

【日本製薬工業協会シンポジウム】

生存時間解析の評価指標に関する最近の展開

—RMST (restricted mean survival time) を理解する—

6. RMSTの差/比とハザード比の経験的な比較 RMSTとハザード比に基づく検定の比較



2018年6月13日

医薬品評価委員会 データサイエンス部会 タスクフォース4
生存時間解析チーム

日本ベーリンガーインゲルハイム(株) 鵜飼 裕之

- 本発表は、先日公開された「生存時間型応答の評価指標 -RMST(restricted mean survival)を理解する-」について、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タスクフォース4 生存時間解析チームが本シンポジウムの趣旨を踏まえ、再構成したものである

➤ 本シンポジウム準備時点でハザード比と比較した結果がまとめられていた下記の論文について紹介

1. RMSTの差/比とハザード比の経験的な比較
 - Trinquart et al(2016)
2. RMSTとハザード比に基づく検定の比較
 - Tian et al(2017)

1 RMSTの差/比とハザード比の経験的な比較 ～Trinquart et al(2016)～

- **Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. J Clin Oncol. 2016; 34:1813–19**
- **がん臨床試験を対象**
- **ハザード比とRMSTの差及び比によって測定された治療効果を比較した**

- 対象: 5つの主要学術誌から2014年7月1日～12月31日に掲載されたがん臨床試験
- 抽出方法: 2人のreview authorが独立にタイトルとアブストラクトを精査

選択基準

- 優越性試験および非劣性試験のランダム化比較試験
- 主要または副次評価項目においてKaplan-Meier曲線の結果がある など

除外基準

- 第I相, 第II相および第IV相試験
- 多群, メタアナリシスや併合解析, クロスオーバー試験 など

- Digitizelt softwareを利用し、各患者の個人データを再構築
- ハザード比とRMSTの差(及び比)を算出し、治療効果の値を比較(図1-3)
- 境界時間 τ の設定
 - 各群で観察されるイベント発現までの最長期間のうち、短い方と定義
- ハザード比もRMSTの比も1より大きい場合、RMSTの差が0より大きい場合に実薬群が優れているとした
- 本論文の評価指標として:RMSTの比に対するハザード比の平均を利用(図4)
 - 1より大きい場合にハザード比のほうが治療効果が大きい

Trinquart et al (2016) の結果: 対象試験の背景情報

➤ 54のランダム化比較試験が選択された

特徴		結果
試験 デザイン	優越性	51 (94%)
	非劣性	3 (6%)
症例数	中央値 (最小, 最大)	503 (168, 2559)
評価項目 (抜粋)	OS	21 (39%)
	PFS	23 (43%)
雑誌	<i>New England Journal of Medicine</i>	11 (20%)
	<i>Lancet</i>	4 (7%)
	<i>Journal of the American Medical Association</i>	2 (4%)
	<i>Journal of Clinical Oncology</i>	22 (41%)
	<i>Lancet Oncology</i>	15 (28%)

Trinquart et al (2016) の結果: 対象試験の有効性の結果



- 24試験(44%)でlog-rank検定により有意差
 - 全て実薬群が優れている
- 13試験(24%)が比例ハザード性が成立していないと考えられる
 - Grambsch-Therneau test (有意水準10%)
- 図1: RMSTの差とハザード比のプロット
- 図2: RMSTの比とハザード比のプロット
- 図3: 54試験のRMSTの比とハザード比及び信頼区間
- 図4: RMSTの比に対するハザード比

図1: RMSTの差とハザード比のプロット

図2: RMSTの比とハザード比のプロット

Trinquart et al (2016)のFig1およびFig2を参照

図3: 54試験のRMSTの比とハザード比及び信頼区間

Trinquart et al (2016)のFig3を参照

- 1試験を除き, 統計学的有意性は一貫していた.
 - ✓ RMSTの差及び比では実薬群が統計学的に有意であったが, ハザード比では有意ではなかった.
- 41試験(76%):
 - $HR > \text{Ratio of RMST}$
- 13試験(24%):
 - $HR < \text{Ratio of RMST}$
 - 統計学的有意差を示した試験はない

図4: RMSTの比に対するハザード比

Trinquart et al (2016)のFig4を参照

- 「RMSTの比に対するハザード比」の比の平均値: 1.11 (95%CI: 1.07, 1.15)
 - ✓ RMSTのほうが保守的な推定値
 - ✓ OSでも同傾向
 - ✓ 比例ハザード性の成否にも依存していない結果
- 試験間の変動は大きかった($I^2=86\%$)
- 24%の試験で比例ハザード性が成立していないと考えられるため, RMSTの差も合わせて報告されるべき
- イベント数の少ない試験においてはハザード比の推定は不安定になる

Trinquart et al (2016)のまとめ

- ハザード比とRMSTの差(及び比)に基づく治療効果の指標は統計学的有意性という観点では一致
- ハザード比は平均的にRMSTの比よりも大きな治療効果を示す傾向にあった
- Time-to-eventのランダム化試験ではRMSTに基づく治療効果の推定結果も定期的に報告されるべき
- 本研究のLimitation
 - 再構築された患者の個人データに基づくため、オリジナルのデータからの一貫性を主張することは困難
 - しかしながら正確性や再現性に優れた手法を用い、再構築されたデータからの中央値やハザード比はオリジナルの論文に近い推定値
 - Kaplan-Meier曲線を積分し、RMSTを推定。境界時間 τ を超えたあたりでat riskが小さくなるため、Kaplan-Meier推定値が不安定になっている可能性
 - 公表された影響力の大きいがんランダム化比較試験を対象としているため、一般化可能性に欠けるかもしれない

2 RMSTとハザード比に基づく検定の比較 ～Tian et al(2017)～

- Tian L, Fu H, Ruberg SJ, Uno H, Wei LJ. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. Biometrics. 2017; doi:10.1111/biom.12770.
- Trinquart et al(2016)により経験的に示唆された結果を受けて、より理論的な状況でRMSTとハザード比の比較を検討
- 本論文内ではRMSTがt-year mean survival timeと表現されているが、RMSTと読み替えて記載する

- ハザード比に基づく検定とRMSTの差に基づく検定についての比較
- 評価指標
 - Asymptotical relative efficiency (ARE)
 - 漸近的な効率性をそれぞれの検定で導出し比をとったもの
 - 必要なサンプルサイズの比の逆数と解釈
 - 効果の差が大きい場合には精度が高い近似にならない
 - Empirical relative efficiency (ERE)
 - 効果の差が大きい場合に、効果の大きさの期待値をそれぞれの標準誤差で割って2乗した比
- Rajkumar et al. (2010)の試験データを参考に、シミュレーションを実施し、特性を確認

Rajkumar(2010)の試験データ

Tian et al (2017)のFig2-aを参照

使用されたシミュレーションデータのKaplan-Meier曲線例

Tian et al (2017)のFig2-b, Fig2-cを参照

- 打ち切り等の条件を変更し評価を行った

Tian et al(2017)の結果(1/2)

比例ハザード性が成り立つ状況

比例ハザード性が成り立つ状況下でのARE
およびEREのシミュレーション結果 (HR=0.7)

Censoring	ARE(ERE)	
	Light	Heavy
$S_0(t)=0.90$	0.95(0.94)	1.05(1.03)
$S_0(t)=0.80$	0.96(0.96)	1.05(1.03)
$S_0(t)=0.70$	0.97(0.93)	1.05(1.01)
$S_0(t)=0.60$	0.98(0.96)	1.06(1.03)
$S_0(t)=0.50$	0.99(0.96)	1.06(1.02)
$S_0(t)=0.40$	1.01(0.97)	1.06(1.02)
$S_0(t)=0.30$	1.02(0.97)	1.06(1.02)
$S_0(t)=0.20$	1.04(0.99)	1.05(1.02)
$S_0(t)=0.10$	1.05(1.01)	1.04(1.01)

- 比例ハザード性が成り立つ状況下において、ハザード比に基づく検定とRMSTの差に基づく検定のAREはおおよそ同様

ARE(ERE)>1の場合、RMSTのほうが効率性が高い
Light: 打ち切りが少ない場合
Heavy: 打ち切りが多い場合
(年間26%に相当)

EREは5000回のシミュレーションデータに基づく

Tian et al(2017)の結果(2/2)

比例ハザード性が成り立っていない状況

Censoring	$S_0(t)=0.30$		$S_0(t)=0.50$		$S_0(t)=0.70$	
	ARE(ERE)		ARE(ERE)		ARE(ERE)	
	Light	Heavy	Light	Heavy	Light	Heavy
$\alpha(u) = \log(0.46) + (u/t)^s$						
s=0.5	1.76(1.77)	1.16(1.13)	2.02(2.05)	1.24(1.22)	2.30(2.37)	1.34(1.31)
s=1.0	1.34(1.33)	1.11(1.08)	1.42(1.41)	1.15(1.13)	1.49(1.47)	1.20(1.17)
s=2.0	1.18(1.17)	1.09(1.07)	1.21(1.19)	1.11(1.10)	1.23(1.23)	1.13(1.12)
$\alpha(u) = -c(\alpha)(t-u)^s$						
s=0.5	1.10(1.05)	1.07(1.05)	1.10(1.06)	1.09(1.06)	1.10(1.07)	1.10(1.06)
s=1.0	1.20(1.17)	1.09(1.06)	1.22(1.19)	1.11(1.09)	1.24(1.21)	1.13(1.11)
s=2.0	1.37(1.34)	1.11(1.09)	1.43(1.38)	1.16(1.13)	1.48(1.45)	1.19(1.17)
$\alpha(u) = -c(\alpha)u^s$						
s=0.5	0.84(0.80)	1.02(0.99)	0.81(0.77)	1.00(0.96)	0.78(0.76)	0.98(0.94)
s=1.0	0.74(0.70)	1.00(0.96)	0.70(0.67)	0.96(0.92)	0.67(0.64)	0.92(0.90)
s=2.0	0.61(0.58)	0.96(0.93)	0.56(0.54)	0.89(0.86)	0.53(0.52)	0.84(0.82)

- 比例ハザード性が成り立っていない状況下では打ち切りが少なく、試験終盤近くの生存曲線の差が大きい場合 □を除き、RMSTの差に基づく検定はハザード比に基づく検定より効率がよかった

- RMSTの差に基づく検定は早期に効果が期待できる場合、ハザード比に基づく検定は遅発性の効果を検討する場合に検出力が高いと言えた
- 試験終盤近くに生存曲線の差が大きい場合にはハザード比に基づく検定のほうが、より検出力が高くなるかもしれない
 - しかしながら、上記の場合には比例ハザード性が成り立っていないと考えられるため、ハザード比の推定値自体の単純な臨床的な解釈が難しい可能性がある

- **Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. J Clin Oncol. 2016; 34:1813–19**
- **Tian L, Fu H, Ruberg SJ, Uno H, Wei LJ. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. Biometrics. 2017; doi:10.1111/biom.12770.**
- **Rajkumar SV, Jacobus S, Callander NS, Fonseca R, Vesole DH, Williams ME, *et al.* Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. Lancet Oncol. 2010; 11(1):29–37.**