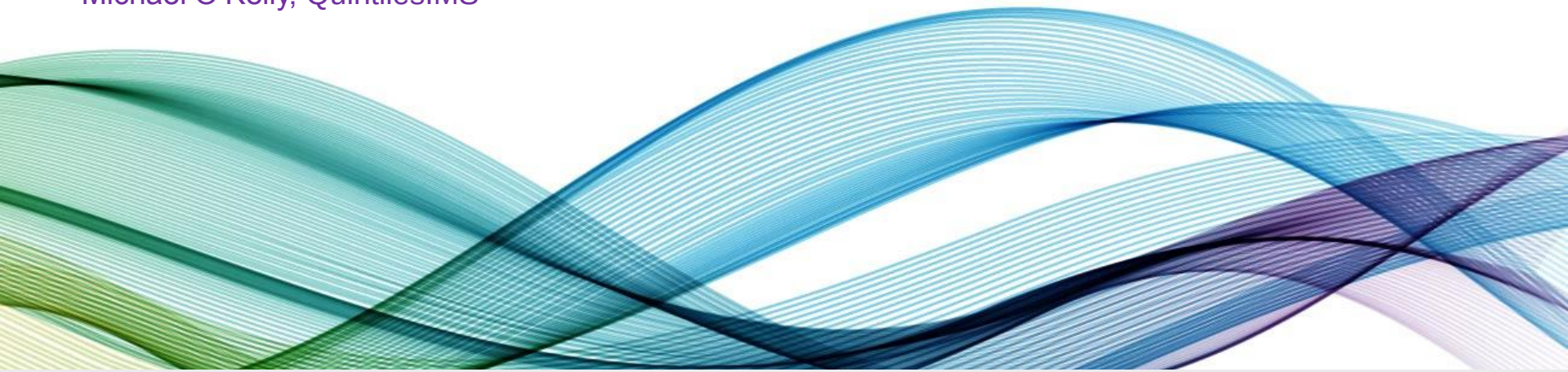
Analysis of Health-Related Quality of Life Outcomes in Cancer Trials when Data are Missing after Disease Progression or Treatment Discontinuation

2017 ASA Biopharmaceutical Section Statistics Workshop

Bohdana Ratitch, QuintilesIMS
Cristina Ivanescu, QuintilesIMS
Michael O'Kelly, QuintilesIMS



Outline

- Motivating example: typical historical designs of prostate cancer trials -

  HR-QoL data collection, endpoints, analyses … and estimands

- HR-QoL deterioration after disease progression and toxicity; imbalance in reasons for treatment discontinuation between treatments

- Accounting for likely HR-QoL deterioration after disease progression and toxicity when data are missing

Health-Related Quality of Life (HR-QoL) in Cancer Trials

- Prolongation of survival remains the basis for drug approval in advanced cancer
- HR-QoL outcomes are also important measures of cancer therapies
  - Considered by regulatory and Health Technology Assessment agencies
- Progressive disease (PD) and toxicity have an important impact on HR-QoL
- New treatments should delay PD and improve how patients feel and function

Motivating Example: HR-QoL in Prostate Cancer Trials

- Typical prostate cancer trial:
  - Study treatment administered until PD or toxicity
  - Post-treatment follow-up: mainly safety/survival follow-up
  - HR-QoL instruments often not collected after end-of-treatment (EOT) visit
  - Subsequent anti-cancer therapies may be started after end of study treatment

- PD and toxicity are treatment-related, intercurrent events
  - There is often a significant imbalance in the proportions and timing of these events between treatment groups
  - If not accounted for in the HR-QoL analyses, will result in confounding and bias

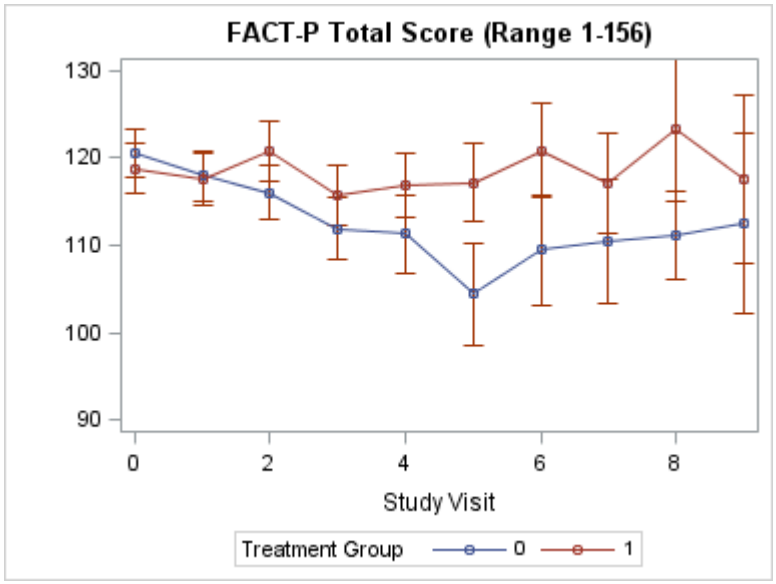
Motivating Example: HR-QoL in Prostate Cancer Trials

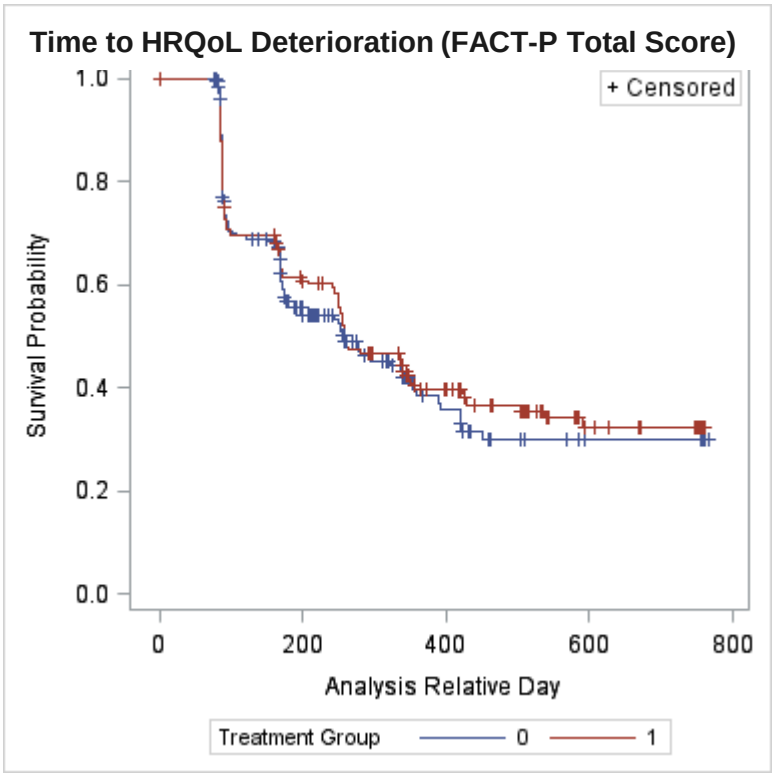
- Functional Assessment on Cancer Therapy-Prostate (FACT-P) – a validated questionnaire for metastatic castration-resistant prostate cancer
- Includes a general functional status scale (physical wellbeing, social and family wellbeing, emotional wellbeing, and functional wellbeing subscales) and a prostate-cancer-specific subscale
- Ranges from 0 to 156 with higher scores indicating better functional status
- HR-QoL deterioration is typically predefined on the basis of score changes from baseline judged clinically meaningful to patients, e.g., decrease of 10 points from baseline in total score

Objectives, Endpoints, Estimands, and Missingness

- Objective (typical protocol language):
  - To evaluate the benefit of experimental treatment A as compared to B on quality of life as assessed by FACT-P
- Endpoints
  - Change from baseline to Month X in FACT-P total score and other sub-scales
  - Time to HR-QoL deterioration, e.g., decrease of 10 points from baseline in FACT-P total score
- Estimand – rarely specified in the protocol or SAP … more on this in a moment
- Missingness – what is considered missing depends on what needs to be estimated…
  - HR-QoL after death?
  - HR-QoL after study treatment discontinuation:
    - Do we need those data?
    - Would it be confounded with the effect of subsequent anti-cancer therapies?

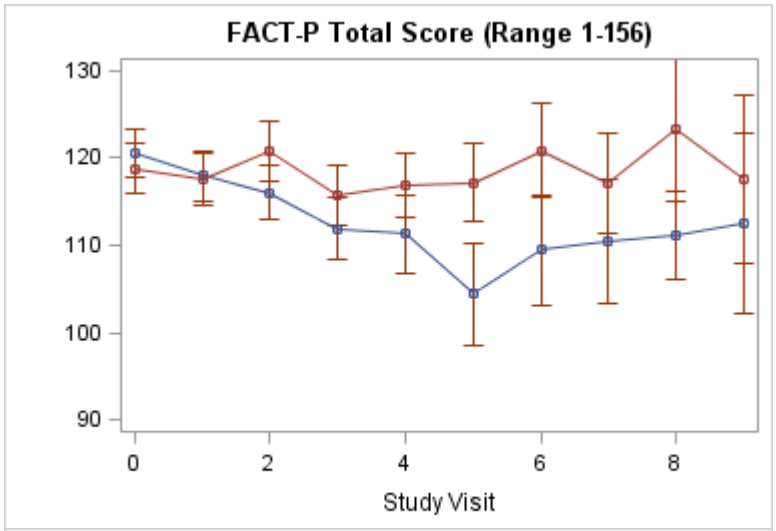
Motivating Example: Summary of On-Treatment Data

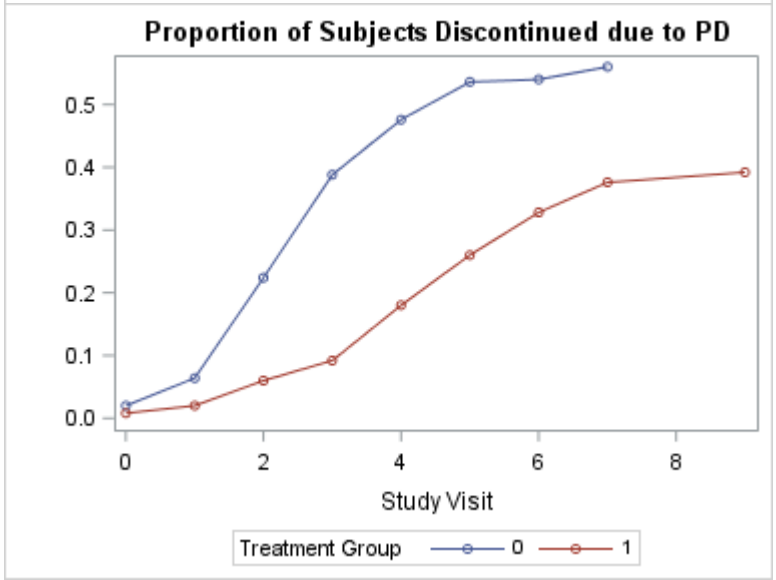


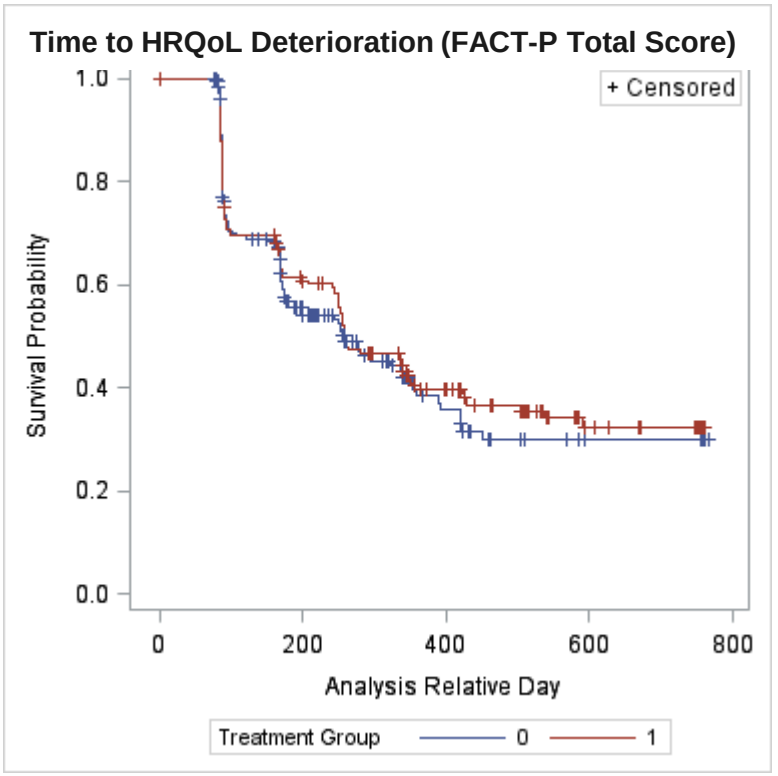


- Small differences between treatments based on observed data prior to treatment discontinuation.

Motivating Example: Summary of On-Treatment Data

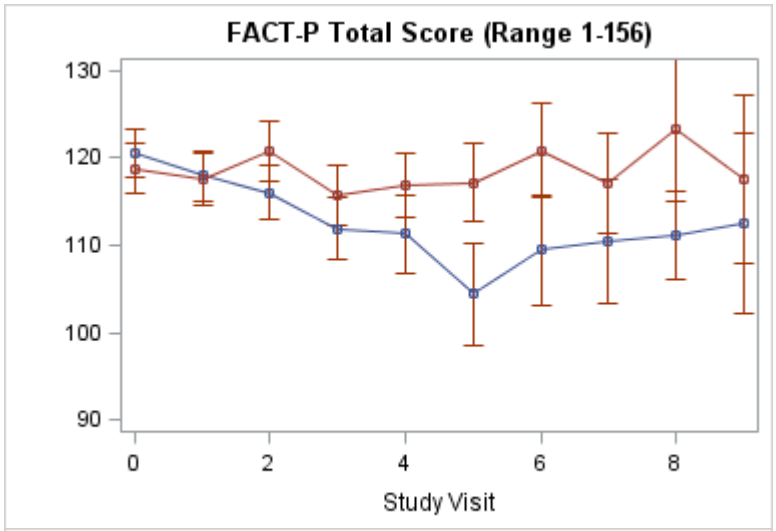


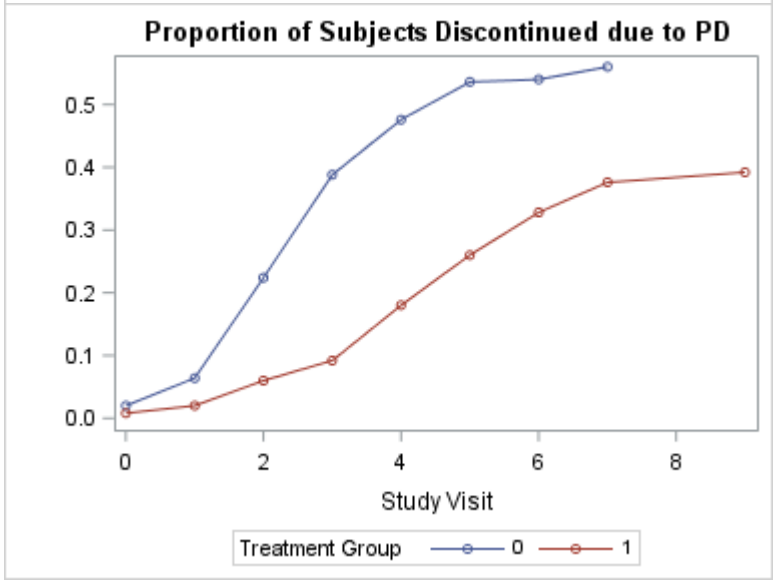


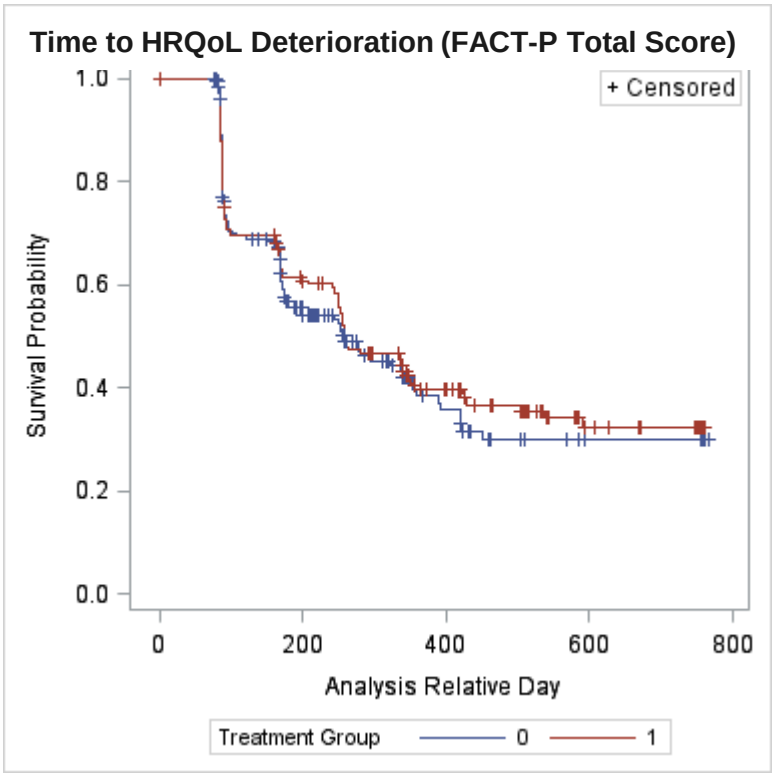


- Small differences between treatments based on observed data prior to treatment discontinuation.
- Significantly different proportions of discontinued subjects due to PD.

Motivating Example: Summary of On-Treatment Data







Can on-treatment data tell us the full story?

What about a likely HR-QoL deterioration after PD?

QuintilesIMS™

Typical Analyses and Their Assumptions

- Mixed Model with Repeated Measures (MMRM) using on-treatment data for change from baseline in HR-QoL scores
  - Least Squares Means at specific time points
  - Average treatment difference over time

  - Not all patients have data at all analysis time points – many have discontinued early.
  - Data after discontinuation are missing, and assumed by MMRM to be missing at random
  - Implication: missing HR-QoL scores are modeled based on observed scores within treatment group, after accounting for covariates and pre-discontinuation scores

Typical Analyses and Their Assumptions

- Mixed Model with Repeated Measures (MMRM) using on-treatment data for change from baseline in HR-QoL scores
  - Least Squares Means at specific time points
  - Average treatment difference over time

  - Not all patients have data at all analysis time points – many have discontinued early.
  - Data after discontinuation are missing, and assumed by MMRM to be missing at random
  - Implication: missing HR-QoL scores are modeled based on observed scores within treatment group, after accounting for covariates and pre-discontinuation scores

    In other words: HR-QoL of subjects who discontinued treatment (including for PD and toxicity) is modelled as similar to subjects who continue treatment (had no PD or toxicity)

Typical Analyses and Their Assumptions

- Mixed Model with Repeated Measures (MMRM) using on-treatment data for change from baseline in HR-QoL scores
  - Least Squares Means at specific time points
  - Average treatment difference over time

  - Not all patients have data at all analysis time points – many have discontinued early.
  - Data after discontinuation are missing, and assumed by MMRM to be missing at random
  - Implication: missing HR-QoL scores are modeled based on observed scores within treatment group, after accounting for covariates and pre-discontinuation scores

    In other words: HR-QoL of subjects who discontinued treatment (including for PD and toxicity) is modelled as similar to subjects who continue treatment (had no PD or toxicity)

- Kaplan-Meier / Cox proportional hazards regression for time to HR-QoL deterioration
  - Censored at random / “ignorable censoring” assumption
  - Implication: among those at risk at t, the event hazard is modeled similarly between those who are censored at t and those who are not, after accounting for covariates

Estimand – Reverse-Engineered from Typical Analysis

- Difference between treatments A and B in mean change from baseline in FACT-P total score at Week X due to study treatment if taken as directed through Week X in all randomized subjects

Estimand – Reverse-Engineered from Typical Analysis

- Difference between treatments A and B in mean change from baseline in FACT-P total score at Week X due to study treatment **if taken as directed through Week X** in all randomized subjects

  - To take treatment through Week X, subjects would need NOT to have PD or toxicity before Week X
  - For a large proportion of subjects, we know this is not the case

Estimand – Reverse-Engineered from Typical Analysis

- Hazard ratio of treatments A versus B for HR-QoL deterioration due to study treatment **during study treatment administration** in all randomized subjects

  - Does not take into account that HR-QoL can deteriorate significantly and rapidly after discontinuation due to PD and toxicity and that proportions of subjects with these events may differ between treatment groups
  - Kaplan-Meier / Cox regression analysis does not acknowledge that censoring is treatment–related and is likely not independent of unobserved event times

Alternative Estimand

- Hazard ratio of treatments A versus B for HR-QoL deterioration due to study treatment **regardless of treatment duration** in all randomized subjects

Alternative Estimand

- Hazard ratio of treatments A versus B for HR-QoL deterioration due to study treatment **regardless of treatment duration** in all randomized subjects

- To estimate this estimand, HR-QoL data post treatment discontinuation would be needed.

Alternative Estimand

- Hazard ratio of treatments A versus B for HR-QoL deterioration **due to study treatment** **regardless of treatment duration** in all randomized subjects

- To estimate this estimand, HR-QoL data post treatment discontinuation would be needed.

Note:

- This estimand targets treatment effect solely "due to study treatment" – not confounded with the effect of subsequent anti-cancer therapies
- Such data may be challenging to collect if patients are expected to start alternative therapies that can have a significant effect on HR-QoL

More Alternative Estimands

- Hazard ratio of treatments A versus B for HR-QoL deterioration **due to study treatment and any subsequent anti-cancer therapy regardless of treatment duration** in all randomized subjects

Treatment Effect due to Study Treatment (Only)

- What if adequate data were not / cannot be collected?
- Use methods that account for likely deterioration of HR-QoL post study treatment discontinuation as compared to patients who continue the treatment.

Analyses with “delta-adjustment” (NRC, 2010; Carpenter et al., 2013; Ratitch et al., 2013; Permutt, 2016; Mehrotra et a., 2017, etc.)

(Permutt, 2016):

“… If missing data are not like observed data, what matters is whether they are more on average or less on average than the observed data, in each treatment group.

… Basically, the method is to predict the missing outcomes and then add values Δi to the predictions in group i, varying the Δi over a plausible range. ”

Treatment Effect due to Study Treatment (Only)

Analytical approaches:

- Patient-level imputation applying clinically justifiable adjustments to imputed values to reflect deterioration
  - Impute HR-QoL scores or time of HR-QoL deterioration
  - Use Multiple Imputation (MI) to account for uncertainty of imputations
- No individual patient-level imputation: control-based mean imputation

Pattern-Mixture Model Multiple Imputation

Imputing HR-QoL longitudinal scores

Pattern*	Experimental Treatment	Control Treatment
1: Discontinuation due to death	The worst score (zero) is assigned after death	
2: Discontinuation due to AEs	HR-QoL score worse by δ after discontinuation compared to predictions based on similar patients who continue treatment	- Similar as for experimental treatment, or - Imputed under MAR within their group, if control treatment AEs can be well managed
3: Discontinuation due to PD	HR-QoL score worse by δ after discontinuation compared to predictions based on similar patients who continue treatment	
4: Discontinuation due to other reasons	HR-QoL score imputed using the MAR assumption	

* Discontinuation refers to study treatment discontinuation

Pattern-Mixture Model Multiple Imputation

Imputing time to HR-QoL deterioration

- Similar pattern-based strategy
- Two options:

1. Impute time to deterioration based on imputed HR-QoL scores
   – Need to also impute assessment dates based on planned schedule of assessments

2. Impute event times directly

   δ - hazard ratio of having the event post-discontinuation for censored vs. completers, after adjusting for baseline covariates:

   the hazard at time $t > c_i$ (after censoring) is $\delta \times h(t)$ for some $\delta \neq 1$ compared to the hazard $h(t)$ of "similar" subjects with events or censored for administrative reasons

Imputation Model Choices for Time to Event

- Time-to-event imputation models that can be used with multiple imputation:
  - Kaplan-Meier (KMMI)
  - Cox regression model (COXMI)
  - Piecewise-exponential survival function (PCEMI)
  - Logistic regression within discrete intervals of follow-up period (PCLMI)

- For the parametric models PCEMI and PCLMI, it is possible to generate samples from a Bayesian joint posterior distribution of model parameters estimated based on a specified prior (e.g., uniform or gamma) and available data (Lipkovich et al., 2016)

- For non-parametric KMMI model and semi-parametric COXMI model, a version of approximate Bayesian bootstrap can be used (Taylor et al., 2002; Zhao et al., 2014; Lipkovich et al., 2016)

Some References for MI with Time-to-Event Data

Published in final edited form as:
J Biopharm Stat. 2014 ; 24(2): 229–253. doi:10.1080/10543406.2013.860769.

A MULTIPLE IMPUTATION METHOD FOR SENSITIVITY ANALYSES OF TIME-TO-EVENT DATA WITH POSSIBLY INFORMATIVE CENSORING

Yue Zhao[1], **Amy H. Herring**[2], **Haibo Zhou**[2], **Mirza W. Ali**[3], and **Gary G. Koch**[2]

MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1738

Published online 21 March 2016 in Wiley Online Library

Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints

Ilya Lipkovich,[a] **Bohdana Ratitch,**[b*] **and Michael O'Kelly**[c]

PharmaSUG 2017 - Paper SP05

Combining Survival Analysis Results after Multiple Imputation of Censored Event Times

Jonathan L. Moscovici, QuintilesIMS, Montreal, QC

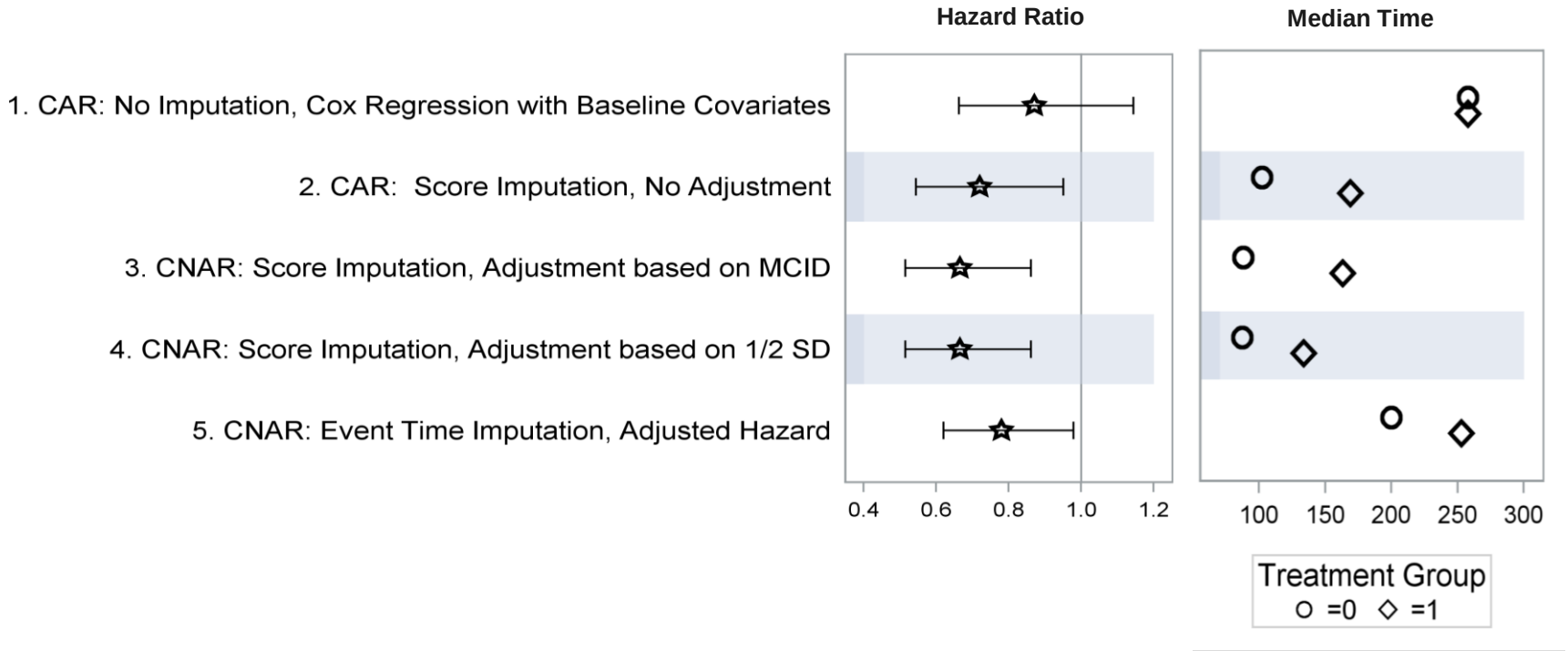
Bohdana Ratitch, QuintilesIMS, Montreal, QC

Choosing Adjustment Parameter δ

Choice of δ may be based on:

- An established Minimally Clinically Important Difference (MCID), e.g.,
  - MCID=6 to10 for FACT-P total score
  - MCID=2 to 3 for FACT-P subscales
    - Cella et al (2009). Estimating Clinically Meaningful Changes for the Functional Assessment of Cancer Therapy - Prostate: Results from a Clinical Trial of Patients with Metastatic Hormone-Refractory Prostate Cancer. Value in health 12: 124-129.
    - Yost & Eton (2005). Combining distribution and anchor-based approaches to determine minimally important differences. The FACIT experience. Eval Health Prof 28: 172.
    - Cella D, Hahn E and Dineen K. Meaningful change in cancer-specific quality of life scores: Differences between improvement and worsening (2002). Quality of Life Research 11: 207–221.
- ½ of SD at baseline
    - Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 2003; **41**(5):582–592.
- Historic studies in a similar indication that collected data post-treatment discontinuation
  - E.g., our analyses suggested that the hazard of HR-QoL deterioration increased 5 to 12 times after discontinuation due to PD or AE

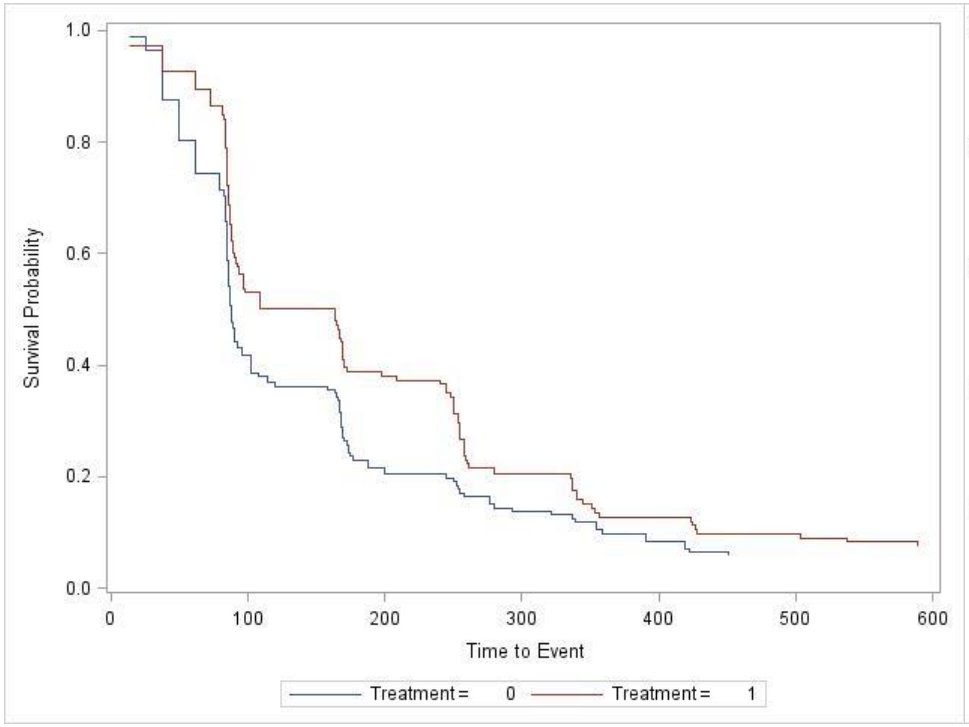
Example: Time to HR-QoL Deterioration – CAR and CNAR Analyses



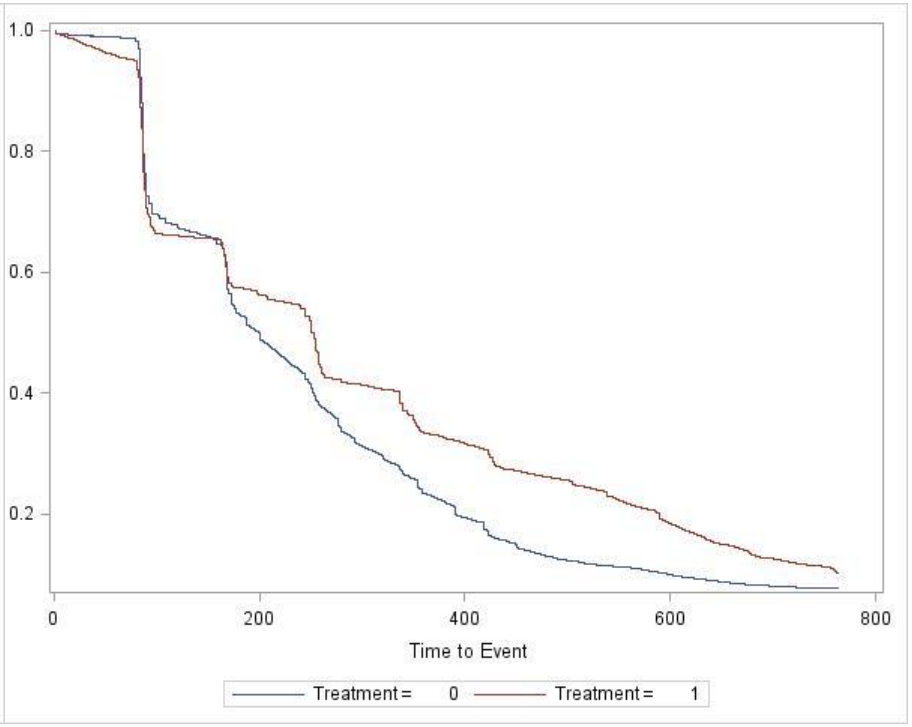
- Score imputation with adjustment based on mid-point of MCID range (δ=-8)
- Event time imputation with adjusted hazard δ=8

Example: Time to HRQoL Deterioration
Kaplan-Meier Curves, Combined after MI

Score Imputation, Adjustment Based on MCID



Event Time Imputation, Adjusted Hazard



NOTE: In practice actual assessment day is often different from the scheduled day.
With score imputation, dates are imputed at discrete scheduled assessment days;
With event time imputation, imputed event times can be between scheduled visit days (as in real data) and produce a smoother curve.

Summary

- Defining the estimand at study design helps to
  - provide clarity on the treatment effect being estimated and consideration of intercurrent events, e.g., discontinuations due to PD, toxicity and alternative therapies;
  - appropriately plan HR-QoL data collection;
  - choose analytical approach aligned with the estimand
- If post-treatment discontinuation HR-QoL data were not collected, for estimands "regardless of treatment duration", analysis should account for likely deterioration of HR-QoL after discontinuation
- If HR-QoL data after study treatment discontinuation are useful and collected, for patients who may still have missing data, perform "retrieved dropout" imputation – based on subjects with post-discontinuation data
- Multiple imputation methods, both for continuous score imputation and time-to-event imputation, provide valuable tools. Other methods, e.g., control-based mean imputation, non-parametric rank analysis, "composite event" analysis, etc., can also be considered.

Extra Slides

General MI Procedure for Time-to-Event Data

Carpenter JR, Kenward MG (2013) Multiple Imputation and its Application. John Wiley & Sons Ltd., West Sussex.

To generate a single imputed value for i^{th} subject, censored at time c_i:

- Estimate imputation model from observed data - survival function $\hat{S}\left(t\middle|\boldsymbol{x},\widehat{\boldsymbol{\beta}}\right)$
  - $\boldsymbol{x}$: covariate vector
  - $\widehat{\boldsymbol{\beta}}$: Bayesian draw of the imputation model's parameter vector
- Impute event times using inverse of failure function:

  (1) Compute $p_i = 1 - \hat{S}\left(c_i\middle|\boldsymbol{x}_i,\widehat{\boldsymbol{\beta}}\right)$

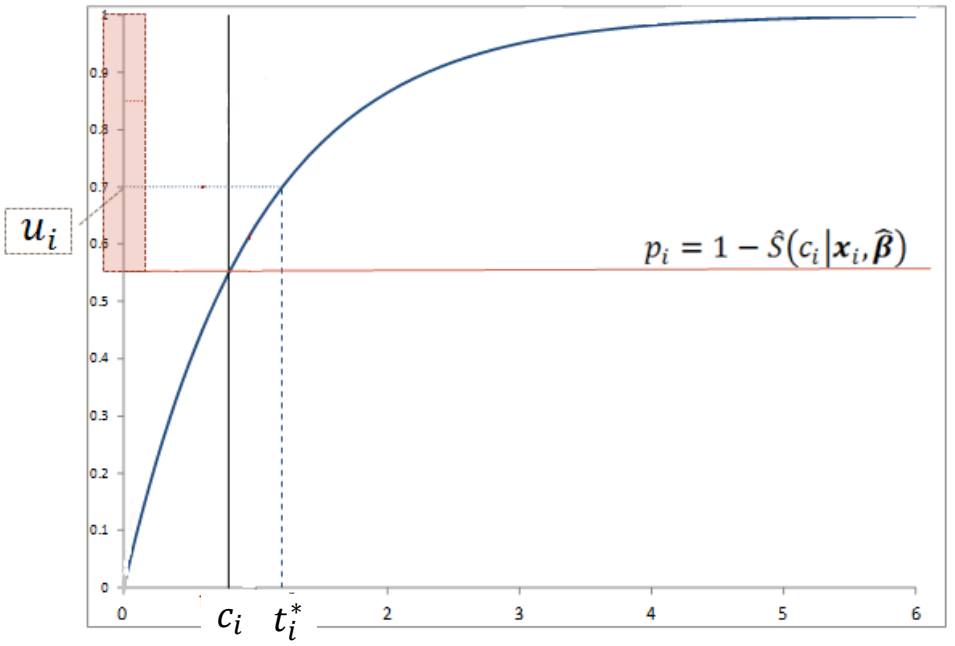
  (2) Draw u_i ~ uniform $[p_i, 1]$

  (3) Obtain the imputed event time t_i^* as

  the solution of $u_i = 1 - \hat{S}\left(t\middle|\boldsymbol{x}_i,\widehat{\boldsymbol{\beta}}\right)$

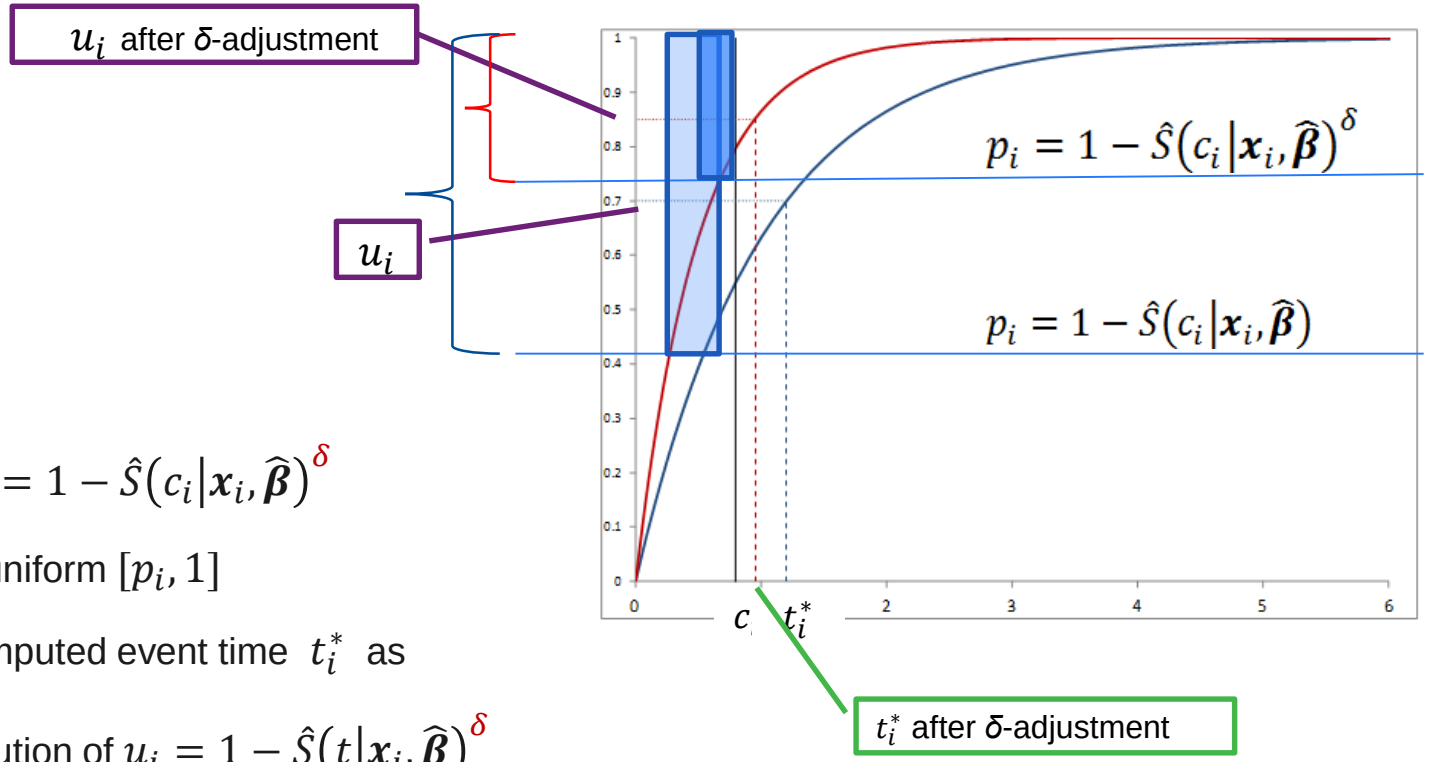
  (4) If $t_i^* > t_{max}$ censor at t_{max}

  (maximum follow-up by study design)



Imputing Time-to-Event Data with Delta Adjustment

To impute event times with delta adjustment, a small modification to the general procedure is needed:



(1) Compute $p_i = 1 - \hat{S}(c_i|\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})^{\delta}$

(2) Draw u_i ~ uniform $[p_i, 1]$

(3) Obtain the imputed event time t_i^* as

the solution of $u_i = 1 - \hat{S}(t|\boldsymbol{x}_i, \hat{\boldsymbol{\beta}})^{\delta}$

(4) If $t_i^* > t_{max}$ censor at t_{max} (maximum follow-up by study design)