Review of analytical methods for analysis of patient reported outcome (PRO) data in the presence of censoring due to death

IMS Health & Quintiles are now IQVIA™

O'Kelly M.[1], Kráľ P.[2], Ratitch B.[3], Niklasson A.[4,] Ivanescu C[5]
[1]IQVIA, Ireland, [2]IQVIA, Slovakia, [3]Eli Lilly, Canada, [4]AstraZeneca, Sweden, [5]IQVIA, Netherlands

AstraZeneca

Introduction

- Evaluation of symptoms and impacts such as health related quality of life (HRQoL) and survival are both important in characterizing disease progression in indications like oncology or heart failure
- There is a growing literature on options for estimating a treatment effect for patient reported outcomes (PROs), where an intercurrent event such as death censors the measure of the effect of interest. The censoring may be regarded simply as making further data collection impossible; but censoring may also impinge directly on what it is desired to estimate (i.e. upon the estimand), and may be required to be taken account of quantitatively
- In practice, estimands are such that death may fall into either of the following categories: for some estimands death may simply render the PRO unascertainable; for other estimands, such as those involving symptoms, impacts and HRQoL, death may be regarded as part of the PRO assessment, albeit a qualitatively different part, and the assessment of the PRO will need to accommodate the event of death
- The objective of this project was to evaluate approaches to analyzing PRO data in the presence of censoring due to death and other events

Methodology

- A targeted literature review was conducted, with advice from subject matter experts (both statistical and PRO experts) to identify and appraise the analytical approaches used in the literature for censoring due to death in the analysis of symptoms and impacts such as health-related quality of life. PubMed and Google were used to identify approaches that account for censoring of PRO data. Methods and their attributes were abstracted and summarised
- The approaches identified were further discussed with experts and a short list was prepared that was evaluated with respect to the assumptions required by each method, their limitations, ease of implementation and interpretability

Results

- We reviewed 26 publications and identified 9 approaches (Table 1)
- Two routes were identified among the methods: 1) **Integrate** the occurrence of death into the PRO measure in some way; or 2) **Stratify** the study population by event status determined irrespective of treatment assignment

Table 1: Summary of attributes of identified statistical approaches to analyze PRO taking account of death, based on the literature survey

Statistical approach	Description	Advantages	Disadvantages
Integrate the event			
Win Ratio[1]	Similar idea to that of the U-statistic: each patient in the experimental arm is compared with each patient in the control arm and is labeled as "winner", "loser", or "tie" in each comparison, with death worse than worst observed PRO at visit	• No categorizing of PRO needed; works directly with ranks • If there are not too many ties, the power is similar to traditional rank-based methods – date of death can be used to minimise ties • Provides hypothesis test and interpretable treatment effect • Relatively easy to implement (SAS code available)	• It loses power when there are too many ties • Interpretation may be difficult because survival and HRQoL are in the same set of ranks
Rank ANCOVA[2]	ANCOVA analysis of outcome ranked as for Win Ratio and covariates; with inference from Cochrane-Mantel-Haenszel test of residuals	• As above: no categorizing of PRO needed; works directly with ranks • Provides hypothesis test • Easy to implement	• It loses power when there are too many ties • Lacks direct estimate of treatment effect; could use quantile regression in conjunction
Continuation ratio modelling[3]	Models the effect of treatment on the probability of worsening in terms of the composite outcome of PRO/death, conditional on attaining a specific outcome category. With logit link, models $\delta_j(\boldsymbol{x})=P(Y_i>j \mid Y_i \geq j, \boldsymbol{x})$;can also model hazard. Allows the odds/hazards to vary with j, but usually assumes effect of, e.g., the treatment is constant across the outcome categories j	• Conditioning on all categories "lower" than j to estimate $P(Y_i>j)$ may not be justifiable, and continuation ratio modelling avoids this. E.g., avoids conditioning on death and non-death statuses together • Odds/hazards for different categories are not assumed the same; homogeneity of odds can be estimated/assessed • Model fit statistics can provide insight as to the goodness of fit with different categorizations • Weight for lower/higher categories can be prespecified	• Not common in clinical trials; its interpretation may be difficult for unexperienced personnel • It needs cut-offs to define categories • Power compared to rank-based methods may be poorer unless categories >=4 and categories evenly distributed • For interpretable overall treatment effect, effect needs to be similar across categories, conditioning on all other explanatory variables
Ordinal regression with cumulative probability model (CPM)[4]	Models an observed ordinal (ranked) outcome, Y, as a monotonic transformation of a latent variable Y*, in turn modeled as linear combination of covariates X via link function G:$G[P(Y \leq y_i \mid \boldsymbol{X})]=\alpha_i - \boldsymbol{\beta}^T\boldsymbol{X}$. Since probability is cumulative, in the model the association between X with death and with the PRO are treated together	• The CPM directly models the conditional cumulative distribution function, from which other components of the conditional distribution, e.g., mean and quantiles, and probabilities associated with specific thresholds of interest can be obtained within each treatment arm • It can be used as a non-parametric or semi-parametric hypothesis test	• It relies on the "parallelism assumption" across all statuses including death, for all covariates including treatment group, as well as a semi-parametric model assumption associated with the link function • Estimate of treatment effect not easily interpretable. • The best link function selection approach is based on model fit – a data driven approach which may be problematic for primary analysis
Responder (binary) analysis[5]	It defines responder-type endpoints, e.g., as ≥X points of deterioration or improvement (analyzed separately) from baseline at timepoint Y, with death counted as failure/non-responder	• Ease of interpretation, if assumptions are acceptable • Ease of implementation	• The power may be lower compared to other methods • Death counted similarly to other levels of non-success • It is not clear if responder definitions are acceptable. Anchor-based analyses may be requested for justification
MMRM with zeroes for death[6]	A standard MMRM analysis can be carried out on a dataset where subject assessments at scheduled time points after death are imputed with a value of zero	• Ease of implementation	• Variance of outcome distorted post-death • It is difficult to interpret treatment effect without a separate consideration of the proportions of deaths in each arm • Increasingly deprecated by regulatory agencies
AUC as a patient-level measure[7]	The HRQoL measurements taken at discrete time points over time for each subject are used to create a subject-specific "curve" over a period of time from baseline at the time point of interest	• It preserves continuous nature of HRQoL measurements • It is commonly used in some analyses (e.g., for QALYs or Q-TWiST) and thus, people are familiar with this method	• It may not discriminate between long survival/poor HRQoL and shorter survival/great HRQoL
Joint modelling of PRO/death[8]	The two components of the joint model, i.e., the survival model and longitudinal mixed effects model of PRO scores, share random effect coefficients, which are estimated simultaneously with other model parameters	• Small differences between the unconditional and partially conditional HRQoL results could serve as an indication that treatment difference is not strongly driven by differences in survival • It may be helpful as supportive analysis illuminating relationship of HRQoL changes and survival in long term	• It may not be useful if the difference between treatments in survival during the time that the PRO is modeled is small • It relies on assumptions about the form of the longitudinal model, e.g., linear or quadratic, albeit with random coefficients for the corresponding effects
Stratify by the event			
Model PRO conditioning on subject's status with regard to event irrespective of treatment assigned (=subject's *principal stratum)*.[9]	Estimate treatment effect in stratum of interest (stratum of interest is usually stratum of those who would always survive irrespective of treatment) Multiple imputation (MI) can be used to allow appropriate uncertainty with regard to stratum. Stratum is modelled via baseline covariates. Death (Y/N) or time-to-death models can be used	• Ease of interpretation: any standard analysis of the PRO can be performed on the stratum identified. • Ease of implementation: software is available to predict stratum membership, i.e., survival and/or death(Y/N), using MI methodology while accounting for modelling uncertainty	• Inference is on stratum, so conclusions do not apply to all randomized subjects • It implies strong assumptions of explainable non-random survival • It is difficult to choose agreed predictors of survival

Discussion

- **To our knowledge, this is the first review reporting and assessing approaches to analyze PRO data in the presence of censoring due to death and other events**
- **Methods that integrate occurrence of death into the PRO measure avoid model assumptions but require ranking or coarsening of outcome. As a result the estimate of treatment effect can be difficult to interpret**
  - **For this route, we recommend considering *Win Ratio; Continuation Ratio Modelling; Rank ANCOVA+quantile regression***
- **Stratifying by the event provides a very nice interpretation in the face of obstructing event, but requires a credible model for membership of stratum**
  - **For this route, we recommend the MI approach. However, its main use is likely to be in supportive analyses because of limited inference (i.e. inference by stratum)**
- **In conclusion, the magnitude of the estimate of treatment effect will vary depending upon the approach taken to the event of death. With increasing potential diversity of estimands and resulting estimates in this context, comparison between treatments may become challenging**

[1]Wang D et al. Pharmaceut. Statist. 2016, 15 238–245; [2]Stokes M. et al Categorical Data Analysis 1995; [3]Agresti A. Categorical Data Analysis 2002; [4]Liu Q et al. Statistics in Medicine. 2017;36:4316–4335; [5]Stokes M. op. cit; [6]Mallinckrodt C. Drug Inf. Journal 2008, 42 308-319; [7]Feldstein ML. Cancer. 1991 Feb 1;67(3 Suppl):851-4; [8]Li Q. et al. J R STAT SOC 2018;67(1):145-163; [9]Rubin D. Statistical Science 2006, 21,(3), 299–309

International Society for Pharmacoeconomics and Outcomes Research, Annual Meeting Europe, poster presentation

Disclaimer: All authors have contributed equally to this research and no payments were made by AstraZeneca for work on the poster. The study was supported by AstraZeneca including financial support.