Centre de Recherche Epidémiologie et Statistique Sorbonne Paris Cité
Inserm UMR 1153

ONCOSTAT Symposium
October 20th, 2016

Restricted mean survival *v* hazard ratios in oncology randomized controlled trials

Raphaël Porcher
Joint work with Ludovic Trinquart, Justine Jacot and Sarah Conner

Equipe METHODS, Centre de Recherche Epidémiologie et Statistique Sorbonne Paris Cité
INSERM U1153, Université Paris Descartes, AP-HP, France

UNIVERSITÉ PARIS DESCARTES
USPC Université Sorbonne Paris Cité
Instituts thématiques Inserm Institut national de la santé et de la recherche médicale
ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Background

- Survival outcomes are most common in cancer trials (OS, PFS, …)
- The hazard ratio (HR) is the preferred measure of the treatment effect on those outcomes
- HRs have been challenged as relevant measures of survival benefit
  - Relative rather than absolute measure of treatment effect
  - Not so easy to interpret for non-statisticians (≠ RR for instance)
  - May not be constant in time (non-proportional hazards)
- Often used in conjunction with e.g. the median to characterize absolute effects

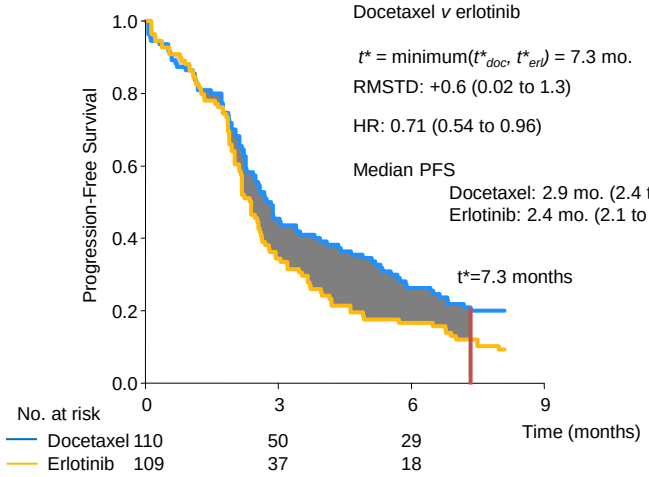
RMST

- Censoring usually precludes estimation of mean survival
- Restricted mean survival time (RMST) has been advocated as an alternative summary for survival curves
- Direct interpretation: $RMST_{t^*}$ is the life expectancy at horizon t*
- Treatment effect
  - Difference in RMST (RMSTD)
  - Ratio of RMST (RMSTR)
  - Difference (ratio) of restricted mean time loss (t* – $RMST_{t^*}$)
- Difference in means more relevant than difference in medians

Karrison *Controlled Clinical Trials* 2007
Royston & Parmar *Statistics in Medicine* 2011
Royston & Parmar *BMC Medical Research Methodology* 2013

Difference in RMST

2nd line *EGFR* wild-type NSCLC



Garassino et al. *Lancet Oncology* 2013

Objective

- To obtain empirical evidence on the comparison of treatment effects measured as HR or RMSTD/R in recent oncology randomized trials

Methods

- Systematic review of phase III parallel-group randomized trials in oncology published last 6 mo. of 2014 in selected high impact journals
- Reconstruction of individual patients data (IPD) from published survival curves
- For the primary outcome whenever possible
  - Priority given to PFS or EFS if co-primary outcomes
- Reanalysis of each trial → HR, RMSTD, RMSTR

Selection of Trials

- Oncology randomized controlled trials
- Published last 6 months of 2014 in

  New England Journal of Medicine
  Lancet
  Journal of the American Medical Association
  Lancet Oncology
  Journal of Clinical Oncology

- Search MEDLINE using Cochrane highly sensitive strategy
- Inclusion criteria
  - Superiority and noninferiority parallel group RCTs
  - Kaplan-Meier curve for a primary or secondary time-to-event outcome
- Exclusion criteria
  - Phase I, II, IV; > 2 arms; meta-analyses; reports of secondary, subgroup, or follow-up analyses;
  - Supportive care, palliative care, or prevention trials;
  - Specific designs (eg. Cluster RCT, cross-over …)

Reconstruction of IPD

- Extraction of times and survival probabilities on KM curves using DigitizeIt software
- Use of an iterative algorithm* to reconstruct individual patient data for each arms using
  - Reconstructed curves
  - Total number of events (when available)
  - Numbers of patients at risk (when available)

* Guyot et al. *BMC Medical Research Methodology* 2012

Estimation of Treatment Effects

- HR
  - Cox proportional hazards regression model
  - Grambsch-Therneau test for the non-proportional hazards
  - Log-rank test of null treatment effect
- RMST
  - t*: minimum of the largest observed event time in each group
  - RMST as area under the KM curve
  - Difference and ratio in RMST
  - Test for null treatment effect: compare RMSTD/SE to a standard normal distribution (same for RMSTR)

Comparison of Treatment Effects

- Comparison of HR and RMSTR (same scale)
- Assess whether a measure is systematically further from the null
- For each trial: ratio HR/RMSTR
- Variance of HR/RMSTR obtained by bootstrap
- Coding so that HR/RMSTR>1 → HR more optimistic than RMSTR
- Meta-analysis of the individual trial ratios
- Subgroup analyses
  - Type of outcome (OS *v* others)
  - (Non-)proportionality of hazards

Characteristics of the Selected RCTs

- 54 trials
- Selected outcome
  - PFS : 23 (43%)
  - OS : 21 (39%)
  - RFS, FFS, DFS, EFS : 10 (18%)

Characteristics of 54 Randomized Trials Included in Analysis

Feature	No. (%)*
Journal	
New England Journal of Medicine	11 (20)
Lancet	4 (7)
Journal of the American Medical Association	2 (4)
Journal of Clinical Oncology	22 (41)
Lancet Oncology	15 (28)
Funding source	
Industry	29 (54)
Nonindustry	17 (31)
Both	8 (15)
Cancer type	
Digestive/GI	11 (20)
Hematologic/blood	9 (17)
Breast	6 (11)
Endocrine and neuroendocrine	6 (11)
Respiratory/thoracic	6 (11)
Skin	6 (11)
Genitourinary	3 (6)
Gynecologic	2 (4)
Head and neck	2 (4)
Other	3 (6)
Type of experimental treatment	
Chemotherapy	8 (15)
Hormonal therapy	6 (11)
Targeted therapy	34 (63)
Chemotherapy plus targeted therapy	3 (6)
Immunotherapy	1 (2)
Radiation therapy	2 (4)
Type of control arm	
Placebo or best supportive care	36 (67)
Active comparator	18 (33)
Trial design	
Superiority	51 (94)
Noninferiority	3 (6)
Randomization ratio	
1:1	50 (93)
1:2	4 (7)
Sample size, median (Q1-Q3)	503 (296-764)
Trials stopped early	12 (22)†

*Except for the entry: sample size.
†Three trials were stopped early for efficacy, five trials for futility, and four trials for enrollment deficiencies.

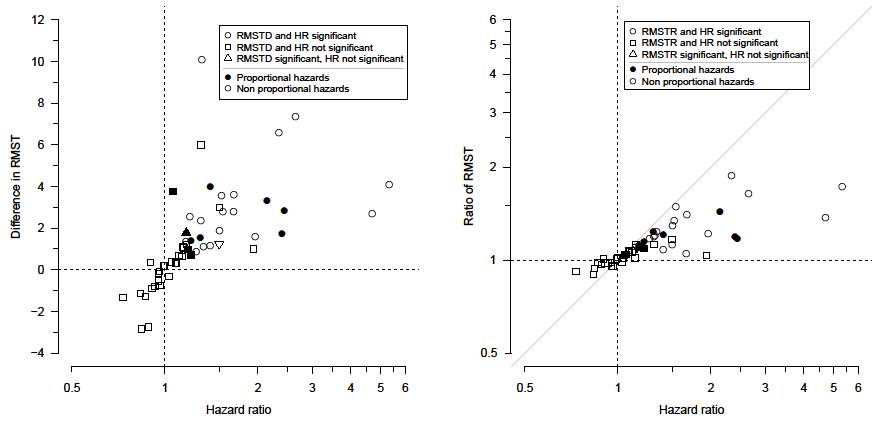
Treatment Effect Estimates: HR

- Median HR (Q_1 to Q_3): 0.84 (0.67 to 0.996)
  - Range from 0.19 to 1.37
- Proportional hazards
  - Reported as checked (and OK) in 5 (9%) trials
  - We found evidence of NPH in 13 (24%) other trials
- Significant difference in 24 (44%) trials (always in favor of experimental)

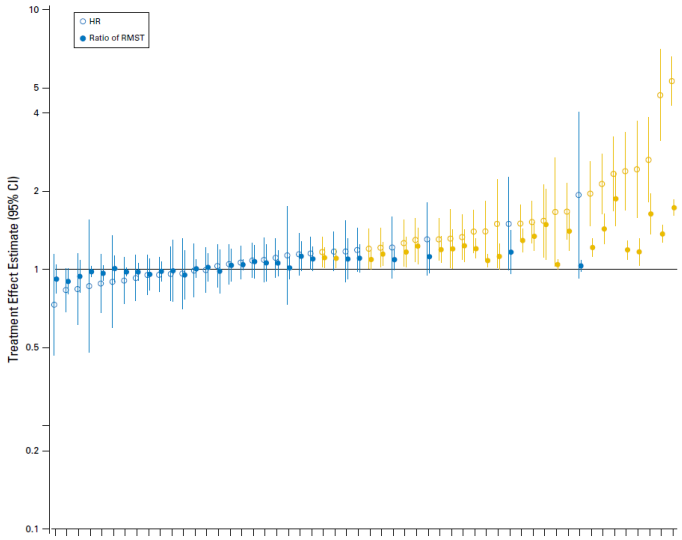
Treatment Effect Estimates: RMST

- Median t* 28.3 mo. (range 11 mo. to 10 y)
- Differences in RMST
  - Range from -2.9 to +10.1 mo.
  - Median (Q_1 to Q_3): +1.1 mo. (0.2 to 2.8)
- Ratios of RMST
  - Range from 0.90 to 1.87
  - Median (Q_1 to Q_3): 1.10 (1.01 to 1.20)
- Significant difference for 25 (46%) of trials

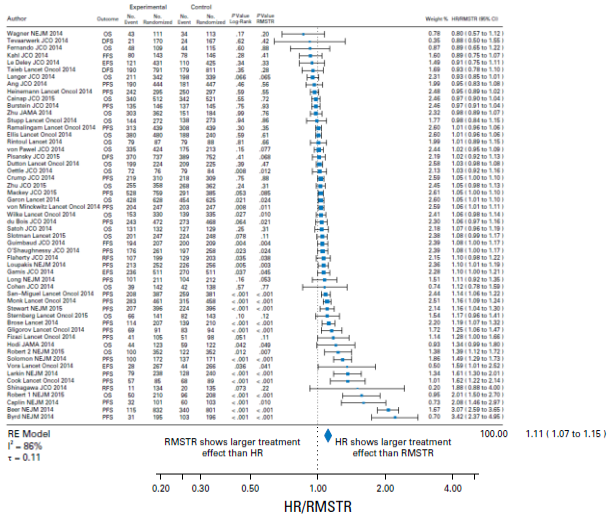
Comparison of Treatment Effects



HRs and RMSTRs with 95% CIs



Meta-analysis of HR/RMSTR



Subgroup analyses

- By type of outcome
  - 1.04 [1.00; 1.07] for OS (n=21 trials)
  - 1.17 [1.10; 1.23] for others (n=33 trials)

  P = .08
- By evidence of non-proportional hazards
  - 1.12 [1.06; 1.19] for NPH (n=13 trials)
  - 1.10 [1.06; 1.15] without NPH (n=41 trials)

  P = .55
- No difference according to sample-size or f-up duration either (post-hoc)

Discussion

- HR and RMST-based measures in agreement overall
- But RMST yielded more conservative estimates
  - For any time-to-event outcome
  - Whatever the evidence for NPH
- Example: dabrafenib + trametinib v vemurafenib in metastatic melanoma with a *BRAF* V600 mutation
  - HR 0.71 (0.55 to 0.93) i.e 1/HR 1.41 (1.08 to 1.82)
  - RMSTD 1.1 mo. (0.3 to 2.0) at t* = 18.1 mo.
  - RMSTR 1.08 (1.02 to 1.15)

RMST-based *v* other measures

- RMST-based measures do not rely on specific assumptions such as PH
- And are quite easily interpreted
- Differences in survival probabilities at a specific time point or in median survival times can also be used
  - Interpretable but summarize a curve by one single point
  - In addition the median survival time may be unestimable (16 [30%] in our sample)
- Limitation of RMST is the choice of t*
  - May be predefined in a protocol according to expected data maturity
  - Curves of RMSTD over time have also been proposed but to not provide a single estimate of the benefit

Limitations of our study

- We used reconstructed IPD
  - Less reliable than the true IPD
  - Unable to perform adjustment / stratification as in some original trials
  - Good reproducibility + agreement with published HRs anyway
- Choice of t*
  - Common strategy for all RCTs without further investigation
  - RMST obtained by KM-integration may become unstable at the right tail
  - Other approaches may be more robust in specific cases (e.g. using flexible models)
- Restricted to trials published in high-impact journals

Thank you for your attention!

JOURNAL OF CLINICAL ONCOLOGY | STATISTICS IN ONCOLOGY

Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials

Ludovic Trinquart, Justine Jacot, Sarah C. Conner, and Raphaël Porcher

Reliability of estimation after IPD reconstruction

- 6 randomly studies were extracted by 2 reviewers and survival probabilities were estimated at 5 time points in each group
  - Over the 60 points, le mean reproducibility error was 0.2% (0.01 to 0.4)
- Bland-Altman analysis for the 27 studies reporting non-adjusted HRs
  - Mean difference of logHR was -0.002 (-0.008 to 0.004)