2020), Redfors *et al.* (2020), Voors *et al.* (2022)). 特に, 心不全領域における EMPULSE 試験では, Win Ratio が事前に規定された主要評価項目に対する解析として用いられた (Voors *et al.* (2022)).

Net Benefit および Win Ratio は異なる背景から提案されたものではあるが, 1) 評価項目の大小関係に基づく比較, 2) 複数の評価項目の階層的な評価, 3) 評価項目の型によらな

い適用可能性という共通点をもつ (Buyse (2010), Buyse and Péron (2020)). これらの指標は連続変数や二値変数などに対しても定義可能であるが, 提案された背景から, 興味のある評価項目として生存時間を含む臨床試験での応用が多い. また, これまで Net Benefit や Win Ratio に関する方法論や応用に関する研究は, それぞれ異なる文脈で発展してきたこともあり, 統一的に整理されることはほとんどなかった. しかし, これらは統計的にも共通点を持つ指標であるため, 生存時間に対する解析方法として一般化ペアワイズ比較の観点から包括的に整理することは有用である. また, Net Benefit や Win Ratio を実装するための R パッケージも充実してきており (Ozenne and Péron (2022), Cui and Huang (2022)), これらの方法の理解は応用的観点からも重要といえる.

本稿では, 生存時間に対する一般化ペアワイズ比較を説明し, Net Benefit および Win Ratio の推測に関する既存方法をレビューに基づき説明する. さらに, いくつかの R パッケージを用いた実装方法を実データ解析に基づき例示する. 2 節では打ち切りの有無に応じた一般化ペアワイズ比較に基づく Net Benefit と Win Ratio の推定量を主に解説し, 3 節ではそれらの漸近分散の推定方法を紹介する. 4 節では, 本稿で解説する方法を実装する R パッケージとして, *BuyseTest* と *WINS* を用いた実データの解析例を示す. 最後に, 考察を 5 節でまとめる. 以下の議論では, 2 群のランダム化比較試験を想定し, 生存時間変数に対しては右側打ち切りのみを仮定する.

2. Net Benefit および Win Ratio の推定

本稿では, 簡単のため, 優先順位の異なる二つの生存時間 (例えば, 全死亡と心不全による入院) を用いて治療を比較する場合を想定する. ただし, その他の型の評価項目や三つ以上の生存時間を用いる場合も同様の推測ができるため, 説明の一般性は失われない.

新治療群の対象者 i ($i = 1, \dots, m$) と標準治療群の対象者 j ($j = 1, \dots, n$) に対し, 優先順位が高いイベント (全死亡) が起きるまでの潜在的な生存時間を $X_i^{(1)}, Y_j^{(1)}$, 優先順位が低いイベント (心不全による入院) が起きるまでの潜在的な生存時間を $X_i^{(2)}, Y_j^{(2)}$ とする. 各群の対象者の潜在的な打ち切りまでの時間を U_i, V_j とする. ただし, 個人を特定する必要がない場合, 添字 i, j を省略することもある. $X^{(l)}, Y^{(l)}$ ($l = 1, 2$) に対する生存関数を $S_{X^{(l)}}(t), S_{Y^{(l)}}(t), U, V$ の生存関数を $G_U(t), G_V(t)$ とする. 打ち切りを伴う生存時間 (観察生存時間) を, $\tilde{X}^{(l)} = \min(X^{(l)}, U)$ および $\tilde{Y}^{(l)} = \min(Y^{(l)}, V)$ としたとき, $\tilde{X}^{(l)}, \tilde{Y}^{(l)}$ の実現値 $\tilde{x}_i^{(l)}, \tilde{y}_j^{(l)}$ に対し, $\delta_i^{(l)} = I(\tilde{X}^{(l)} = X^{(l)})$, $\varepsilon_j^{(l)} = I(\tilde{Y}^{(l)} = Y^{(l)})$ とすると, 実際に観察される生存時間に関するデータの組は $(\tilde{x}_i^{(1)}, \delta_i^{(1)}, \tilde{x}_i^{(2)}, \delta_i^{(2)})$ と $(\tilde{y}_j^{(1)}, \varepsilon_j^{(1)}, \tilde{y}_j^{(2)}, \varepsilon_j^{(2)})$ となる. 本稿では, 打ち切りはイベントまでの時間と独立であると仮定し, 各生存時間を競合リスクとしては考慮しないこととする. また, $\tau^{(l)} (\geq 0)$ は, $X^{(l)}, Y^{(l)}$ に対する勝ち負けのルールを決める際に事前に定める閾値である. 例えば, 新治療群の生存時間が標準治療群より

3 ヶ月以上長い場合に勝ちとするルールであれば、 $\tau^{(1)} = 3$ ヶ月とする。なお、評価項目ごとに異なる閾値を設定することも可能であり、 $\tau^{(1)} \neq \tau^{(2)}$ であってもよい。

以下では、2.1 節で一般化ペアワイズ比較の定義を述べ、2.2 節で打ち切りがない場合、2.3 節で打ち切りがある場合の Net Benefit と Win Ratio の推定法を説明する。2.4 節では生存関数による推定方法を概説する。

2.1 一般化ペアワイズ比較

一般化ペアワイズ比較では、事前に定めた勝ち負けのルールにより、一つまたは複数の評価項目を勝敗がつくまで順番にそれぞれ比較し、その結果から総合的に各ペアにおける勝ち負けを判定する。ここで、 l 番目の評価項目における新治療群の勝ち、引き分け、負けの条件のそれぞれを、 $X^{(l)} > Y^{(l)} + \tau^{(l)}$, $|X^{(l)} - Y^{(l)}| \leq \tau^{(l)}$, $X^{(l)} < Y^{(l)} - \tau^{(l)}$ と表す。このとき、 $\mathbf{X} = (X^{(1)}, X^{(2)})$, $\mathbf{Y} = (Y^{(1)}, Y^{(2)})$ に対する半順序構造を、

$$\mathbf{X} \succ \mathbf{Y} \equiv (X^{(1)} > Y^{(1)} + \tau^{(1)}) \cup \{(|X^{(1)} - Y^{(1)}| \leq \tau^{(1)}) \cap (X^{(2)} > Y^{(2)} + \tau^{(2)})\}$$

と定義し、これにより標準治療群に対する新治療群の「勝ち」を表す (Bebu and Lachin (2016), Péron *et al.* (2021))。同様に、 \mathbf{X}, \mathbf{Y} に対する逆の半順序構造、すなわち標準治療群に対する新治療群の「負け」(新治療群に対する標準治療群の「勝ち」)を、

$$\mathbf{X} \prec \mathbf{Y} \equiv (X^{(1)} < Y^{(1)} - \tau^{(1)}) \cup \{(|X^{(1)} - Y^{(1)}| \leq \tau^{(1)}) \cap (X^{(2)} < Y^{(2)} - \tau^{(2)})\}$$

と表す。評価項目が二つの生存時間である場合以外でも同様の勝ち負けのルールを規定することで一般化ペアワイズ比較の適用は可能であるが、二値の評価項目では $\tau^{(l)}$ を設定できない (Péron *et al.* (2015, 2016a, 2019a)) など、評価項目の型や研究の目的に応じた詳細なルールの検討は必要である。評価項目の型や数によらない一般の場合における一般化ペアワイズ比較については、Bebu and Lachin や Buyse らにより議論されている (Buyse (2010), Bebu and Lachin (2016), Buyse and Péron (2020), Buyse *et al.* (2021))。

一般化ペアワイズ比較に基づき、事前に定めた勝ち ($\mathbf{X} \succ \mathbf{Y}$) および負け ($\mathbf{X} \prec \mathbf{Y}$) のルールに対して、以下の新治療群が勝っている確率 $\Pr[\mathbf{X} \succ \mathbf{Y}]$ と負けている確率 $\Pr[\mathbf{X} \prec \mathbf{Y}]$ を推定することができる。

$$\Pr[\mathbf{X} \succ \mathbf{Y}] = \Pr[X^{(1)} > Y^{(1)} + \tau^{(1)}] + \Pr[(|X^{(1)} - Y^{(1)}| \leq \tau^{(1)}) \cap (X^{(2)} > Y^{(2)} + \tau^{(2)})]$$

$$\Pr[\mathbf{X} \prec \mathbf{Y}] = \Pr[X^{(1)} < Y^{(1)} - \tau^{(1)}] + \Pr[(|X^{(1)} - Y^{(1)}| \leq \tau^{(1)}) \cap (X^{(2)} < Y^{(2)} - \tau^{(2)})]$$

Net Benefit (Δ) と Win Ratio (Ψ) はこれらの差および比として以下のように定義される (Bebu and Lachin (2016), Péron *et al.* (2021))。

$$\Delta = \Pr[\mathbf{X} \succ \mathbf{Y}] - \Pr[\mathbf{X} \prec \mathbf{Y}]$$

$$\Psi = \Pr[\mathbf{X} \succ \mathbf{Y}] / \Pr[\mathbf{X} \prec \mathbf{Y}]$$

$\Pr[\mathbf{X} \succ \mathbf{Y}]$ や $\Pr[\mathbf{X} \prec \mathbf{Y}]$ の推定方法には、ペアごとのスコアに基づく方法と生存関数による方法がある (Buyse (2010), Bebu and Lachin (2016), Oakes (2016), Dong *et al.* (2016, 2020b), Péron *et al.* (2018), Finkelstein and Schoenfeld (2019)). これまでの推定方法の提案の多くはスコアに基づくものであるため、以下では主にスコアに基づく方法を説明し、生存関数による推定法は 2.4 節で簡単に説明する。

2.2 打ち切りがない場合の推定

新治療群と標準治療群からそれぞれ 1 人ずつ対象者を選んでペアを作り、ペアごとに評価項目の大小関係を順番に比較し、当該ペアにおける勝ち負けのスコアを計算する。評価項目の優先順位を考慮するスコアの計算では、1 番目の評価項目の比較によるスコアを計算し、その比較が引き分けであった場合に 2 番目の評価項目の比較によるスコアを計算し、最後にそれぞれの評価項目に対するスコアから総合的に計算されるスコアを求める。計算されたスコアを用いて $\Pr[\mathbf{X} \succ \mathbf{Y}]$ と $\Pr[\mathbf{X} \prec \mathbf{Y}]$ を推定し、これらの差および比として Net Benefit と Win Ratio は推定される。

ここで、 $X_i^{(1)}$ と $Y_j^{(1)}$ の比較における勝ち、負け、引き分けのスコアを $s_{ij}^{(1)+}, s_{ij}^{(1)-}, s_{ij}^{(1)0}$ とし、 $X_i^{(2)}$ と $Y_j^{(2)}$ の比較における勝ち、負け、引き分けのスコアを $s_{ij}^{(2)+}, s_{ij}^{(2)-}, s_{ij}^{(2)0}$ とする。さらに、二つの評価項目の優先順位を考慮した総合的な勝ちと負けのスコアを s_{ij}^+, s_{ij}^- とする。打ち切りがない場合、すべてのペアにおいて各評価項目の大小関係を観察データから確定でき、各評価項目 ($l = 1, 2$) に対する勝ち・負け・引き分けのスコア

$$\begin{aligned} s_{ij}^{(l)+} &= I(x_i^{(l)} > y_j^{(l)} + \tau^{(l)}) \\ s_{ij}^{(l)-} &= I(x_i^{(l)} < y_j^{(l)} - \tau^{(l)}) \\ s_{ij}^{(l)0} &= I(|x_i^{(l)} - y_j^{(l)}| \leq \tau^{(l)}) \end{aligned}$$

が求まる。

優先順位を考慮したペアごとの総合スコアは表 1 のようになる。

表 1 で示したペア (i, j) に対する二つの評価項目による総合スコアは、

$$\begin{aligned} s_{ij}^+ &= I(x_i^{(1)} > y_j^{(1)} + \tau^{(1)}) + I(|x_i^{(1)} - y_j^{(1)}| \leq \tau^{(1)})I(x_i^{(2)} > y_j^{(2)} + \tau^{(2)}) \\ s_{ij}^- &= I(x_i^{(1)} < y_j^{(1)} - \tau^{(1)}) + I(|x_i^{(1)} - y_j^{(1)}| \leq \tau^{(1)})I(x_i^{(2)} < y_j^{(2)} - \tau^{(2)}) \end{aligned}$$

と表現できる。したがって、総合スコア s_{ij}^+, s_{ij}^- は、 $w_{ij}^{(2)} = s_{ij}^{(1)0}$ とすると、以下のように各評価項目に対するスコアの重み付き和として表現できる。

$$s_{ij}^+ = s_{ij}^{(1)+} + w_{ij}^{(2)} \times s_{ij}^{(2)+}, \quad s_{ij}^- = s_{ij}^{(1)-} + w_{ij}^{(2)} \times s_{ij}^{(2)-}$$

ここで、 $w_{ij}^{(2)}$ は、 $X_i^{(2)}$ と $Y_j^{(2)}$ の比較が総合スコアに寄与する重みである。

表 1 打ち切りがない場合の一般化ペアワイズ比較における総合スコア.

上位の評価項目	下位の評価項目	総合的スコア
新治療の勝ち $s_{ij}^{(1)+} = 1, s_{ij}^{(1)-} = 0, s_{ij}^{(1)0} = 0$	考慮しない	新治療の勝ち $s_{ij}^{+} = 1, s_{ij}^{-} = 0$
新治療の負け $s_{ij}^{(1)+} = 0, s_{ij}^{(1)-} = 1, s_{ij}^{(1)0} = 0$	考慮しない	新治療の負け $s_{ij}^{+} = 0, s_{ij}^{-} = 1$
引き分け $s_{ij}^{(1)+} = 0, s_{ij}^{(1)-} = 0, s_{ij}^{(1)0} = 1$	新治療の勝ち $s_{ij}^{(2)+} = 1, s_{ij}^{(2)-} = 0, s_{ij}^{(2)0} = 0$	新治療の勝ち $s_{ij}^{+} = 1, s_{ij}^{-} = 0$
引き分け $s_{ij}^{(1)+} = 0, s_{ij}^{(1)-} = 0, s_{ij}^{(1)0} = 1$	新治療の負け $s_{ij}^{(2)+} = 0, s_{ij}^{(2)-} = 1, s_{ij}^{(2)0} = 0$	新治療の負け $s_{ij}^{+} = 0, s_{ij}^{-} = 1$
引き分け $s_{ij}^{(1)+} = 0, s_{ij}^{(1)-} = 0, s_{ij}^{(1)0} = 1$	引き分け $s_{ij}^{(2)+} = 0, s_{ij}^{(2)-} = 0, s_{ij}^{(2)0} = 1$	引き分け $s_{ij}^{+} = 0, s_{ij}^{-} = 0$

以上の手順により計算されたすべてのペアでの s_{ij}^{+} と s_{ij}^{-} を用いて、勝ち負けの確率を以下のように推定する.

$$\widehat{\Pr}[\mathbf{X} \succ \mathbf{Y}] = \hat{U}^{+} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{+} \quad (1)$$

$$\widehat{\Pr}[\mathbf{X} \prec \mathbf{Y}] = \hat{U}^{-} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{-} \quad (2)$$

Net Benefit と Win Ratio の推定量は、 \hat{U}^{+} と \hat{U}^{-} から

$$\hat{\Delta} = \hat{U}^{+} - \hat{U}^{-}, \quad \hat{\Psi} = \hat{U}^{+} / \hat{U}^{-}$$

となる. スコアに基づくこれらの推定量は、多変量の U 統計量理論により、一致性、漸近正規性がある (Bebu and Lachin (2016), Dong *et al.* (2016), Péron *et al.* (2021), Ozenne *et al.* (2021)).

一般に、 L 個の評価項目の場合は、勝ち負けの総合スコア s_{ij}^{+} および s_{ij}^{-} を L 個の評価項目に対するスコアの重み付き和として計算すれば良く、上記と同様の計算で求まる. 詳細については付録 A に示す.

2.3 打ち切りがある場合の推定

生存時間を評価項目に用いる場合、打ち切りにより、勝ち、負け、引き分けが確定しないペアが存在する場合がある. Buyse や Péron らはこのようなペアを無情報ペア (uninformative pair) としている (Buyse (2010), Péron *et al.* (2018, 2021)). 打ち切りによる無情報ペアが存在する場合、無情報ペアを引き分けと同様に扱う (2.3.1 節)、勝ち負けのスコアを重み付ける (2.3.2 節)、条件付き確率に基づく補完を伴うスコアを用いる (2.3.3 節)、といっ

表 2 一般化ベアワイズ比較における Gehan, Dong, Péron 流のスコアリングのまとめ.

A. $s_{ij}^{(l)+}, s_{ij}^{(l)-}$ に対するスコアリング

		$s_{ij}^{(l)+,v}$		$s_{ij}^{(l)-,v}$	
		$(\delta_i, \varepsilon_i)$		$\tilde{x}_i > \tilde{y}_j + \tau$	$\tilde{x}_i \leq \tilde{y}_j + \tau$
Gehan ($v = G$)	(1,1)	1	0	1	0
	(0,1)	1	0	0	0
	(1,0)	0	0	1	0
	(0,0)	0	0	0	0
Dong ($v = D$)	(1,1)	$1/(\hat{G}_0(\tilde{y}_j + \tau)\hat{G}_v(\tilde{y}_j))$	0	$1/(\hat{G}_0(\tilde{x}_i)\hat{G}_v(\tilde{x}_i + \tau))$	0
	(0,1)	$1/(\hat{G}_0(\tilde{y}_j + \tau)\hat{G}_v(\tilde{y}_j))$	0	0	0
	(1,0)	0	0	$1/(\hat{G}_0(\tilde{x}_i)\hat{G}_v(\tilde{x}_i + \tau))$	0
	(0,0)	0	0	0	0
Péron ($v = P$)	(1,1)	1	0	1	0
	(0,1)	1	$\frac{\hat{S}_x(\tilde{y}_j + \tau)}{\hat{S}_x(\tilde{x}_i)}$	$1 - \frac{\hat{S}_x(\tilde{y}_j - \tau)}{\hat{S}_x(\tilde{x}_i)}$	0
	(1,0)	$1 - \frac{\hat{S}_y(\tilde{x}_i - \tau)}{\hat{S}_y(\tilde{y}_j)}$	0	1	$\frac{\hat{S}_y(\tilde{x}_i + \tau)}{\hat{S}_y(\tilde{y}_j)}$
	(0,0)	$1 - \frac{\hat{S}_y(\tilde{x}_i - \tau)}{\hat{S}_y(\tilde{y}_j)}$ $-\frac{\int_{t>\tilde{x}_i-\tau}^{\infty} \hat{S}_x(t+\tau) d\hat{S}_y(t)}{\hat{S}_x(\tilde{x}_i)\hat{S}_y(\tilde{y}_j)}$	$-\frac{\int_{t>\tilde{y}_i}^{\infty} \hat{S}_y(t+\tau) d\hat{S}_x(t)}{\hat{S}_x(\tilde{x}_i)\hat{S}_y(\tilde{y}_j)}$	$1 - \frac{\hat{S}_x(\tilde{y}_j - \tau)}{\hat{S}_x(\tilde{x}_i)}$ $-\frac{\int_{t>\tilde{y}_j-\tau}^{\infty} \hat{S}_y(t+\tau) d\hat{S}_x(t)}{\hat{S}_x(\tilde{x}_i)\hat{S}_y(\tilde{y}_j)}$	$-\frac{\int_{t>\tilde{x}_i}^{\infty} \hat{S}_y(t+\tau) d\hat{S}_x(t)}{\hat{S}_x(\tilde{x}_i)\hat{S}_y(\tilde{y}_j)}$

表 2 (continued).

B. $s_{ij}^{(0)0}$ に対するスコアリング		$s_{ij}^{(l)0,v}$	
$(\delta_i, \varepsilon_j)$	$ \tilde{x}_i - \tilde{y}_j \leq \tau$	$\tilde{x}_i > \tilde{y}_j + \tau$	$\tilde{x}_i < \tilde{y}_j - \tau$
Gehan [†] ($v = G$)	(0, 1)	1	0
	(1, 0)	0	0
	(0, 0)	0	0
	(0, 1)	0	0
Péron ($v = P$)	(1, 1)	1	0
	(0, 1)	$1 - \frac{\hat{S}_X(\tilde{y}_j - \tau)}{\hat{S}_X(\tilde{x}_i)}$	$\frac{\hat{S}_X(\tilde{y}_j - \tau) - \hat{S}_X(\tilde{y}_j + \tau)}{\hat{S}_X(\tilde{x}_i)}$
	(1, 0)	$1 - \frac{\hat{S}_Y(\tilde{x}_i - \tau)}{\hat{S}_Y(\tilde{y}_j)}$	$\frac{\hat{S}_Y(\tilde{x}_i - \tau) - \hat{S}_Y(\tilde{x}_i + \tau)}{\hat{S}_Y(\tilde{y}_j)}$
(0, 0)		$1 + \frac{\int_{t > \tilde{y}_j}^{\infty} \hat{S}_X(t + \tau) d\hat{S}_Y(t)}{\hat{S}_X(\tilde{x}_i) \hat{S}_Y(\tilde{y}_j)}$	$\frac{\hat{S}_Y(\tilde{x}_i - \tau)}{\hat{S}_Y(\tilde{y}_j)} + \frac{\int_{t > \tilde{y}_j}^{\infty} \hat{S}_X(t + \tau) d\hat{S}_Y(t)}{\hat{S}_X(\tilde{x}_i)} + \frac{\int_{t > \tilde{y}_j}^{\infty} \hat{S}_X(t + \tau) d\hat{S}_Y(t)}{\hat{S}_X(\tilde{x}_i) \hat{S}_Y(\tilde{y}_j)}$
		$1 + \frac{\int_{t > \tilde{x}_i}^{\infty} \hat{S}_Y(t + \tau) d\hat{S}_X(t)}{\hat{S}_Y(\tilde{x}_i) \hat{S}_Y(\tilde{y}_j)}$	$\frac{\int_{t > \tilde{x}_i}^{\infty} \hat{S}_Y(t + \tau) d\hat{S}_X(t)}{\hat{S}_Y(\tilde{x}_i) \hat{S}_Y(\tilde{y}_j)} + \frac{\int_{t > \tilde{x}_i}^{\infty} \hat{S}_Y(t + \tau) d\hat{S}_X(t)}{\hat{S}_X(\tilde{x}_i) \hat{S}_Y(\tilde{y}_j)}$

*記法の簡略化のため, 表内における評価項目の番号を表す添字 l と Péron 流スコアにおける $\hat{\eta}_X^{(l)}, \hat{\eta}_Y^{(l)}$ を省略している. 各方法の解説は, 本文 2.3.1 ~ 2.3.3 節を参照のこと. $\delta_i = 1$ は観察生存時間 \tilde{x}_i がイベントであることを, $\delta_i = 0$ は \tilde{x}_i が打ち切りであることを示す. 同様に $\varepsilon_j = 1$ は観察生存時間 \tilde{y}_j がイベントであることを, $\varepsilon_j = 0$ は観察生存時間 \tilde{y}_j が打ち切りであることを示す.

[†] $s_{ij}^{(0)0}$ に対する Gehan 流のスコアは, 本文 2.3.2 節の Dong 流スコアの計算にも同様に用いられる.

た提案がなされている (Péron *et al.* (2018, 2021), Dong *et al.* (2020c, 2021)). 手法間の各評価項目に対するスコアの違いを表 2 に示す.

2.3.1 Gehan 流スコア

Buyse は打ち切りがある場合の Net Benefit の推定方法として, 大小関係が確定するペアのみを利用するスコアリングを提案した (Buyse (2010)). このスコアリングは, 評価項目がただ一つで $\tau = 0$ である場合に Gehan による一般化 Wilcoxon 検定のスコアと等しくなることから, いくつかの研究で Gehan 流スコアとよばれている (Péron *et al.* (2021), Ozenne *et al.* (2021)).

Gehan 流スコアでは, 各評価項目 $l = 1, 2$ に対して, 勝ち, 負け, 引き分けのスコアを以下のように計算する.

$$\begin{aligned} s_{ij}^{(l)+,G} &= I(\tilde{x}_i^{(l)} > \tilde{y}_j^{(l)} + \tau^{(l)})\varepsilon_j^{(l)} \\ s_{ij}^{(l)-,G} &= I(\tilde{x}_i^{(l)} < \tilde{y}_j^{(l)} - \tau^{(l)})\delta_i^{(l)} \\ s_{ij}^{(l)0,G} &= I(|\tilde{x}_i^{(l)} - \tilde{y}_j^{(l)}| \leq \tau^{(l)})\delta_i^{(l)}\varepsilon_j^{(l)} \end{aligned}$$

例えば, $\delta_i^{(l)} = 0, \varepsilon_j^{(l)} = 1$ であった場合, $x_i^{(l)} > \tilde{x}_i^{(l)}, y_j^{(l)} = \tilde{y}_j^{(l)}$ であることから, $\tilde{x}_i^{(l)} > \tilde{y}_j^{(l)} + \tau^{(l)}$ ならば $x_i^{(l)} > y_j^{(l)} + \tau^{(l)}$ となり, 当該ペアにおいて新治療群の勝ちが確定する ($s_{ij}^{(l)+,G} = 1$). 打ち切りによって勝ち, 負け, 引き分けのいずれも確定できない無情報ペアの指示変数は以下ようになる.

$$u_{ij}^{(l),G} = 1 - (s_{ij}^{(l)+,G} + s_{ij}^{(l)-,G} + s_{ij}^{(l)0,G})$$

各評価項目に対するスコアに基づき, 二つの評価項目から勝ち負けを決める Gehan 流の総合スコアは,

$$s_{ij}^{+,G} = s_{ij}^{(1)+,G} + w_{ij}^{(2),G} \times s_{ij}^{(2)+,G}, \quad s_{ij}^{-,G} = s_{ij}^{(1)-,G} + w_{ij}^{(2),G} \times s_{ij}^{(2)-,G} \quad (3)$$

となる. ただし, 2 番目の評価項目に対する重みは $w_{ij}^{(2),G} = s_{ij}^{(1)0,G} + u_{ij}^{(1),G}$ である. (3) 式では, 上位の評価項目での引き分けに加えて, 無情報の場合も下位の評価項目の比較に進むという手順になっている (Buyse (2010)). 一方で, 最下位 (2 番目) の評価項目での無情報ペアは (3) 式のスコアには寄与しない.

Gehan 流スコアに基づく Net Benefit および Win Ratio の推定は, (3) 式で計算されるスコアを用いて (1)(2) 式の U 統計量を算出し, それらの差や比から求めればよい. Gehan 流スコアに基づく Net Benefit および Win Ratio はその漸近分布が打ち切り分布に依存することが示されており (Bebu and Lachin (2016), Oakes (2016)), 推定量にバイアスがあることが経験的にも示されている (Péron *et al.* (2018, 2021), Dong *et al.* (2020c), Ozenne *et al.* (2021)). Gehan 流スコアに基づく Win Ratio の推定量は, Pocock らが unmatched

win ratio として提案したもの (Pocock *et al.* (2012)), あるいは Bebu and Lachin や Dong らによる一連の研究で議論されている生存時間データに対する Win Ratio の推定量 (Bebu and Lachin (2016), Dong *et al.* (2016, 2018, 2020b)) と実質的に同値である。ただし、これらの Win Ratio に関する研究では、無情報ペアと引き分けペアの区別は明確にはしていない。その理由として、全死亡の観察には長期追跡が必要となるために第一優先の評価項目での勝ち負けの評価がそもそも困難な状況であること、勝ち負けのルールにおいて明示的には τ を考慮しない ($\tau = 0$) ことなど、Win Ratio を応用する場面である循環器疾患の特徴を加味した方法論の発展がなされてきたことによると考えられる (Pocock *et al.* (2012))。また、 $\tau = 0$ とした (3) 式の Gehan 流スコアに基づき、 $\sum_{i=1}^m \sum_{j=1}^n (s_{ij}^+ - s_{ij}^-)$ と計算した統計量は、Finkelstein-Schoenfeld 検定 (Finkelstein and Schoenfeld (1999)) のスコアと同等であることが知られている (Verbeeck *et al.* (2019))。Finkelstein-Schoenfeld 検定は、AIDS の臨床試験において、評価項目の優先度を考慮して生存時間データと経時データを統合した検定として提案されたものである (Finkelstein and Schoenfeld (1999))。循環器領域の ATTR-ACT 試験 (Maurer *et al.* (2018)) などでは、治療効果の推定に Win Ratio を用い、その検定として Finkelstein-Schoenfeld 検定を用いている。また、治療中止を伴う高カリウム血症に対する治療評価における Win Ratio や Finkelstein-Schoenfeld 検定の有用性の議論がされている (Tsuchikawa and Sakamaki (2022))。

2.3.2 Dong 流スコア

Dong らは打ち切りがある場合の Win Ratio の推定方法として、逆確率重み付け (Inverse Probability of Censoring Weight: IPCW) 法を応用した以下のスコアの利用を提案している (Dong *et al.* (2020c))。

$$s_{ij}^{(l)+,D} = \frac{I(\tilde{x}_i^{(l)} > \tilde{y}_j^{(l)} + \tau^{(l)})\varepsilon_j^{(l)}}{\hat{G}_U(\tilde{y}_j^{(l)})\hat{G}_V(\tilde{y}_j^{(l)})}$$

$$s_{ij}^{(l)-,D} = \frac{I(\tilde{x}_i^{(l)} < \tilde{y}_j^{(l)} - \tau^{(l)})\delta_i^{(l)}}{\hat{G}_U(\tilde{x}_i^{(l)})\hat{G}_V(\tilde{x}_i^{(l)})}$$

\hat{G}_U, \hat{G}_V は各群における打ち切り時間の生存関数の推定量であり、Dong らは Kaplan-Meier 法により推定することを提案している (Dong *et al.* (2020c))。このスコアは Gehan 流スコアに観察確率の逆数を重み付けた形になっている。Dong らは $\tau = 0$ とした Win Ratio の場合のみを議論しているが (Dong *et al.* (2020c)), $\tau > 0$ の場合、 $s_{ij}^{(l)+,D}$ の分母を $\hat{G}_U(\tilde{y}_j^{(l)} + \tau^{(l)})\hat{G}_V(\tilde{y}_j^{(l)})$ と、 $s_{ij}^{(l)-,D}$ の分母を $\hat{G}_U(\tilde{x}_i^{(l)})\hat{G}_V(\tilde{x}_i^{(l)} + \tau^{(l)})$ と修正する必要がある。

二つの評価項目から勝ち負けを総合的に評価した Dong 流スコアを、

$$s_{ij}^{+,D} = s_{ij}^{(1)+,D} + w_{ij}^{(2),G} \times s_{ij}^{(2)+,D}, \quad s_{ij}^{-,D} = s_{ij}^{(1)-,D} + w_{ij}^{(2),G} \times s_{ij}^{(2)-,D} \quad (4)$$

とする。ただし、2 番目の評価項目に対する重みは、Gehan 流スコアと同じく、引き分け

と無情報ペアに基づく (3) 式の $w_{ij}^{(2),G}$ を用いる. 最下位 (2 番目) の評価項目での無情報ペアの情報は勝ちおよび負けのペアに重み付けることで考慮されるため, 最下位の評価項目での無情報ペアは (4) 式のスコアには直接的には寄与しない. (4) 式のスコアを (1)(2) 式の統計量の計算に用いることで Net Benefit および Win Ratio の推定量が計算できる.

Dong 流スコアを用いる推定量は打ち切り時間と生存時間が独立であるという独立打ち切りの仮定が成立する状況においては, 一致性のある推定方法であることが示されている (Dong *et al.* (2020c)). また, Dong らは, $s_{ij}^{(l)+,D}, s_{ij}^{(l)-,D}$ における重みの推定において, 打ち切り時間と関連する共変量を含めた Cox 回帰を用いることで, 打ち切りに対する仮定を共変量条件付き独立に緩めた推定量も提案している (Dong *et al.* (2021)).

2.3.3 Péron 流スコア

Péron らは, 打ち切りにより大小関係が確定しないペア (Gehan 流スコアにおける無情報ペア) に対して, 以下のような勝ち負けの確率の推定値を用いて当該スコアを補完する方法を提案した (Péron *et al.* (2018), Ozenne *et al.* (2021)).

$$\begin{aligned} s_{ij}^{(l)+,P} &= \widehat{\Pr}[X_i^{(l)} > Y_j^{(l)} + \tau^{(l)} \mid \tilde{x}_i^{(l)}, \tilde{y}_j^{(l)}, \delta_i^{(l)}, \varepsilon_j^{(l)}] \\ s_{ij}^{(l)-,P} &= \widehat{\Pr}[X_i^{(l)} < Y_j^{(l)} - \tau^{(l)} \mid \tilde{x}_i^{(l)}, \tilde{y}_j^{(l)}, \delta_i^{(l)}, \varepsilon_j^{(l)}] \\ s_{ij}^{(l)0,P} &= \widehat{\Pr}[|X_i^{(l)} - Y_j^{(l)}| \leq \tau^{(l)} \mid \tilde{x}_i^{(l)}, \tilde{y}_j^{(l)}, \delta_i^{(l)}, \varepsilon_j^{(l)}] \end{aligned}$$

$s_{ij}^{(l)+,P}, s_{ij}^{(l)-,P}, s_{ij}^{(l)0,P}$ の推定量は, $\tilde{x}_i^{(l)}, \tilde{y}_j^{(l)}$ の大小関係と $\delta_i^{(l)}, \varepsilon_j^{(l)}$ の状態により異なる (表 2). 例えば, $\delta_i^{(l)} = 0, \varepsilon_j^{(l)} = 1$ かつ $\tilde{x}_i^{(l)} < \tilde{y}_j^{(l)} - \tau^{(l)}$ であった場合, 条件付き確率の定義より, $s_{ij}^{(l)+,P}$ は

$$\begin{aligned} s_{ij}^{(l)+,P} &= \widehat{\Pr}[X_i^{(l)} > Y_j^{(l)} + \tau^{(l)} \mid X_i^{(l)} > \tilde{x}_i^{(l)}, Y_j^{(l)} = \tilde{y}_j^{(l)}] \\ &= \frac{\widehat{\Pr}[X_i^{(l)} > \tilde{y}_j^{(l)} + \tau^{(l)}, X_i^{(l)} > \tilde{x}_i^{(l)}]}{\widehat{\Pr}[X_i^{(l)} > \tilde{x}_i^{(l)}]} \\ &= \frac{\hat{S}_{X^{(l)}}(\tilde{y}_j^{(l)} + \tau^{(l)}; \hat{\eta}_X^{(l)})}{\hat{S}_{X^{(l)}}(\tilde{x}_i^{(l)}; \hat{\eta}_X^{(l)})} \end{aligned} \quad (5)$$

となる. このとき, (5) 式の $s_{ij}^{(l)+,P}$ は, $\hat{S}_{X^{(l)}}$ に適当な推定量を用いることで計算でき, $\hat{\eta}_X^{(l)}$ は l 番目の評価項目の生存関数 $\hat{S}_{X^{(l)}}$ の推定におけるパラメータの推定量を表す (Ozenne *et al.* (2021))

また, $\delta_i^{(l)} = 1, \varepsilon_j^{(l)} = 0$ かつ $\tilde{x}_i^{(l)} \geq \tilde{y}_j^{(l)} - \tau^{(l)}$ であった場合, $s_{ij}^{(l)-,P}$ は

$$\begin{aligned} s_{ij}^{(l)-,P} &= \widehat{\Pr}[X_i^{(l)} < Y_j^{(l)} - \tau^{(l)} \mid X_i^{(l)} = \tilde{x}_i^{(l)}, Y_j^{(l)} > \tilde{y}_j^{(l)}] \\ &= \frac{\widehat{\Pr}[Y_j^{(l)} > \tilde{x}_i^{(l)} + \tau^{(l)}, Y_j^{(l)} > \tilde{y}_j^{(l)}]}{\widehat{\Pr}[Y_j^{(l)} > \tilde{y}_j^{(l)}]} \\ &= \frac{\hat{S}_{Y^{(l)}}(\tilde{x}_i^{(l)} + \tau^{(l)}; \hat{\boldsymbol{\eta}}_Y^{(l)})}{\hat{S}_{Y^{(l)}}(\tilde{y}_j^{(l)}; \hat{\boldsymbol{\eta}}_Y^{(l)})} \end{aligned} \quad (6)$$

となる. ただし, $\hat{\boldsymbol{\eta}}_Y^{(l)}$ は $\hat{S}_{Y^{(l)}}$ のパラメータの推定量である (Ozenne *et al.* (2021)).

他の場合の $s_{ij}^{(l)+,P}, s_{ij}^{(l)-,P}$ も同様に導出可能であり, 詳細は Péron らの論文で確認できる (Péron *et al.* (2018), Ozenne *et al.* (2021)). Péron らは, $\hat{S}_{X^{(l)}}(t; \hat{\boldsymbol{\eta}}_X^{(l)}), \hat{S}_{Y^{(l)}}(t; \hat{\boldsymbol{\eta}}_Y^{(l)})$ の推定に Kaplan-Meier 法を用いることを提案しており (Péron *et al.* (2018, 2021)), 以降本稿でも Péron 流スコアでは Kaplan-Meier 法を用いるものとする.

Kaplan-Meier 推定に基づく Péron 流スコアでは, 場合によって, $s_{ij}^{(l)+,P}, s_{ij}^{(l)-,P}, s_{ij}^{(l)0,P}$ が計算できないペアが存在することがある. 例えば, (5) 式の $\hat{S}_{X^{(l)}}(\tilde{y}_j^{(l)} + \tau^{(l)}; \hat{\boldsymbol{\eta}}_X^{(l)})$ において, 新治療群における評価項目 $X^{(l)}$ の最大観察値 $\max_i(\tilde{x}_i^{(l)})$ が打ち切りであった場合, $\tilde{y}_j^{(l)} + \tau^{(l)} > \max_i(\tilde{x}_i^{(l)})$ となる j に対して $\hat{S}_{X^{(l)}}(\tilde{y}_j^{(l)} + \tau^{(l)}; \hat{\boldsymbol{\eta}}_X^{(l)})$ は推定できないため, スコアが計算できない (Péron *et al.* (2018, 2021)). このようなペアを Péron 流スコアにおける無情報ペアと定義し, 以下の指示変数で表す.

$$u_{ij}^{(l),P} = 1 - (s_{ij}^{(l)+,P} + s_{ij}^{(l)-,P} + s_{ij}^{(l)0,P})$$

Péron 流スコアの無情報ペア $u_{ij}^{(l),P}$ は, Gehan 流スコアの無情報ペア $u_{ij}^{(l),G}$ とは異なり, 一般には Gehan 流スコアと比べて, Péron 流スコアの無情報ペアの総数は少なくなる (Péron *et al.* (2018, 2021)). これは, Péron 流スコアは, Gehan 流スコアにおける無情報ペア ($u_{ij}^{(l),G} = 1$) に対して, Kaplan-Meier 推定量を用いた補完により 0 から 1 の間の値をとることによる (表 2). なお, Gehan 流スコアにおける有情報ペア ($u_{ij}^{(l),G} = 0$) では, Péron 流スコアは Gehan 流スコアと等しく, 1 もしくは 0 の値をとる.

二つの評価項目から勝ち負けを総合的に評価した Péron 流スコアは,

$$s_{ij}^{+,P} = s_{ij}^{(1)+,P} + w_{ij}^{(2),P} \times s_{ij}^{(2)+,P}, \quad s_{ij}^{-,P} = s_{ij}^{(1)-,P} + w_{ij}^{(2),P} \times s_{ij}^{(2)-,P} \quad (7)$$

となる. ただし, 2 番目の評価項目に対する重みは, $w_{ij}^{(2),P} = s_{ij}^{(1)0,P} + u_{ij}^{(1),P}$ とする. Gehan 流スコアと同様, 最下位 (2 番目) の評価項目での無情報ペアは (7) 式のスコアには寄与しない. (7) 式のスコアを (1)(2) 式の統計量の計算に用いることで Net Benefit および Win Ratio の推定量が計算できる.

Péron らは, 独立打ち切りの条件下において, Péron 流スコアに基づく Net Benefit の推定量は Gehan 流スコアに基づく推定量よりもバイアスを軽減することをシミュレーショ

ン実験により示している (Péron *et al.* (2018, 2021)). ただし, 打ち切りの割合が多い場合は, Péron 流スコアに基づく Net Benefit の推定量においても残存するバイアスは少なくない (Péron *et al.* (2021)).

2.3.4 補正版 Péron 流スコア

2.3.3 節で述べた Péron 流のスコアに対し, Péron 流の無情報ペアのスコアを他のペアに割り振る補正方法を Péron らは提案している (Péron *et al.* (2021)). 提案では, 無情報ペアでの勝ち, 負け, 引き分けの確率が, 有情報ペアの勝ち, 負け, 引き分けの確率と同様になるという仮定に基づき, 以下のように計算される有情報ペアの補正版スコア $s_{ij}^{(l)+,Pc}$, $s_{ij}^{(l)-,Pc}$, $s_{ij}^{(l)0,Pc}$ を Net Benefit や Win Ratio の推定に用いる.

$$\begin{aligned} s_{ij}^{(l)+,Pc} &= s_{ij}^{(l)+,P} + u_{ij}^{(l),P} \frac{\hat{U}^{(l)+}}{\hat{U}^{(l)+} + \hat{U}^{(l)-} + \hat{U}^{(l)0}} \\ s_{ij}^{(l)-,Pc} &= s_{ij}^{(l)-,P} + u_{ij}^{(l),P} \frac{\hat{U}^{(l)-}}{\hat{U}^{(l)+} + \hat{U}^{(l)-} + \hat{U}^{(l)0}} \\ s_{ij}^{(l)0,Pc} &= s_{ij}^{(l)0,P} + u_{ij}^{(l),P} \frac{\hat{U}^{(l)0}}{\hat{U}^{(l)+} + \hat{U}^{(l)-} + \hat{U}^{(l)0}} \end{aligned}$$

ここで,

$$\hat{U}^{(l)+} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{(l)+,P}, \quad \hat{U}^{(l)-} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{(l)-,P}, \quad \hat{U}^{(l)0} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{(l)0,P}$$

は, 有情報ペアから計算される, 勝ち, 負け, 引き分けの確率の推定値である.

二つの評価項目から勝ち負けを総合的に評価した補正版 Péron 流のスコアは,

$$s_{ij}^{+,Pc} = s_{ij}^{(1)+,Pc} + w_{ij}^{(2),Pc} \times s_{ij}^{(2)+,Pc}, \quad s_{ij}^{-,Pc} = s_{ij}^{(1)-,Pc} + w_{ij}^{(2),Pc} \times s_{ij}^{(2)-,Pc} \quad (8)$$

となる. ただし, l 番目の評価項目に対する重みは, $w_{ij}^{(2),P} = s_{ij}^{(1)0,Pc}$ である. 上記の補正により, Péron 流スコアにおける無情報ペアのスコアを, 勝ち, 負け, 引き分けのペアに配分したため, (8) 式の重みには補正後の引き分けのペアのみが反映されることになる.

Péron らは, 無情報ペアにおける勝ち, 負け, 引き分けの確率が, 有情報ペアにおけるそれぞれの確率と同じであるという仮定が正しければ, 補正版 Péron 流スコアに基づく Net Benefit の推定量が補正なしの Péron 流スコアよりもバイアスが軽減されるということをシミュレーション実験により示している (Péron *et al.* (2021)). 一方で, この仮定が成立しているかどうかは観察データから確認できないため, 盲目的に補正版を用いるべきではないとも指摘している (Péron *et al.* (2021)). 補正版 Péron 流スコアの仮定の妥当性は, 生存時間が指数分布に従う場合は正当化されるが, 生存曲線が右端で平坦になる (治癒がみられる) 状況においては仮定の成立は疑わしいと議論されている (Péron *et al.* (2021)).

2.4 生存関数による Net Benefit および Win Ratio の推定

2.2 節, 2.3 節で説明したスコアに基づく Net Benefit や Win Ratio の推定方法以外に, 生存関数による推定方法が提案されている (Oakes (2016), Finkelstein and Schoenfeld (2019), Dong *et al.* (2020b)). 生存関数による方法は, 評価項目が一つの場合, 勝ち, 負けの確率が, 各群の生存関数を用いて,

$$\Pr[X > Y + \tau] = \int_t -S_X(t + \tau) dS_Y(t), \quad \Pr[Y > X + \tau] = \int_t -S_Y(t + \tau) dS_X(t)$$

と表されることを利用して, 適当な分布関数の推定量 (例えば, Kaplan-Meier 推定量など) を上式に代入し, Net Benefit あるいは Win Ratio の推定量を求める方法である (Oakes (2016), Finkelstein and Schoenfeld (2019), Dong *et al.* (2020b)). この推定量は, 評価項目が二つ以上の場合への拡張も提案されているが (Oakes (2016), Finkelstein and Schoenfeld (2019)), 複数評価項目における同時分布の推定や, 漸近論に基づく分散の推定が困難になることが指摘されている (Dong *et al.* (2020b)).

3. Net Benefit および Win Ratio の推定量の分散

Net Benefit あるいは Win Ratio に対する仮説検定や区間推定は, ブートストラップ法や並べ替え検定によるリサンプリング法に基づくものと, U 統計量理論に基づく漸近分散の推定量を用いるものに大別される (Verbeek *et al.* (2020), Ozenne *et al.* (2021)). 一般化ペアワイズ比較に基づく Net Benefit および Win Ratio の推定量は, (1)(2) 式による二つの U 統計量の差および比として表すことができるため, 一般の U 統計量理論をもとに漸近分散の推定量を導出できる.

3.1 節では, (1)(2) 式で示した $\Pr[\mathbf{X} \succ \mathbf{Y}]$, $\Pr[\mathbf{X} \prec \mathbf{Y}]$ の推定量 \hat{U}^+ , \hat{U}^- が従う漸近分布と, 打ち切りがない場合の \hat{U}^+ , \hat{U}^- の漸近分散が 2.2 節のスコアを用いて推定できることを説明し, Net Benefit や Win Ratio の分散の推定との関連を説明する. 3.2 節では, 打ち切りがある場合の漸近分散の推定量を 2.3 節で導入した Gehan, Dong, Péron 流スコアを用いて計算する方法を説明する.

3.1 勝ち負けの確率の推定量の漸近分散

まず, 打ち切りのない場合における (1)(2) 式の \hat{U}^+ , \hat{U}^- の漸近分布を考える. \hat{U}^+ と \hat{U}^- はそれぞれ 2 変量 U 統計量であることから, それらの漸近同時分布は以下の 2 変量正規分布となる (Bebu and Lachin (2016))

$$\begin{pmatrix} \hat{U}^+ \\ \hat{U}^- \end{pmatrix} \sim N \left(\begin{pmatrix} \Pr[\mathbf{X} \succ \mathbf{Y}] \\ \Pr[\mathbf{X} \prec \mathbf{Y}] \end{pmatrix}, \begin{pmatrix} \sigma_{\hat{U}^+, \hat{U}^+} & \sigma_{\hat{U}^+, \hat{U}^-} \\ \sigma_{\hat{U}^-, \hat{U}^+} & \sigma_{\hat{U}^-, \hat{U}^-} \end{pmatrix} \right) \quad (9)$$

$\sigma_{\hat{U}^+, \hat{U}^+}$ は以下のように分解できる (Bebu and Lachin (2016)).

$$\sigma_{\hat{U}^+, \hat{U}^+} = \frac{\sigma_{+, +'}}{m} + \frac{\sigma_{+', +}}{n} \quad (10)$$

ただし,

$$\begin{aligned} \sigma_{+, +' } &= \text{Cov}[I(\mathbf{X} \succ \mathbf{Y}), I(\mathbf{X} \succ \mathbf{Y}')] \\ &= E[I(\mathbf{X} \succ \mathbf{Y})I(\mathbf{X} \succ \mathbf{Y}')] - E[I(\mathbf{X} \succ \mathbf{Y})]E[I(\mathbf{X} \succ \mathbf{Y}')] \\ &= \Pr[\mathbf{X} \succ \mathbf{Y}, \mathbf{X} \succ \mathbf{Y}'] - (\Pr[\mathbf{X} \succ \mathbf{Y}])^2 \\ \sigma_{+', +} &= \text{Cov}[I(\mathbf{X} \succ \mathbf{Y}), I(\mathbf{X}' \succ \mathbf{Y})] \\ &= E[I(\mathbf{X} \succ \mathbf{Y})I(\mathbf{X}' \succ \mathbf{Y})] - E[I(\mathbf{X} \succ \mathbf{Y})]E[I(\mathbf{X}' \succ \mathbf{Y})] \\ &= \Pr[\mathbf{X} \succ \mathbf{Y}, \mathbf{X}' \succ \mathbf{Y}] - (\Pr[\mathbf{X} \succ \mathbf{Y}])^2 \end{aligned}$$

である. \mathbf{X}', \mathbf{Y}' はそれぞれ \mathbf{X}, \mathbf{Y} と同一の確率分布に従う確率変数である. $\sigma_{\hat{U}^+, \hat{U}^-}$ と $\sigma_{\hat{U}^-, \hat{U}^-}$ についても同様に分解可能であり, 詳細は付録 B に記載した.

Net Benefit の推定量の漸近分散は, (9) 式より,

$$\sigma_{\hat{\Delta}, \hat{\Delta}} = \text{Var}[\hat{U}^+ - \hat{U}^-] \approx \sigma_{\hat{U}^+, \hat{U}^+} + \sigma_{\hat{U}^-, \hat{U}^-} - 2\sigma_{\hat{U}^+, \hat{U}^-} \quad (11)$$

である. Win Ratio の推定量は \hat{U}^+ と \hat{U}^- の比であるため, 対数変換後の漸近分散 $\sigma_{\ln \hat{\Psi}, \ln \hat{\Psi}}$ を以下のように求める.

$$\begin{aligned} \sigma_{\ln \hat{\Psi}, \ln \hat{\Psi}} &= \text{Var}[\log(\hat{U}^+ / \hat{U}^-)] \\ &= \text{Var}[\log(\hat{U}^+)] + \text{Var}[\log(\hat{U}^-)] - 2 \text{Cov}[\log(\hat{U}^+), \log(\hat{U}^-)] \\ &\approx \frac{\sigma_{\hat{U}^+, \hat{U}^+}}{(\Pr[\mathbf{X} \succ \mathbf{Y}])^2} + \frac{\sigma_{\hat{U}^-, \hat{U}^-}}{(\Pr[\mathbf{X} \prec \mathbf{Y}])^2} - \frac{2\sigma_{\hat{U}^+, \hat{U}^-}}{\Pr[\mathbf{X} \succ \mathbf{Y}] \Pr[\mathbf{X} \prec \mathbf{Y}]} \end{aligned} \quad (12)$$

(12) 式の近似にはデルタ法による一次近似を用いている (Bebu and Lachin (2016), Dong *et al.* (2016)). したがって, Net Benefit および Win Ratio の推定量の漸近分散をデータから推定するには, $\sigma_{\hat{U}^+, \hat{U}^+}, \sigma_{\hat{U}^+, \hat{U}^-}, \sigma_{\hat{U}^-, \hat{U}^-}$ の推定, すなわち, $\sigma_{+, +'}$ 等をデータから推定する必要がある. ここでは Ozenne らによる方法 (Ozenne *et al.* (2021)) と Dong らによる方法 (Dong *et al.* (2016)) を説明する.

3.1.1 Ozenne らによる推定量

Ozenne らは,

$$\widehat{\Pr}[\mathbf{X} \succ \mathbf{Y}, \mathbf{X} \succ \mathbf{Y}'] = \frac{1}{mn^2} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1}^n s_{ij}^+ s_{ij'}^+, \quad \widehat{\Pr}[\mathbf{X} \succ \mathbf{Y}] = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^+$$

などの推定量を用い, $\sigma_{+,+}$ と $\sigma_{+,+}$ の推定量として以下を提案した (Ozenne *et al.* (2021))

$$\begin{aligned}\hat{\sigma}_{+,+} &= \frac{1}{mn^2} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1}^n s_{ij}^+ s_{ij'}^+ - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^+ \right)^2 \\ \hat{\sigma}_{+,+} &= \frac{1}{m^2 n} \sum_{i=1}^m \sum_{i'=1}^m \sum_{j=1}^n s_{ij}^+ s_{i'j}^+ - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^+ \right)^2\end{aligned}$$

これらを (10) 式に代入し, $\hat{\sigma}_{\hat{U}^+, \hat{U}^+}$ を推定する. $\hat{\sigma}_{\hat{U}^-, \hat{U}^-}$ および $\hat{\sigma}_{\hat{U}^+, \hat{U}^-}$ についても同様に推定するが, その詳細は付録 B に記載した. これらの推定量を (11)(12) 式に代入することで, Net Benefit および対数変換後の Win Ratio の漸近分散を推定できる.

Ozenne らは, 打ち切りのない状況におけるシミュレーション実験により, 提案する漸近分散の推定量は, 小標本であっても経験分散に近い推定値を与えることを示している (Ozenne *et al.* (2021)).

3.1.2 Dong らによる推定量

Dong らの推定方法は, Ozenne らとは, $\hat{\sigma}_{+,+}$ および $\hat{\sigma}_{+,+}$ の計算方法, $\Pr[\mathbf{X} \succ \mathbf{Y}]$ の推定における仮定が異なる. 例えば, $\hat{\sigma}_{+,+}$ は

$$\widehat{\Pr}[\mathbf{X} \succ \mathbf{Y}, \mathbf{X} \succ \mathbf{Y}'] = \frac{1}{mn(n-1)} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1, j' \neq j}^n s_{ij}^+ s_{ij'}^+$$

と \mathbf{Y} と \mathbf{Y}' が同じ対象者であることを許容しない点が異なる. また, $\Pr[\mathbf{X} \succ \mathbf{Y}]$ の推定量は, $\Pr[\mathbf{X} \succ \mathbf{Y}] = \Pr[\mathbf{X} \prec \mathbf{Y}]$ という帰無仮説の下で

$$\widehat{\Pr}[\mathbf{X} \succ \mathbf{Y}] = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2}$$

という推定量を考える. このような推定量を用いて, $\sigma_{+,+}$ と $\sigma_{+,+}$ を以下のように計算する (Dong *et al.* (2016)).

$$\begin{aligned}\hat{\sigma}_{+,+} &= \frac{1}{mn(n-1)} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1, j' \neq j}^n s_{ij}^+ s_{ij'}^+ - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2 \\ \hat{\sigma}_{+,+} &= \frac{1}{m(m-1)n} \sum_{i=1}^m \sum_{i'=1, i' \neq i}^m \sum_{j=1}^n s_{ij}^+ s_{i'j}^+ - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2\end{aligned}$$

これらを (10) 式に代入することで $\hat{\sigma}_{\hat{U}^+, \hat{U}^+}$ を推定する. $\hat{\sigma}_{\hat{U}^-, \hat{U}^-}$ および $\hat{\sigma}_{\hat{U}^+, \hat{U}^-}$ の推定量の詳細は付録 B に記載した. これらの推定量を, (11)(12) 式に代入することで Net Benefit および対数変換後の Win Ratio の漸近分散の推定量が得られる.

Verbeek らは, 打ち切りのない状況でのシミュレーション実験により, Dong らの推定量に基づく Net Benefit および Win Ratio の漸近分散推定量の性能を, 信頼区間の被覆確

率および検定の第一種の過誤確率の観点から評価している (Verbeeck *et al.* (2020)). U 統計量理論の漸近論に基づく方法であることから、両群合わせた例数が 50 例以下の場合では推定量の性能が低下することがあると指摘している。また、帰無仮説の下での推定量を用いていることから、Net Benefit および Win Ratio の真値が帰無仮説の値から大きく離れる設定においても性能が落ち、Net Benefit の信頼区間では被覆確率の上昇が、Win Ratio の信頼区間としては被覆確率の低下が見られたと報告している (Verbeeck *et al.* (2020)).

3.2 漸近分散の推定における打ち切りの考慮

打ち切りのある状況下で、2.3 節で述べた Gehan, Dong, Péron 流のスコアリングによる Net Benefit および Win Ratio の推定量を用いる場合の漸近分散の推定量を考える。

3.2.1 Gehan 流スコアに基づく推定量の漸近分散

Gehan 流スコアに基づく Net Benefit および Win Ratio 統計量に対しては、3.1 節で説明した Ozenne らまたは Dong らによる漸近分散の推定量における s_{ij}^+, s_{ij}^- に対して、(3) 式の Gehan 流スコア $s_{ij}^{+,G}, s_{ij}^{-,G}$ を代入することで漸近分散を推定する方法がそれぞれ提案されている (Dong *et al.* (2016), Ozenne *et al.* (2021)). Ozenne らによる方法は R パッケージ *BuyseTest* で (Ozenne and Péron (2022)), Dong らによる方法は R パッケージ *WINS* で実装されている (Cui and Huang (2022)).

Ozenne らの方法による Net Benefit 統計量に対する漸近分散の推定量は、打ち切りを考慮した設定のシミュレーション実験により統計的な性能が評価されており、 $\Delta = 0$ という帰無仮説の下でのみ推定分散に基づく信頼区間の被覆確率が名目水準に近くなることが報告されている (Ozenne *et al.* (2021)). なお、 $\Delta \neq 0$ の対立仮説の下では、打ち切りがある場合、Gehan 流スコアに基づく Net Benefit の点推定値にバイアスがあるため、Ozenne らの分散の性能評価は困難である (Ozenne *et al.* (2021)). Dong らの方法による Win Ratio 統計量に対する漸近分散の推定量については、 $\Psi \neq 1$ の対立仮説の下で特に打ち切り割合が多いときに推定分散に基づく信頼区間の被覆確率が名目水準を下回ることがシミュレーション実験により報告されている (Dong *et al.* (2020c)).

3.2.2 Dong 流スコアに基づく推定量の漸近分散

IPCW 法に基づく Dong 流スコアに基づく Win Ratio に対しては、3.1.2 節での Dong らによる推定量に (4) 式の $s_{ij}^{+,D}, s_{ij}^{-,D}$ を代入することで、漸近分散を推定する方法が提案されている (Dong *et al.* (2020c, 2021)). この方法は、R パッケージ *WINS* で実装されている (Cui and Huang (2022)). Dong 流スコアに基づく Win Ratio 統計量に対する漸近分散の推定量については、推定分散に基づく信頼区間の被覆確率が名目水準に近くなることがシミュレーション実験により報告されている (Dong *et al.* (2020c)).

3.2.3 Péron 流スコアに基づく推定量の漸近分散

Péron 流スコアに基づく Net Benefit の推定量に対して、Ozenne らは U 統計量に推定パラメータを含む場合の漸近分散を導出している (Ozenne *et al.* (2021)). 簡単のため、本節では評価項目がただ一つの場合における Ozenne らの方法を説明し、評価項目を表す添字 l も省略する. 二つの評価項目の場合への拡張については、Ozenne らの論文で確認できる (Ozenne *et al.* (2021)).

Ozenne らの方法では、2.3.3 節で説明した Péron 流スコアに Kaplan-Meier 推定量が用いられていることを考慮している. Péron 流スコアが推定量に基づくことを明示するため、本節では Péron 流スコアを $s_{ij}^{+,P}(\boldsymbol{\eta}), s_{ij}^{-,P}(\boldsymbol{\eta})$ (ただし、 $\boldsymbol{\eta} = \{\boldsymbol{\eta}_X, \boldsymbol{\eta}_Y\}$ である) と表記する. Péron 流スコアに対する漸近分散を導く上で、Net Benefit の真値 Δ に対する推定量 $\hat{\Delta}(\hat{\boldsymbol{\eta}})$ の誤差を以下のように分解する.

$$\hat{\Delta}(\hat{\boldsymbol{\eta}}) - \Delta = (\hat{\Delta}(\hat{\boldsymbol{\eta}}) - \hat{\Delta}(\boldsymbol{\eta})) + (\hat{\Delta}(\boldsymbol{\eta}) - \Delta) \quad (13)$$

(13) 式の右辺第 1 項は、Péron 流スコアにおける生存関数の推定に伴う誤差であり、以下のように近似的に表される.

$$\hat{\Delta}(\hat{\boldsymbol{\eta}}) - \hat{\Delta}(\boldsymbol{\eta}) \approx \frac{1}{m} \sum_{i=1}^m IF^{\hat{\boldsymbol{\eta}}_X}(x_i) \nabla_{\boldsymbol{\eta}_X} \hat{\Delta}(\boldsymbol{\eta}) \big|_{\boldsymbol{\eta}_X = \hat{\boldsymbol{\eta}}_X} + \frac{1}{n} \sum_{j=1}^n IF^{\hat{\boldsymbol{\eta}}_Y}(y_j) \nabla_{\boldsymbol{\eta}_Y} \hat{\Delta}(\boldsymbol{\eta}) \big|_{\boldsymbol{\eta}_Y = \hat{\boldsymbol{\eta}}_Y}$$

ここで、 $IF^{\hat{\boldsymbol{\eta}}_X}, IF^{\hat{\boldsymbol{\eta}}_Y}$ は生存関数のパラメータの推定量 $\hat{\boldsymbol{\eta}}_X, \hat{\boldsymbol{\eta}}_Y$ に対する影響関数 (influence function) である (Ozenne *et al.* (2021)). また、 $\nabla_{\boldsymbol{\eta}_X}$ は $\boldsymbol{\eta}_X$ の要素に対する偏微分を列ベクトルとしたものである. この式は影響関数に対する連鎖率 (chain rule) を用いることで導かれる (Ozenne *et al.* (2021)).

(13) 式の右辺第 2 項は、Péron 流スコアの算出に用いる生存関数のパラメータの真値を所与とした下での Net Benefit の推定量と真値との誤差であり、 U 統計量理論により以下のように分解できる.

$$\hat{\Delta}(\boldsymbol{\eta}) - \Delta \approx \frac{1}{m} \sum_{i=1}^m h_1^{\hat{\Delta}}(x_i; \boldsymbol{\eta}) + \frac{1}{n} \sum_{j=1}^n h_1^{\hat{\Delta}}(y_j; \boldsymbol{\eta})$$

ただし、

$$\begin{aligned} h_1^{\hat{\Delta}}(x_i; \boldsymbol{\eta}) &= (E_Y[s_{ij}^{+,P}(\boldsymbol{\eta}) \mid x_i] - \Pr[X > Y + \tau]) - (E_Y[s_{ij}^{-,P}(\boldsymbol{\eta}) \mid x_i] - \Pr[X < Y - \tau]), \\ h_1^{\hat{\Delta}}(y_j; \boldsymbol{\eta}) &= (E_X[s_{ij}^{+,P}(\boldsymbol{\eta}) \mid y_j] - \Pr[X > Y + \tau]) - (E_X[s_{ij}^{-,P}(\boldsymbol{\eta}) \mid y_j] - \Pr[X < Y - \tau]) \end{aligned}$$

である.

以上から, Péron 流スコアに基づく Net Benefit の推定量を以下のように分解できる.

$$\begin{aligned}\hat{\Delta}(\hat{\eta}) - \Delta &\approx \frac{1}{m} \sum_{i=1}^m \{h_1^{\hat{\Delta}}(x_i; \eta) + IF^{\hat{\eta}^X}(x_i) \nabla_{\eta^X} \hat{\Delta}(\eta) |_{\eta^X = \hat{\eta}^X}\} \\ &\quad + \frac{1}{n} \sum_{j=1}^n \{h_1^{\hat{\Delta}}(y_j; \eta) + IF^{\hat{\eta}^Y}(y_j) \nabla_{\eta^Y} \hat{\Delta}(\eta) |_{\eta^Y = \hat{\eta}^Y}\}\end{aligned}$$

この近似を用いて, Ozenne らは以下の Net Benefit の漸近分散の推定量を導いた.

$$\begin{aligned}\hat{\sigma}_{\hat{\Delta}(\hat{\eta}), \hat{\Delta}(\hat{\eta})} &= \frac{1}{m^2} \sum_{i=1}^m \{\hat{h}_1^{\hat{\Delta}}(x_i; \hat{\eta}) + IF^{\hat{\eta}^X}(x_i) \nabla_{\eta^X} \hat{\Delta}(\eta) |_{\eta^X = \hat{\eta}^X}\}^2 \\ &\quad + \frac{1}{n^2} \sum_{j=1}^n \{\hat{h}_1^{\hat{\Delta}}(y_j; \hat{\eta}) + IF^{\hat{\eta}^Y}(y_j) \nabla_{\eta^Y} \hat{\Delta}(\eta) |_{\eta^Y = \hat{\eta}^Y}\}^2\end{aligned}$$

ただし,

$$\begin{aligned}h_1^{\hat{\Delta}}(x_i; \hat{\eta}) &= \left(\frac{1}{n} \sum_{j=1}^n s_{ij}^{+,P}(\hat{\eta}) - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{+,P}(\hat{\eta}) \right) \\ &\quad - \left(\frac{1}{n} \sum_{j=1}^n s_{ij}^{-,P}(\hat{\eta}) - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{-,P}(\hat{\eta}) \right) \\ h_1^{\hat{\Delta}}(y_j; \hat{\eta}) &= \left(\frac{1}{m} \sum_{i=1}^m s_{ij}^{+,P}(\hat{\eta}) - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{+,P}(\hat{\eta}) \right) \\ &\quad - \left(\frac{1}{m} \sum_{i=1}^m s_{ij}^{-,P}(\hat{\eta}) - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^{-,P}(\hat{\eta}) \right)\end{aligned}$$

である.

実際に Péron 流スコアを用いる場合の影響関数 $IF^{\hat{\eta}^X}$, $IF^{\hat{\eta}^Y}$ や $\hat{h}_1^{\hat{\Delta}}(x_i; \hat{\eta})$, $\hat{h}_1^{\hat{\Delta}}(y_j; \hat{\eta})$ の詳細な導出については Ozenne らおよびその引用文献を参照されたい (Ozenne *et al.* (2021)).

Ozenne らは, シミュレーション実験により提案した分散の推定量の性能を信頼区間の被覆確率の観点から評価している (Ozenne *et al.* (2021)). $\tau = 0$ の場合では被覆確率が十分に担保される一方で, τ が大きくなるにともなって数%の被覆確率の低下が見られたこと, 各群の例数が 400 例超となる場合では $\tau > 0$ であっても被覆確率が名目上の値に近くなったこと等を報告している. なお, Péron 流スコアに対する Ozenne らの分散推定量は, R パッケージ *BuyseTest* で実装されている (Ozenne and Péron (2022)).

3.2.4 補正版 Péron 流スコアに基づく推定量の漸近分散

3.2.3 節で説明した Ozenne らの分散の推定量は, 補正なしの Péron 流スコアに基づく Net Benefit に対する性能評価がされている. 2.2.4 節で述べた補正版 Péron 流スコアに基づく推定量に対しては, U 統計量理論に基づく分散の推定量などの妥当性は評価されて

いない (Ozenne *et al.* (2021)). Péron らは、補正版 Péron 流スコアの分散の推定や信頼区間の算出においては、リサンプリングによるブートストラップ法を用いるべきであると述べている (Péron *et al.* (2021)). 実際、Péron 流スコアに基づく推定量を実装している *BuyseTest* パッケージにおいては、*BuyseTest* 関数のオプションにおいて、2.2.4 節による補正 (`correction.uninf = TRUE`) と *U* 統計量に基づく漸近分散 (`inference = 'u-statistic'`) を同時に指定すると Warning メッセージが出る仕様となっている。

4. Net Benefit と Win Ratio の実データへの適用

本節では、一般化ペアワイズ比較に基づく Net Benefit および Win Ratio が実装されている R パッケージのうち、2 節と 3 節で紹介した方法を実装している *BuyseTest* パッケージおよび *WINS* パッケージを用いて、実際に実施された、がん臨床試験の V325 試験 (4.1 節)、循環器臨床試験の HF-ACTION 試験 (4.2 節) のデータに適用した結果を例示する。V325 試験の個別データは、R の *surrosurv* パッケージ (Rotolo *et al.* (2018)) に含まれる *gastadv* データセットより利用できる。HF-ACTION 試験の個別データは、非虚血性高リスク心不全患者の部分集団 426 例からなるデータが、R の *WR* パッケージにおける *hfaction.cpx9* データセットより利用できる。なお、解析で用いたパッケージのバージョンは、*BuyseTest* (ver. 2.3.11) および *WINS* (ver. 1.3.2) である。

4.1 がん臨床試験への適用

V325 試験は、進行胃癌患者に対する 1 次治療として、Docetaxel + Cisplatin + 5-FU 療法 (新治療群) と Cisplatin + 5-FU 療法 (標準治療群) を比較したランダム化比較試験である (Van Cutsem *et al.* (2006)). 公開されているデータからは、全生存期間 (Overall Survival; OS) および無増悪生存期間 (progression-free-survival; PFS) の生存時間評価項目を利用できる。ここでは、上位の評価項目を死亡までの時間、下位の評価項目を腫瘍増悪までの時間とし、これらを代替する OS と PFS のデータを用いて Net Benefit と Win Ratio を求めた結果を報告する。

表 3 および図 1 に OS と PFS に対する記述的な解析の結果を示す。いずれの群においても、OS および PFS ともに 7 割以上の対象者でイベントが発生していた。Cox 回帰によるハザード比の推定値は、OS では 0.81 (95%信頼区間; 0.65–1.00)、PFS では 0.81 (95%信頼区間; 0.66–0.99) であった。

表 4 に Net Benefit および Win Ratio を推定した結果を示す。ここでは、OS (PFS) が 1 日でも長い場合には「勝ち」とするルールを想定し、いずれの評価項目に対しても τ を 0 とした。表 4 の 3–6 列目は、比較ペアの組み合わせ総数 52210 のうち、各評価項目で勝ち、負け、引き分け、無情報と分類されるペア数の合計と割合を、Gehan 流、Péron 流、補正版

表 3 V325 試験の生存時間データの要約.

		新治療群 (n = 227)	標準治療群 (n = 230)
OS	イベント数 (%)	167 (74%)	176 (77%)
	中央値 (カ月) (95%CI)	9.0 (8.2, 10.6)	8.6 (7.3, 9.5)
	ハザード比 (95%CI)	0.81 (0.65, 1.00)	
PFS	イベント数 (%)	194 (85%)	197 (86%)
	中央値 (カ月) (95%CI)	5.7 (5.0, 6.3)	4.0 (3.7, 5.4)
	ハザード比 (95%CI)	0.81 (0.66, 0.99)	

Survival パッケージ(ver 3.2.13)を用い, 中央値は Kaplan-Meier 法, ハザード比は Cox 回帰により推定した.

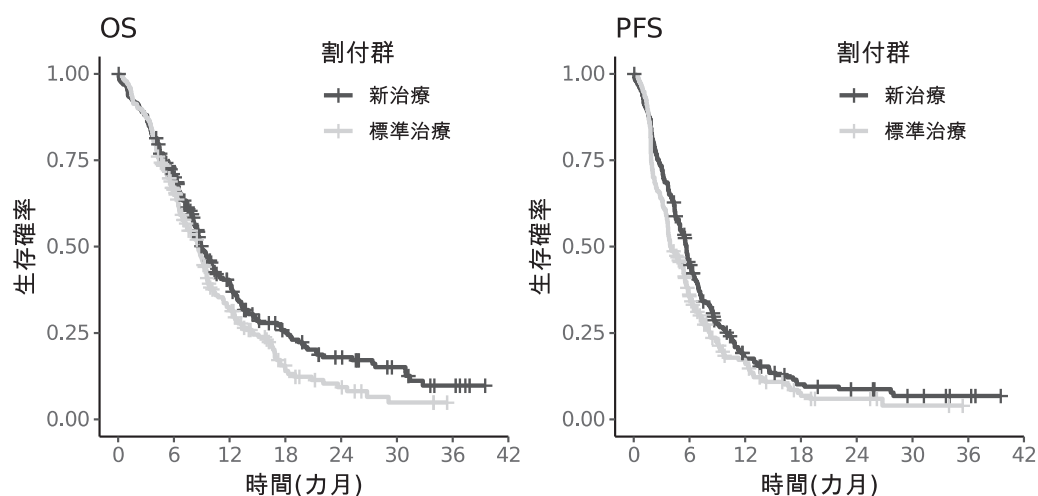


図 1 V325 試験の生存時間データに対する Kaplan-Meier 推定値.

Péron 流のスコアリング方法ごとに示している. Dong 流スコアを計算できる *WINS* パッケージは, 評価項目ごとに分類, 計算した各ペアの結果を出力しないため, Net Benefit と Win Ratio の推定値のみ示している.

表 4 から, OS に対する Gehan 流スコアでは, 打ち切りにより勝ち, 負け, 引き分けに分類できない Gehan 流スコアにおける無情報ペアが全ペアの 18.2% 存在したことが分かる. 一方で, Péron 流スコアにおける無情報ペアは 0.5% まで減少した. 補正版 Péron 流スコア

表 4 V325 試験への一般化ペアワイズ比較の適用結果 (比較ペアの総数 52210).

スコアリング方法	評価項目	勝ち	負け	引き分け	無情報	Net Benefit (95% CI)	Win Ratio (95% CI)
Gehan	OS	43.9% (22902)	37.8% (19755)	0.1% (54)	18.2% (9499)		
	PFS	5.8% (3011)	5.4% (2838)	0.0% (14)	7.1% (3690)		
	総合	49.6% (25913)	43.3% (22593)	0.1% (68)		6.36% (-3.9%, 16.5%)	1.15 (0.92, 1.43)
Péron	OS	54% (28171)	45.4% (23722)	0.1% (65)	0.5% (251)		
	PFS	0.3% (147)	0.2% (127)	0.0% (6)	0.1% (37)		
	総合	54.2% (28318)	45.7% (23849)	0.1% (71)		8.56% (-2.5%, 19.4%)	1.19 (0.95, 1.48)
補正版 Péron	OS	54.2% (28308)	45.7% (23837)	0.1% (66)	0.0% (0)		
	PFS	0.1% (32)	0.0% (28)	0.0% (6)	0.0% (0)		
	総合	54.3% (28340)	45.7% (23864)	0.1% (72)		8.57% (-2.6%, 19.4%)	1.19 (0.95, 1.48)
Dong	総合					6.28% (-4.0%, 16.6%)	1.14 (0.92, 1.42)

* 推定値に対する 95%信頼区間は, Gehan および Péron に対しては, 3.2.1 節および 3.2.3 節の Ozenne らによる漸近分散の推定量を, 補正版 Péron に対してはブートストラップ法に基づく方法を, Dong に対しては 3.2.2 節の Dong らによる漸近分散の推定量をそれぞれ用いた. Dong 流スコアは WINS パッケージ, それ以外の方法は BuyseTest パッケージにより計算した.

では, Péron 流スコアにおける 251 の無情報ペアのスコアを,

$$\text{勝ち} : 54.2\% (= 28171 / (28171 + 23722 + 65))$$

$$\text{負け} : 45.7\% (= 23722 / (28171 + 23722 + 65))$$

$$\text{引き分け} : 0.1\% (= 65 / (28171 + 23722 + 65))$$

という割合で各カテゴリーに配分することで, 無情報ペア数を 0 とみなし, 勝ち, 負け, 引き分けのペア数を増加させた. なお, OS での比較結果が引き分けもしくは無情報のペアで PFS の比較が行われるため, 優先順位が 2 番目の評価項目 (PFS) の比較が行われたペア数はスコアリング方法で異なり, それぞれ 9553 ペア (Gehan 流), 316 ペア (Péron 流), 66 ペア (補正版 Péron 流) であった.

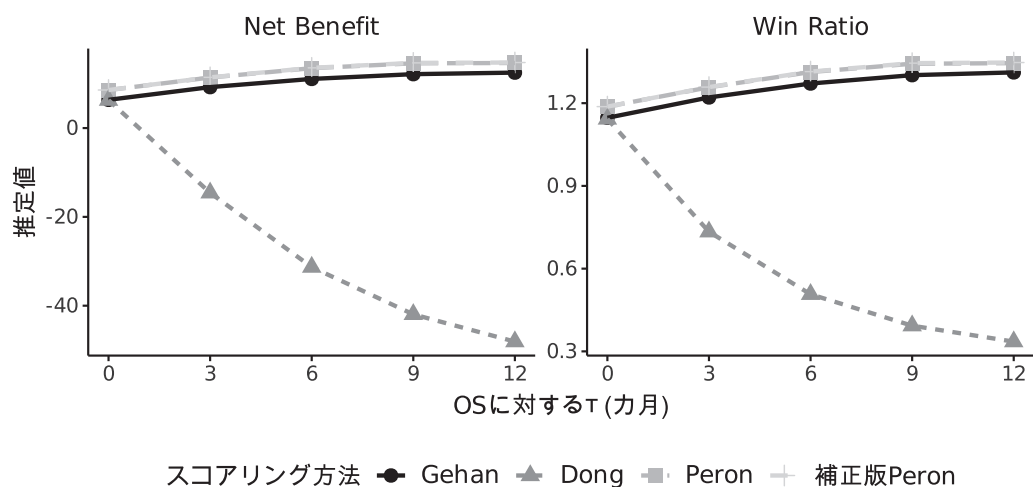


図2 V325 試験に対する一般化ペアワイズ比較の感度分析.

各スコアリング方法による Net Benefit の点推定値はそれぞれ 6.36% (Gehan 流), 6.28% (Dong 流), 8.56% (Péron 流), 8.57% (補正版 Péron 流), Win Ratio の点推定値はそれぞれ 1.15 (Gehan 流), 1.14 (Dong 流), 1.19 (Péron 流), 1.19 (補正版 Péron 流) であった. Gehan 流スコアと Dong 流スコアが近い結果であったことから, Dong 流スコアにおける重みづけの影響が小さかった可能性が示唆される. Péron 流スコアと補正版 Péron 流スコアの結果は非常に近かったが, これは Péron 流スコアにおける無情報ペア数が全ペア数の 0.5% と少なかったためである.

Net Benefit および Win Ratio に対する感度分析として, OS に対する τ の値を 0 から 12 カ月まで 3 か月ずつ変化させた場合の推定値の変化を図 2 に示した. 簡単のため, PFS に対する τ の値は 0 のままとした. Gehan 流, Péron 流, 補正版 Péron 流のスコアにおいては, τ の値が大きくなるにつれて, Net Benefit および Win Ratio の推定値は増加した. 一方で, Dong 流スコアに基づく推定値は減少し, 他の方法との結果の乖離が大きくなった. この原因としては, 2.3.2 節で述べた τ が 0 でないときに必要となる Dong 流スコアの重みの修正が WINS パッケージ (Ver 1.3.2) において実装されていないことが考えられる.

4.2 循環器臨床試験への適用

HF-ACTION 試験は, 慢性心不全患者に対する運動療法 (新治療群) の有用性を通常治療 (標準治療群) と比較したランダム化比較試験である (O'Connor *et al.* (2009)). 主要評価項目は, 理由を問わない全死亡と入院の複合イベントと設定された. ここでは, 「全死亡までの時間」, 「初発入院までの時間」という優先順位を想定し, それぞれを評価項目として用いた治療の評価を行う.

表 5 および図 3 に全死亡と入院に対する記述的な解析の結果を示す. 本試験では全死亡

表5 HF-ACTION 試験の生存時間データの要約.

		新治療群, N = 205	標準治療群, N = 221
追跡期間 (ヵ月)	中央値 (範囲)	28 (19, 40)	29 (18, 39)
全死亡	イベント数 (%)	36 (18%)	57 (26%)
	中央値 (ヵ月)	未到達	未到達
	ハザード比 (95%CI)	0.67 (0.44, 1.02)	
初発入院	イベント数 (%)	145 (71%)	170 (77%)
	中央値 (ヵ月) (95%CI)	15.3 (11.3, 17.8)	11.4 (8.5, 14.4)
	ハザード比 (95%CI)	0.83 (0.66, 1.03)	
複合イベント	イベント数 (%)	151 (74%)	175 (79%)
	全死亡 (%)	6 (2.9%)	5 (2.3%)
	初発入院 (%)	145 (71%)	170 (77%)
	打ち切り数 (%)	54 (26%)	46 (21%)
	中央値 (ヵ月) (95%CI)	15.1 (10.4, 17.6)	10.5 (8.3, 14.1)
	ハザード比 (95%CI)	0.84 (0.67, 1.04)	

Survival パッケージ(ver 3.2.13)を用い, 中央値は Kaplan-Meier 法, ハザード比は Cox 回帰により推定した.

イベントは新治療群で 18%, 標準治療群で 26%と少なく, 図 3 の Kaplan-Meier 推定値からも打ち切りが多数発生していることがわかる. 一方で, 初発入院までの時間に対しては, 両群とも 7 割以上の対象者でイベントが生じていた. Cox 回帰によるハザード比の推定値は, 全死亡までの時間では 0.67 (95%信頼区間; 0.44–1.02), 初発入院までの時間では 0.83 (95%信頼区間; 0.66–1.03) であった. 複合評価項目として, 全死亡または入院のいずれか最初のイベントまでの時間に対する解析を行った場合, いずれの群においても発生した複合イベントのほとんどは入院で規定されており, ハザード比の推定値は 0.84 (95%信頼区間; 0.67–1.04) と, 初発入院までの時間に対するハザード比と近い結果であった.

表 6 に Net Benefit および Win Ratio を推定した結果を示す. ここでは, いずれの評価項目に対しても $\tau = 0$ とした. 全死亡までの時間の比較では, Gehan 流スコアにおける無情報ペアが 69.1%, Péron 流スコアにおいても 48.6%あり, 多くのペアが無情報ペアと分類された. これは, 全死亡までの時間の打ち切り割合が 7 割以上と多かったことに起因して

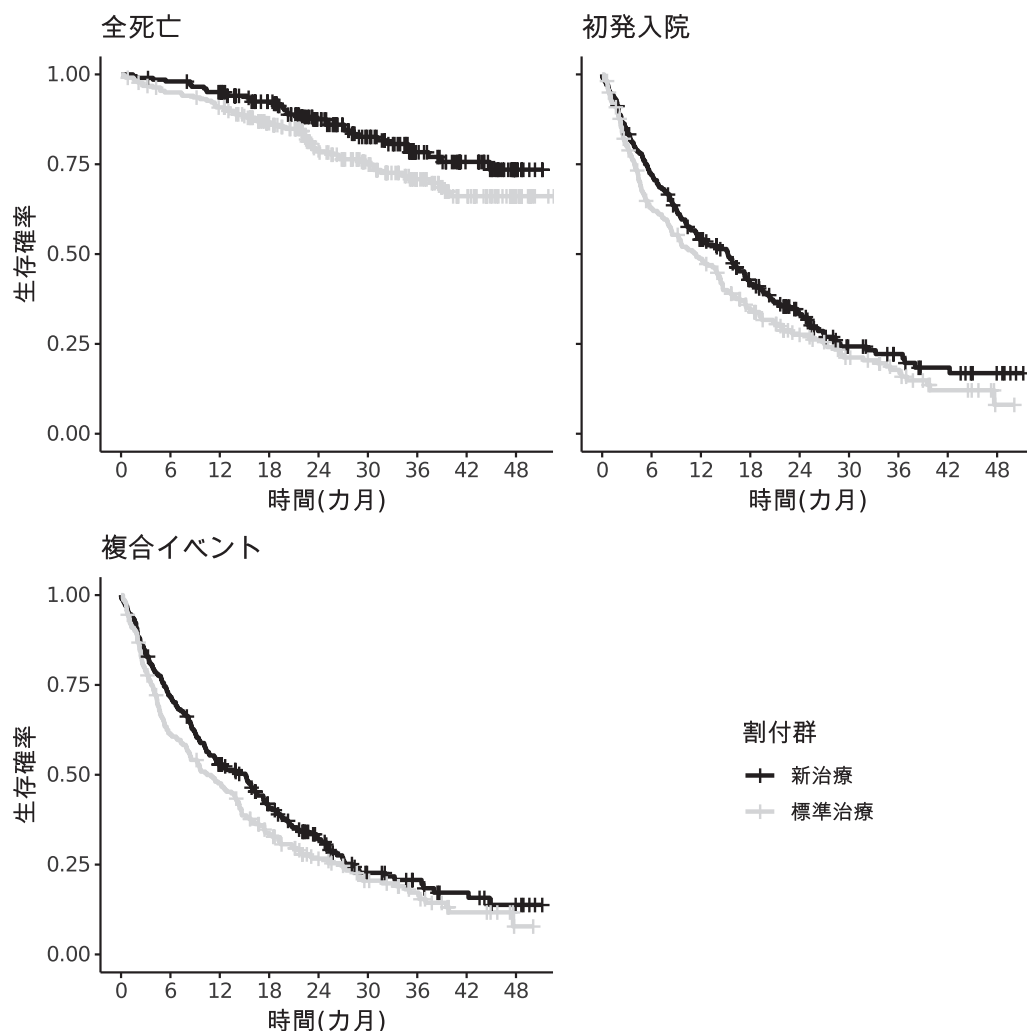


図3 HF-ACTION 試験の生存時間データに対する Kaplan-Meier 推定値.

いる。全死亡までの時間が等しいペアは存在しなかったため、 $\tau = 0$ とした場合、引き分けに分類されるペアは存在しなかった。したがって、Gehan 流スコアおよび P ron 流スコアでは、それぞれの方法で無情報ペアに分類された 31301 ペアおよび 22012 ペアが初発入院までの時間での比較に用いられた。一方で、補正版 P ron 流スコアにおいては、P ron 流スコアで無情報となった 22012 ペアのスコアが、勝ち(負け)にそれぞれ 58% (42%) 配分された。つまり、 $\tau = 0$ とした補正版 P ron 流スコアでは、全死亡までの時間の比較における引き分けと無情報のペアはなくなり、初発入院での比較は実施されなかったことになる。各スコアリングによる Net Benefit の推定値はそれぞれ 10.33% (Gehan 流), 11.20% (Dong 流), 11.39% (P ron 流), 16.98% (補正版 P ron 流), Win Ratio の推定値はそれぞれ 1.26 (Gehan 流), 1.26 (Dong 流), 1.26 (P ron 流), 1.41 (補正版 P ron 流)であった。

表 6 HF-ACTION 試験への一般化ペアワイズ比較の適用結果 (比較ペアの総数 45305).

スコアリング方法	評価項目	勝ち	負け	引き分け	無情報	Net Benefit (95% CI)	Win Ratio (95% CI)
Gehan	全死亡	18.9% (8576)	12% (5428)	0.0% (0)	69.1% (31301)		
	初発入院	30.6% (13865)	27.2% (12335)	0.1% (24)	11.2% (5077)		
	総合	49.5% (22441)	39.2% (17763)	0.1% (24)		10.33% (0%, 20.4%)	1.26 (1.00, 1.60)
Péron	全死亡	30.1% (13624)	21.3% (9669)	0.0% (0)	48.6% (22012)		
	初発入院	25.1% (11376)	22.4% (10172)	0.0% (19)	1.0% (445)		
	総合	55.2% (25000)	43.8% (19841)	0.0% (19)		11.39% (-1.3%, 23.7%)	1.26 (0.97, 1.63)
補正版 Péron	全死亡	58.5% (26499)	41.5% (18806)	0.0% (0)	0.0% (0)		
	初発入院	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)		
	総合	58.5% (26499)	41.5% (18806)	0.0% (0)		16.98% (-3.9%, 37.5%)	1.41 (0.92, 2.20)
Dong	総合					11.20% (0.1%, 22.2%)	1.26 (1.00, 1.57)

* 推定値に対する 95%信頼区間は, Gehan および Péron に対しては, 3.2.1 節および 3.2.3 節の Ozenne らによる漸近分散の推定量を, 補正版 Péron に対してはブートストラップ法に基づく方法を, Dong に対しては 3.2.2 節の Dong らによる漸近分散の推定量をそれぞれ用いた. Dong 流スコアは *WINS* パッケージ, それ以外の方法は *BuyseTest* パッケージにより計算した.

補正版 Péron 流スコアでは初発入院までの時間の情報が実質的に用いられなかったことが値の大きさに影響したと考えられる.

感度分析として, 全死亡までの時間に対する τ の値を 0 から 30 カ月まで 6 か月ずつ変化させた場合における推定値の変化を図 4 に示した. 簡単のため, 初発入院までの時間に対する τ の値は 0 のままとした. V325 試験とは異なり, Gehan 流, Péron 流のスコアリング方法では, τ の値が大きくなるにつれて Net Benefit および Win Ratio の推定値は減少した. 補正版 Péron 流スコアに基づく推定値は τ に対して非単調な動きを示した. Dong 流スコアに関しては, 4.1 節で述べた問題が生じており, 解釈はできない.

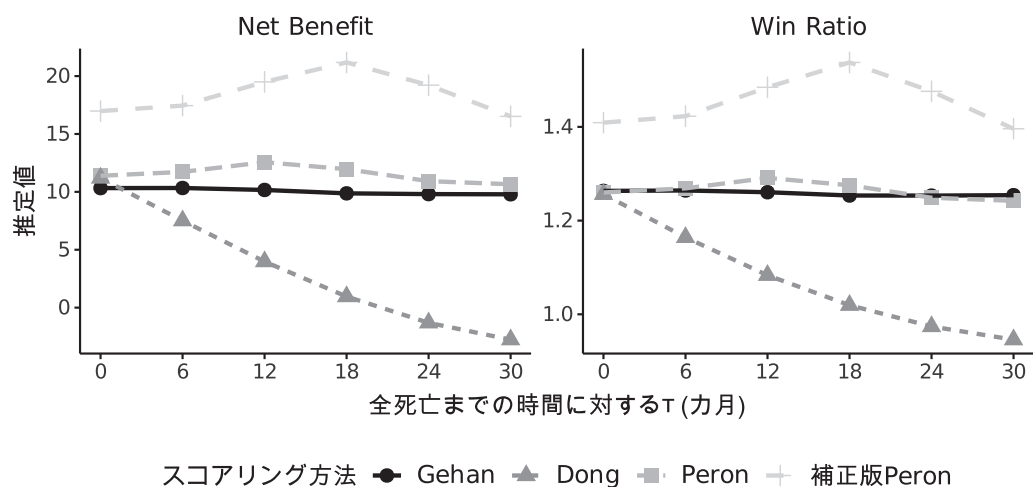


図4 HF-ACTION 試験に対する一般化ペアワイズ比較の感度分析.

5. 考察

本稿では、一般化ペアワイズ比較に基づく生存時間解析として、打ち切りを含むデータに対して提案されているスコアリング方法に焦点を当てながら、Net Benefit や Win Ratio に関する推定方法をレビューした。一般化ペアワイズ比較という観点から統一的に整理できる一方、評価項目が打ち切りの存在する生存時間である場合、スコアリング方法によって、同じ効果指標 (Net Benefit もしくは Win Ratio) であっても一部結果が大きく異なることが、実データ解析からも明らかとなった。

Gehan 流のスコアリングによる推定量は、その推定対象が打ち切り分布に依存したものになることが知られている (Buyse (2010), Péron *et al.* (2018, 2021), Dong *et al.* (2020c)). 複数評価項目に対する Net Benefit や Win Ratio の推定に Gehan 流スコアを用いる場合、上位の評価項目の打ち切り割合により、下位の評価項目による推定量への寄与度合いが変わるという特性を持つ (Péron *et al.* (2018)). この性質により、打ち切り割合が大きく異なる試験から得られた Net Benefit や Win Ratio の推定値を互いに参照して解釈することが難しいという指摘がある (Rauch *et al.* (2014)). 一般に、打ち切り割合は当該試験の観察期間によるため、Win Ratio の推定値を観察期間の関数として図示することを Finkelstein and Schoenfeld は提案している (Finkelstein and Schoenfeld (2019)). なお、4 節で紹介した *BuyseTest* パッケージおよび *WINS* パッケージでは観察期間を考慮した Net Benefit および Win Ratio の推定が可能である。Péron 流スコアと Dong 流スコアは、いずれも独立打ち切りの仮定の下、Kaplan-Meier 法に基づく生存時間あるいは打ち切り時間の推定量を利用した方法である (Péron *et al.* (2018, 2021), Dong *et al.* (2020c, 2021)). Péron 流スコアは、Gehan 流スコアでの無情報ペアに対して調整をする (生存確率の推定値を利用し 0 から

1の間の値で補完を行う)方法である。対照的に、Dong 流スコアは、Gehan 流スコアでの有情報ペアに対して調整をする(重み付けにより当該ペアのスコアを増幅させる)方法である。いずれの方法においても、生存時間あるいは打ち切り時間に対する Kaplan-Meier 法による推定が妥当に行える範囲、すなわち、当該試験の最大観察時間までの範囲のみで調整が可能となる (Péron *et al.* (2018, 2021), Dong *et al.* (2020c, 2021)). したがって、 τ を考慮した Net Benefit および Win Ratio の推定を Péron 流スコアにより行う場合、Kaplan-Meier 法による生存確率が推定不能であることに起因する無情報ペアが、 $\tau = 0$ の場合よりも相対的に増加する可能性が高くなる (Péron *et al.* (2018, 2021)). 補正版 Péron 流スコアはこの問題に対する一つの対処法として提案されているが、無情報ペアに対する強い仮定に基づく方法であることに留意すべきである (Péron *et al.* (2021)). Dong 流スコアで τ を考慮する場合には、IPCW 法における重みを適切に修正する必要があるが、Dong 流スコアを実装している WINS パッケージでは、そのような修正は本稿執筆時 (2022/10/18) における最新のバージョン (Ver 1.3.2) においても実装されていないため、解析に使用する際は十分注意が必要である。

Net Benefit および Win Ratio 統計量に対する漸近分散の推定量は、3節で紹介した方法の他にもいくつか提案されており、Verbeeck らによってシミュレーション実験に基づく性能評価とともにまとめられている (Verbeeck *et al.* (2020)). 本稿では用いなかったが、Bebu and Lachin による推定量は、R パッケージ *WinRatio* で実装されている方法であり、打ち切りのない場合 (2.2 節) や Gehan 流スコア (2.3.1 節) に基づく Net Benefit や Win Ratio 統計量の漸近分散の推定量として提案されている (Bebu and Lachin (2016)). この方法は、3.1.2 節の Dong らによる漸近分散の推定量と同じく、 $\Pr[\mathbf{X} \succ \mathbf{Y}, \mathbf{X} \succ \mathbf{Y}']$ の推定において \mathbf{Y} と \mathbf{Y}' が同じ対象者であることを許容しない方法である。一方で、Dong らの方法とは異なり、 $\Pr[\mathbf{X} \succ \mathbf{Y}]$ の推定量には対立仮説の下での推定量を用いており、この点に関しては 3.1.1 節の Ozenne らによる漸近分散の推定量と共通である。しかしながら、Bebu and Lachin による方法は、分散の推定値が負になりうることが数値実験により示されているなど、統計的な性質に問題がある推定量と考えられる (Verbeeck *et al.* (2020), Ozenne *et al.* (2021)). なお、2.3 節の Dong 流スコアや Péron 流スコアのように、打ち切りに対する補正を行った統計量に対する漸近分散の推定量は少なく、3.2 節で紹介した方法に限られる。

生存時間解析の文脈における、既存の一般化ペアワイズ比較に関する研究の課題と今後の展望をいくつか述べる。まず、本稿で紹介した Péron 流スコア (補正あり, なし) と Dong 流スコアに基づく推定量およびそれらの漸近分散の推定量に対する統計的性能を網羅的に評価した研究は未だ存在しない。特に Net Benefit や Win Ratio の主たる応用場面である複数の評価項目による治療の評価においては、異なる方法間での性能比較はまだ十分

になされていないため、今後の研究が望まれる。次に、既存の推定方法は主に独立打ち切りの仮定に基づくものが多い (Péron *et al.* (2018, 2021), Dong *et al.* (2020c)). Dong 流スコアでは、共変量を考慮した Cox 回帰による重みを用いることにより、打ち切りに対する仮定を条件付き独立に緩める提案がなされている (Dong *et al.* (2021)). 一方で、Péron 流スコアに対しては未だそのような拡張は提案されていない。今後の課題として、例えば共変量を考慮した調整済み Kaplan-Meier 推定量を Péron 流スコアの計算に取り入れることなどの可能性が考えられる。また、2.4 節で述べたように、Net Benefit や Win Ratio はスコアに基づく方法以外に生存関数による推定方法が考えられる。しかしながら、生存関数による推定方法に関する既存の研究は多くなく、複数の評価項目を用いる場合に利用できる推定量は Finkelstein and Schoenfeld による提案に限られる (Finkelstein and Schoenfeld (2019)). 生存関数による推定量および信頼区間・検定の構築方法や、スコアに基づく推定量との性能比較などはさらなる研究の余地があると考えられる。

本稿では十分に紹介できなかった関連する研究としては、競合リスクがある場合における Aalen-Johansen 推定量を用いた Péron 流スコアに基づく Net Benefit の推定量の提案 (Cantagallo *et al.* (2021)), 再発事象データに対する Win Ratio の推定 (Mao *et al.* (2022)), 複数評価項目を統合する際の重みの拡張 (Luo *et al.* (2017), Yang and Troendle (2021), Yang *et al.* (2022)), Win Ratio の検定に対する例数設計の方法論 (Mao *et al.* (2021)), Win Ratio に対する回帰分析 (Mao and Wang (2021)), クラスタランダム化試験への適用 (Zhang and Jeong (2021)), 等が挙げられる。今後、一般化ペアワイズ比較に基づく Net Benefit や Win Ratio 等の統計量に関する方法および適用事例の研究がさらに発展していくことが望まれる。

謝辞

本特集の企画者である江村剛志先生には、貴重な執筆の機会を頂きましたことに心から感謝申し上げます。また、3名の匿名査読者から有益な助言を頂きましたことに感謝申し上げます。本稿の作成にあたり有益なコメントを頂いた田栗正隆先生に感謝申し上げます。本研究は JSPS21H03401 の助成を受けたものです。

付録 A L 個の評価項目に対する勝ち・負けの統合スコアの計算

初めに L 個の評価項目の場合における Net Benefit と Win Ratio の定義を与える。 $\mathbf{X} = (X^{(1)}, \dots, X^{(L)})$, $\mathbf{Y} = (Y^{(1)}, \dots, Y^{(L)})$ に対する新治療群の「勝ち」と「負け」をそれぞれ以下のように定義する。

$$\mathbf{X} \succ \mathbf{Y} \equiv (X^{(1)} > Y^{(1)} + \tau^{(1)}) \cup \bigcup_{l=2}^L \left\{ (X^{(l)} > Y^{(l)} + \tau^{(l)}) \cap \bigcap_{k=1}^{l-1} (|X^{(k)} - Y^{(k)}| \leq \tau^{(k)}) \right\}$$

$$\mathbf{X} \prec \mathbf{Y} \equiv (X^{(1)} < Y^{(1)} - \tau^{(1)}) \cup \bigcup_{l=2}^L \left\{ (X^{(l)} < Y^{(l)} - \tau^{(l)}) \cap \bigcap_{k=1}^{l-1} (|X^{(k)} - Y^{(k)}| \leq \tau^{(k)}) \right\}$$

したがって、新治療群が勝っている確率 $\Pr[\mathbf{X} \succ \mathbf{Y}]$ と負けている確率 $\Pr[\mathbf{X} \prec \mathbf{Y}]$ は以下で与えられる。

$$\begin{aligned} \Pr[\mathbf{X} \succ \mathbf{Y}] &= \Pr[X^{(1)} > Y^{(1)} + \tau^{(1)}] \\ &\quad + \sum_{l=2}^L \Pr \left[(X^{(l)} > Y^{(l)} + \tau^{(l)}) \cap \bigcap_{k=1}^{l-1} (|X^{(k)} - Y^{(k)}| \leq \tau^{(k)}) \right] \\ \Pr[\mathbf{X} \prec \mathbf{Y}] &= \Pr[X^{(1)} < Y^{(1)} - \tau^{(1)}] \\ &\quad + \sum_{l=2}^L \Pr \left[(X^{(l)} < Y^{(l)} - \tau^{(l)}) \cap \bigcap_{k=1}^{l-1} (|X^{(k)} - Y^{(k)}| \leq \tau^{(k)}) \right] \end{aligned}$$

2.1 節で述べたように、これらの確率の差として Net Benefit が、比として Win Ratio が定義される。

さらに、2.2 節で説明した、ペア (i, j) に対する二つの評価項目による総合スコア s_{ij}^+ および s_{ij}^- を、 L 個の評価項目の場合に拡張するには、以下のように計算すればよい。

$$s_{ij}^+ = \sum_{l=1}^L w_{ij}^{(l)} s_{ij}^{(l)+}, \quad s_{ij}^- = \sum_{l=1}^L w_{ij}^{(l)} s_{ij}^{(l)-}$$

ここで、 l 番目の評価項目に対する重み $w_{ij}^{(l)}$ は

$$w_{ij}^{(l)} = \begin{cases} 1 & (l = 1) \\ \prod_{k=1}^{l-1} s_{ij}^{(k)0} & (l = 2, \dots, L) \end{cases}$$

とし、2 番目以降の評価項目では、上位の項目全てが引き分けであった場合に当該項目が寄与することとする。評価項目が 1 つの場合においては、当該の評価項目における勝ち・負けのみを考えればよく、

$$s_{ij}^+ = I(x_i^{(1)} > y_j^{(1)} + \tau^{(1)}), \quad s_{ij}^- = I(x_i^{(1)} < y_j^{(1)} - \tau^{(1)})$$

とすればよい。計算された s_{ij}^+, s_{ij}^- を 2.2 節の (1)(2) 式に代入した推定量から、Net Benefit や Win Ratio を推定することができる。なお、 U 統計量理論による (1)(2) 式の推定量の妥当性は評価項目の数によらない (Bebu and Lachin (2016), Dong *et al.* (2016), Péron *et al.* (2021), Ozenne *et al.* (2021))。

付録 B \hat{U}^+, \hat{U}^- の漸近分散とその推定量に関する補足

3.1 節の (9) 式における $\sigma_{\hat{U}^-, \hat{U}^-}$ と $\sigma_{\hat{U}^+, \hat{U}^-}$ は以下のように表される。

$$\sigma_{\hat{U}^-, \hat{U}^-} = \frac{\sigma_{-, -'}}{m} + \frac{\sigma_{-, -}}{n} \quad (14)$$

ただし,

$$\begin{aligned}
 \sigma_{-, -'} &= \text{Cov}[I(\mathbf{X} \prec \mathbf{Y}), I(\mathbf{X} \prec \mathbf{Y}')] \\
 &= E[I(\mathbf{X} \prec \mathbf{Y})I(\mathbf{X} \prec \mathbf{Y}')] - E[I(\mathbf{X} \prec \mathbf{Y})]E[I(\mathbf{X} \prec \mathbf{Y}')] \\
 &= \Pr[\mathbf{X} \prec \mathbf{Y}, \mathbf{X} \prec \mathbf{Y}'] - (\Pr[\mathbf{X} \prec \mathbf{Y}])^2 \\
 \sigma_{-', -} &= \text{Cov}[I(\mathbf{X} \prec \mathbf{Y}), I(\mathbf{X}' \prec \mathbf{Y})] \\
 &= E[I(\mathbf{X} \prec \mathbf{Y})I(\mathbf{X}' \prec \mathbf{Y})] - E[I(\mathbf{X} \prec \mathbf{Y})]E[I(\mathbf{X}' \prec \mathbf{Y})] \\
 &= \Pr[\mathbf{X} \prec \mathbf{Y}, \mathbf{X}' \prec \mathbf{Y}] - (\Pr[\mathbf{X} \prec \mathbf{Y}])^2
 \end{aligned}$$

また,

$$\sigma_{\hat{U}^+, \hat{U}^-} = \frac{\sigma_{+, -'}}{m} + \frac{\sigma_{+', -}}{n} \quad (15)$$

ただし,

$$\begin{aligned}
 \sigma_{+, -'} &= \text{Cov}[I(\mathbf{X} \succ \mathbf{Y}), I(\mathbf{X} \prec \mathbf{Y}')] \\
 &= E[I(\mathbf{X} \succ \mathbf{Y})I(\mathbf{X} \prec \mathbf{Y}')] - E[I(\mathbf{X} \succ \mathbf{Y})]E[I(\mathbf{X} \prec \mathbf{Y}')] \\
 &= \Pr[\mathbf{X} \succ \mathbf{Y}, \mathbf{X} \prec \mathbf{Y}'] - \Pr[\mathbf{X} \succ \mathbf{Y}] \Pr[\mathbf{X} \prec \mathbf{Y}] \\
 \sigma_{+', -} &= \text{Cov}[I(\mathbf{X} \succ \mathbf{Y}), I(\mathbf{X}' \prec \mathbf{Y})] \\
 &= E[I(\mathbf{X} \succ \mathbf{Y})I(\mathbf{X}' \prec \mathbf{Y})] - E[I(\mathbf{X} \succ \mathbf{Y})]E[I(\mathbf{X}' \prec \mathbf{Y})] \\
 &= \Pr[\mathbf{X} \succ \mathbf{Y}, \mathbf{X}' \prec \mathbf{Y}] - \Pr[\mathbf{X} \succ \mathbf{Y}] \Pr[\mathbf{X}' \prec \mathbf{Y}]
 \end{aligned}$$

である.

Ozenne らによる推定量

Ozenne らによる $\sigma_{\hat{U}^-, \hat{U}^-}$ の推定量は, 以下の $\sigma_{-, -'}$ と $\sigma_{-', -}$ に対する推定量を (14) 式に代入することで得られる (Ozenne *et al.* (2021)).

$$\begin{aligned}
 \hat{\sigma}_{-, -'} &= \frac{1}{mn^2} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1}^n s_{ij}^- s_{ij'}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^- \right)^2 \\
 \hat{\sigma}_{-', -} &= \frac{1}{m^2 n} \sum_{i=1}^m \sum_{i'=1}^m \sum_{j=1}^n s_{ij}^- s_{i'j}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^- \right)^2
 \end{aligned}$$

また, $\sigma_{\hat{U}^+, \hat{U}^-}$ の推定量は, 以下の推定量を (15) 式に代入すればよい.

$$\begin{aligned}
 \hat{\sigma}_{+, -'} &= \frac{1}{mn^2} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1}^n s_{ij}^+ s_{ij'}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^+ \right) \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^- \right) \\
 \hat{\sigma}_{+', -} &= \frac{1}{m^2 n} \sum_{i=1}^m \sum_{i'=1}^m \sum_{j=1}^n s_{ij}^+ s_{i'j}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^+ \right) \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n s_{ij}^- \right)
 \end{aligned}$$

Dong らによる推定量

Dong らによる $\sigma_{\hat{U}^-, \hat{U}^-}$ および $\sigma_{\hat{U}^+, \hat{U}^-}$ の推定量は、以下の $\sigma_{-, -'}$ 等の推定量をそれぞれ (14)(15) 式に代入することで求められる (Dong *et al.* (2016)).

$$\begin{aligned}\hat{\sigma}_{-, -'} &= \frac{1}{mn(n-1)} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1, j' \neq j}^n s_{ij}^- s_{ij'}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2 \\ \hat{\sigma}_{-, -'} &= \frac{1}{m(m-1)n} \sum_{i=1}^m \sum_{i'=1, i' \neq i}^m \sum_{j=1}^n s_{ij}^- s_{i'j}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2 \\ \hat{\sigma}_{+, -'} &= \frac{1}{mn(n-1)} \sum_{i=1}^m \sum_{j=1}^n \sum_{j'=1, j' \neq j}^n s_{ij}^+ s_{ij'}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2 \\ \hat{\sigma}_{+, -'} &= \frac{1}{m(m-1)n} \sum_{i=1}^m \sum_{i'=1, i' \neq i}^m \sum_{j=1}^n s_{ij}^+ s_{i'j}^- - \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \frac{s_{ij}^+ + s_{ij}^-}{2} \right)^2\end{aligned}$$

参 考 文 献

- Acion, L., Peterson, J. J., Temple, S. and Arndt, S. (2006). Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects, *Statistics in Medicine*, **25**, 591–602.
- Bebu, I. and Lachin, J. M. (2016). Large sample inference for a win ratio analysis of a composite outcome based on prioritized components, *Biostatistics*, **17**, 178–187.
- Buyse, M. (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem, *Statistics in Medicine*, **29**, 3245–3257.
- Buyse, M. and Péron, J. (2020). Generalized pairwise comparisons for prioritized outcomes, *Principles and Practice of Clinical Trials*, Piantadosi, S. and Meinert, C. L. ed., Springer, 1–25.
- Buyse, M., Saad, E. D., Burzykowski, T. and Péron, J. (2020). Assessing treatment benefit in immuno-oncology, *Statistics in Biosciences*, **12**, 83–103.
- Buyse, M., Saad, E. D., Péron, J., Chiem, J. C., De Backer, M., Cantagallo, E. and Ciani, O. (2021). The net benefit of a treatment should take the correlation between benefits and harms into account, *Journal of Clinical Epidemiology*, **137**, 148–158.
- Cantagallo, E., De Backer, M., Kicinski, M., Ozenne, B., Collette, L., Legrand, C., Buyse, M. and Péron, J. (2021). A new measure of treatment effect in clinical trials involving competing risks based on generalized pairwise comparisons, *Biometrical Journal*, **63**, 272–288.
- Chamseddine, A. N., Oba, K., Buyse, M., Boku, N., Bouché, O., Satar, T., Auperin, A. and Paoletti, X. (2021). Impact of follow-up on generalized pairwise comparisons for estimating the irinotecan benefit in advanced/metastatic gastric cancer, *Contemporary Clinical Trials*, **105**, 106400.
- Cui, Y. and Huang, B. (2022). WINS: The R WINS package, R package version 1.3.2.
- Dobler, D. and Pauly, M. (2018). Bootstrap- and permutation-based inference for the Mann–Whitney effect for right-censored and tied data, *Test*, **27**, 639–658.
- Dong, G., Hoaglin, D. C., Qiu, J., Matsouaka, R. A., Chang, Y. W., Wang, J. and Vandemeulebroecke, M. (2020a). The win ratio: on interpretation and handling of ties, *Statistics in Biopharmaceutical Research*, **12**, 99–106.
- Dong, G., Huang, B., Chang, Y. W., Seifu, Y., Song, J. and Hoaglin, D. C. (2020b). The win ratio: impact of censoring and follow-up time and use with nonproportional hazards, *Pharmaceutical Statistics*, **19**, 168–177.
- Dong, G., Huang, B., Wang, D., Verbeek, J., Wang, J. and Hoaglin, D. C. (2021). Adjusting win statistics for dependent censoring, *Pharmaceutical Statistics*, **20**, 440–450.

- Dong, G., Li, D., Ballerstedt, S. and Vandemeulebroecke, M. (2016). A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components, *Pharmaceutical Statistics*, **15**, 430–437.
- Dong, G., Mao, L., Huang, B., Gamalo-Siebers, M., Wang, J., Yu, G. and Hoaglin, D. C. (2020c). The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring, *Journal of Biopharmaceutical Statistics*, **30**, 882–899.
- Dong, G., Qiu, J., Wang, D. and Vandemeulebroecke, M. (2018). The stratified win ratio, *Journal of Biopharmaceutical Statistics*, **28**, 778–796.
- Efron, B. (1967). The two sample problem with censored data, *Proc Fifth Berkeley Symp Math Stat Probab*, **4**, 831–853.
- European Medicines Agency (2011). Benefit-risk methodology project. Work package 1 report: description of the current practice of benefit-risk assessment for centralised procedure products in the EU regulatory network, https://www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project-work-package-1-report-description-current-practice-benefit-risk_en.pdf.
- Fergusson, N. A., Ramsay, T., Chassé, M., English, S. W. and Knoll, G. A. (2018). The win ratio approach did not alter study conclusions and may mitigate concerns regarding unequal composite end points in kidney transplant trials, *Journal of Clinical Epidemiology*, **98**, 9–15.
- Ferreira, J. P., Jhund, P. S., Duarte, K., Claggett, B. L., Solomon, S. D., Pocock, S., Petrie, M. C., Zannad, F. and McMurray, J. J. V. (2020). Use of the win ratio in cardiovascular trials, *JACC. Heart failure*, **8**, 441–450.
- Finkelstein, D. M. and Schoenfeld, D. A. (1999). Combining mortality and longitudinal measures in clinical trials, *Statistics in Medicine*, **18**, 1341–1354.
- Finkelstein, D. M. and Schoenfeld, D. A. (2019). Graphing the win ratio and its components over time, *Statistics in Medicine*, **38**, 53–61.
- Food and Drug Administration (2021). Benefit-risk assessment for new drug and biological products, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/benefit-risk-assessment-new-drug-and-biological-products>.
- Giai, J., Maucourt-Boulch, D., Ozenne, B., Chiêm, J. C., Buyse, M. and Péron, J. (2021). Net benefit in the presence of correlated prioritized outcomes using generalized pairwise comparisons: a simulation study, *Statistics in Medicine*, **40**, 553–565.
- Kotalik, A., Eaton, A., Lian, Q., Serrano, C., Connett, J. and Neaton, J. D. (2019). A win ratio approach to the re-analysis of Multiple Risk Factor Intervention Trial, *Clinical Trials*, **16**, 626–634.
- Luo, X., Qiu, J., Bai, S. and Tian, H. (2017). Weighted win loss approach for analyzing prioritized outcomes, *Statistics in Medicine*, **36**, 2452–2465.
- Luo, X., Tian, H., Mohanty, S. and Tsai, W. Y. (2015). An alternative approach to confidence interval estimation for the win ratio statistic, *Biometrics*, **71**, 139–145.
- Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, **18**, 50–60.
- Mao, L., Kim, K. and Li, Y. (2022). On recurrent-event win ratio, *Statistical Methods in Medical Research*, **31**, 1120–1134.
- Mao, L., Kim, K. and Miao, X. (2021). Sample size formula for general win ratio analysis, *Biometrics*, <https://doi.org/10.1111/biom.13501>.
- Mao, L. and Wang, T. (2021). A class of proportional win-fractions regression models for composite outcomes, *Biometrics*, **77**, 1265–1275.
- Maurer, M. S., Schwartz, J. H., Gundapaneni, B., Elliott, P. M., Merlini, G., Waddington-Cruz, M., Kristen, A. V., Grogan, M., Witteles, R., Damy, T., Drachman, B. M., Shah, S. J., Hanna, M., Judge, D. P., Barsdorf, A. I., Huber, P., Patterson, T. A., Riley, S., Schumacher, J., Stewart, M., Sultan, M. B., Rapezzi, C. and ATTR-ACT Study Investigators (2018). Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy, *The New England Journal of Medicine*, **379**, 1007–1016.
- Oakes, D. (2016). On the win-ratio statistic in clinical trials with multiple types of event, *Biometrika*, **103**, 742–745.

- O'Connor, C. M., Whellan, D. J., Lee, K. L., Keteyian, S. J., Cooper, L. S., Ellis, S. J., Leifer, E. S., Kraus, W. E., Kitzman, D. W., Blumenthal, J. A., Rendall, D. S., Miller, N. H., Fleg, J. L., Schulman, K. A., McKelvie, R. S., Zannad, F., Piña, I. L. and HF-ACTION Investigators (2009). Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial, *JAMA: the journal of the American Medical Association*, **301**, 1439–1450.
- Ozenne, B., Budtz-Jørgensen, E. and Péron, J. (2021). The asymptotic distribution of the net benefit estimator in presence of right-censoring, *Statistical Methods in Medical Research*, **30**, 2399–2412.
- Ozenne, B. and Péron, J. (2022). BuyseTest: implementation of the generalized pairwise comparisons, R package version 2.3.11.
- Péron, J., Buyse, M., Ozenne, B., Roche, L. and Roy, P. (2018). An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring, *Statistical Methods in Medical Research*, **27**, 1230–1239.
- Péron, J., Giai, J., Maucourt-Boulch, D. and Buyse, M. (2019a). The benefit-risk balance of nab-paclitaxel in metastatic pancreatic adenocarcinoma, *Pancreas*, **48**, 275–280.
- Péron, J., Idlhaj, M., Maucourt-Boulch, D., Giai, J., Roy, P., Collette, L., Buyse, M. and Ozenne, B. (2021). Correcting the bias of the net benefit estimator due to right-censored observations, *Biometrical Journal*, **63**, 893–906.
- Péron, J., Lambert, A., Munier, S., Ozenne, B., Giai, J., Roy, P., Dalle, S., Machingura, A., Maucourt-Boulch, D. and Buyse, M. (2019b). Assessing long-term survival benefits of immune checkpoint inhibitors using the net survival benefit, *Journal of the National Cancer Institute*, **111**, 1186–1191.
- Péron, J., Roy, P., Conroy, T., Desseigne, F., Ychou, M., Gourgou-Bourgade, S., Stanbury, T., Roche, L., Ozenne, B. and Buyse, M. (2016a). An assessment of the benefit-risk balance of FOLFIRINOX in metastatic pancreatic adenocarcinoma, *Oncotarget*, **7**, 82953–82960.
- Péron, J., Roy, J., Ding, P., Parulekar, W. R., Roche, L. and Buyse, M. (2015). Assessing the benefit–risk of new treatments using generalised pairwise comparisons: the case of erlotinib in pancreatic cancer, *British Journal of Cancer*, **112**, 971–976.
- Péron, J., Roy, P., Ozenne, B., Roche, L. and Buyse, M. (2016b). The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials, *JAMA Oncology*, **2**, 901–905.
- Pocock, S. J., Ariti, C. A., Collier, T. J. and Wang, D. (2012). The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities, *European Heart Journal*, **33**, 176–182.
- Rauch, G., Jahn-Eimermacher, A., Brannath, W. and Kieser, M. (2014). Opportunities and challenges of combined effect measures based on prioritized outcomes, *Statistics in Medicine*, **33**, 1104–1120.
- Redfors, B., Gregson, J., Crowley, A., McAndrew, T., Ben-Yehuda, O., Stone, G. W. and Pocock, S. J. (2020). The win ratio approach for composite endpoints: practical guidance based on previous experience, *European Heart Journal*, **41**, 4391–4399.
- Rotolo, F., Paoletti, X. and Michiels, S. (2018). surroSurv: an R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials, *Computer Methods and Programs in Biomedicine*, **155**, 189–198.
- Saad, E. D., Zalberg, J. R., Peron, J., Coart, E., Burzykowski, T. and Buyse, M. (2018). Understanding and communicating measures of treatment effect on survival: can we do better?, *Journal of the National Cancer Institute*, **110**, 232–240.
- Tsuchikawa, M. and Sakamaki, K. (2022). Estimands for continuous longitudinal outcomes in the presence of treatment discontinuation—a simulation study in hyperkalemia treatments, *Statistics in Biopharmaceutical Research*, <https://doi.org/10.1080/19466315.2022.2050289>.
- Van Cutsem, E., Moiseyenko, V. M., Tjulandin, S., Majlis, A., Constenla, M., Boni, C., Rodrigues, A., Fodor, M., Chao, Y., Voznyi, E., Risse, M. L., Ajani, J. A. and V325 Study Group (2006). Phase III study of docetaxel and cisplatin plus fluorouracil compared with cisplatin and fluorouracil as first-line therapy for advanced gastric cancer: a report of the V325 Study Group, *Journal of Clinical Oncology*, **24**, 4991–4997.
- Verbeeck, J., Ozenne, B. and Anderson, W. N. (2020). Evaluation of inferential methods for the net benefit and win ratio statistics, *Journal of Biopharmaceutical Statistics*, **30**, 765–782.

- Verbeeck, J., Spitzer, E., de Vries, T., van Es, G. A., Anderson, W. N., Van Mieghem, N. M., Leon, M. B., Molenberghs, G. and Tijssen, J. (2019). Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints, *Statistics in Medicine*, **38**, 5641–5656.
- Voors, A. A., Angermann, C. E., Teerlink, J. R., Collins, S. P., Kosiborod, M., Biegus, J., Ferreira, J. P., Nassif, M. E., Psotka, M. A., Tromp, J., Borleffs, C., Ma, C., Comin-Colet, J., Fu, M., Janssens, S. P., Kiss, R. G., Mentz, R. J., Sakata, Y., Schirmer, H., Schou, M., Schulze, P. C., Spinarova, L., Volterrani, M., Wranicz, J. K., Zeymer, U., Zieroth, S., Brueckmann, M., Blatchford, J. P., Salsali, A. and Ponikowski, P. (2022). The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial, *Nature Medicine*, **28**, 568–574.
- Yang, S. and Troendle, J. (2021). Event-specific win ratios and testing with terminal and non-terminal events, *Clinical Trials*, **18**, 180–187.
- Yang, S., Troendle, J., Pak, D. and Leifer, E. (2022). Event-specific win ratios for inference with terminal and non-terminal events, *Statistics in Medicine*, **41**, 1225–1241.
- Zhang, D. and Jeong, J. H. (2021). Inference on win ratio for cluster-randomized semi-competing risk data, *Japanese Journal of Statistics and Data Science*, **4**, 1263–1292.