ISPOR EU - Issue Panel
How shall we analyze change from baseline in COA endpoints? Is MMRM bowing out?

Talk 3: How should death be dealt with when estimating the treatment effect in repeatedly collected COAs?

Konstantina Skaltsa, *Director, Statistical Services, Patient-Centered Solutions (PCS)*

7th November 2022

Questions for which we will seek answers in this talk

1. Are we currently using MMRM in clinical trials where death is an occurrence?
2. Why is this not an appropriate method in this setting?
3. What do regulators say about MMRM when death happens?
4. Shall we be thinking of the estimand first?
5. Are there options in the current statistical literature?

MMRM: Mixed Model Repeated Measures

Let's all align on an example setting

Setting

- 2-arm (active vs control) phase 3 clinical trial in a late-phase solid-tumour oncology indication
- Primary endpoint is PFS or OS
- **Change from baseline in QoL or symptoms X at Week Y** is a (key) secondary endpoint – there may be label claims, but not relevant to the discussion
- Patients are treated until disease progression, unacceptable toxicity, investigator's decision etc
- Death may occur in these trials **prior to Week Y** resulting in missing data at the timepoint of interest

How should death be dealt with when estimating the treatment effect in repeatedly collected COAs? Konstantina Skaltsa, IQVIA, 7th November 2022, ISPOR EU Conference

PFS: Progression-Free Survival; QoL: Quality of Life; OS: Overall Survival

Q1: Are we currently using MMRM in clinical trials where death is an occurrence?

Yes, we are.

- There is plenty of literature reporting clinical trial results with this method implemented.

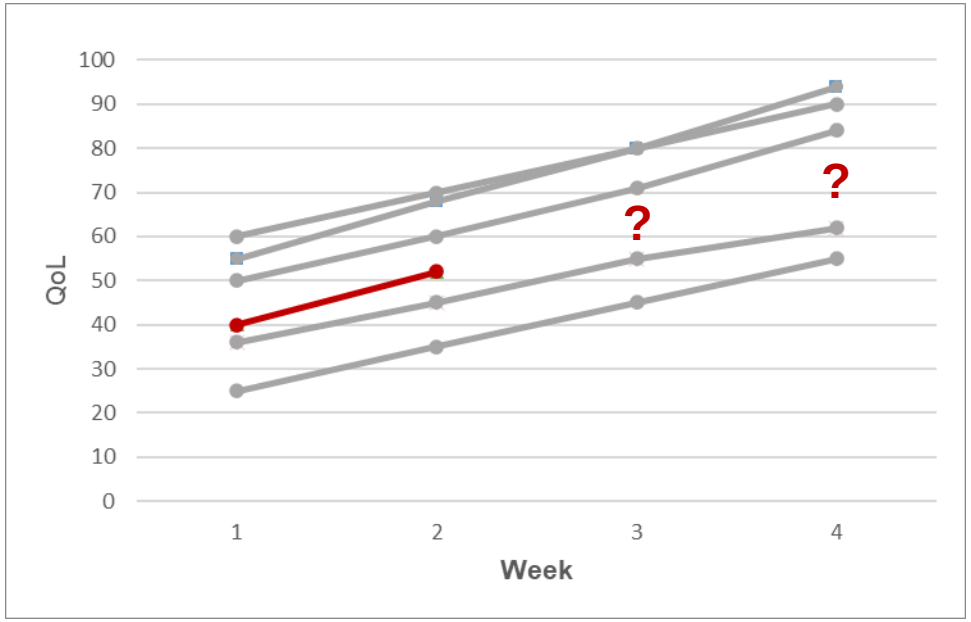
"We analyzed quality of life with mixed effects models for repeated measures with baseline values as a covariate… 670 patients were randomized; 657 patients died at long-term FU"[1]

"For symptom analysis, comparisons of LCS between arms were conducted using a linear mixed-effects model in which the missing data depend on the observed LCS… LCS data were missing in 111 surveys because of death or severe impairment of the patient's general condition; this accounted for 6.2% of the total number of surveys scheduled."[2]

[1]https://doi.org/10.1016/S1470-2045(12)70412-6.
[2]Journal of Clinical Oncology 2010 28:5, 753-760

Q2: Why is this not an appropriate method in this setting?

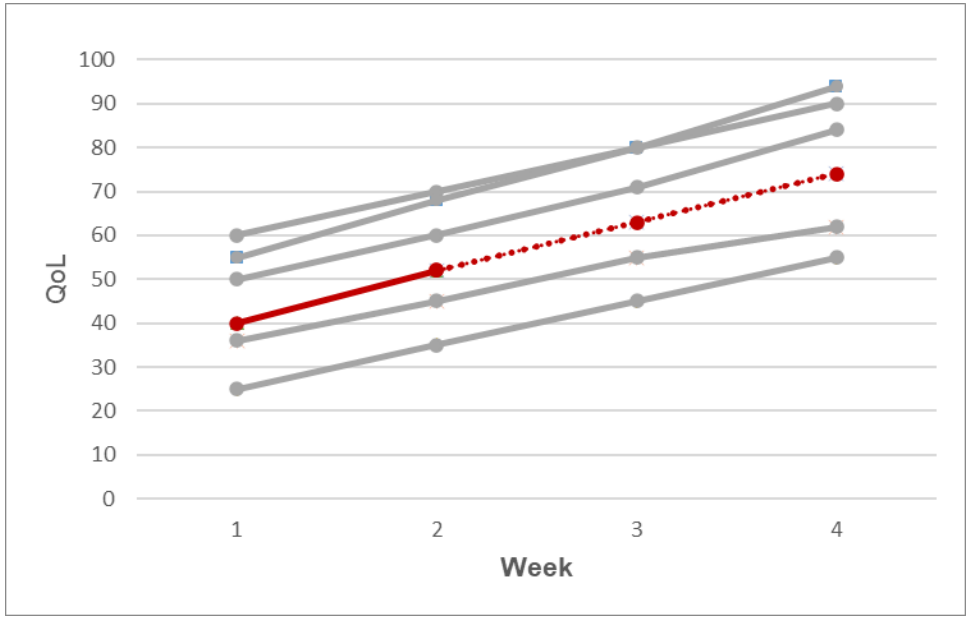
MMRM makes the assumption that the missing data are missing at random (MAR)



Red solid line
Observations of subject that has an ICE at week 2

Grey solid line
observations of subjects in the same treatment arm who have similar baseline characteristics

Red dashed line
Inferred values under MAR



Under MAR, the MMRM model estimates the mean treatment effect assuming that "*. . . after withdrawal, subjects* ***would have continued*** *just like their peers in the same arm who have the same covariates and same observed data (so far)".*

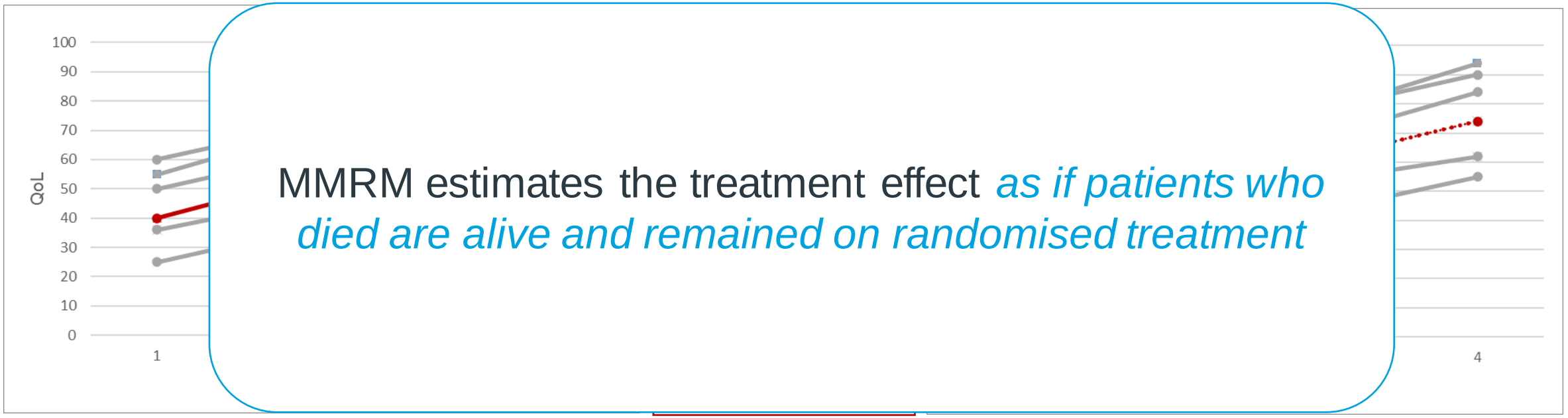
Hypothetical language ☺

Slide presented at ISOQoL 2021 "An estimand perspective on the Mixed Model Repeated Measures (MMRM) for the analysis of longitudinal PRO data in clinical trials". Oral presentation by Konstantina Skaltsa.

Quote by James Roger. https://www.psiweb.org/docs/default-source/resources/psi-subgroups/scientific/2015/estimands-28-09-2015/jamesroger.pdf?sfvrsn=bba3d0db_2
Graph inspired by presentation by Jiawei Wei "On the role of hypothetical estimand in clinical trials and its estimation" (PSI One-day meeting: Missing data in clinical trials – Past, present and future, 4th May 2021)

ICE: Intercurrent Event; MAR: Missing At Random; MMRM: Mixed Model Repeated Measures

MMRM makes the assumption that the missing data are missing at random (MAR)



MMRM estimates the treatment effect *as if patients who died are alive and remained on randomised treatment*

Under MAR, the MMRM model estimates the mean treatment effect assuming that "*. . . after withdrawal, subjects* ***would have continued*** *just like their peers in the same arm who have the same covariates and same observed data (so far)".*

Hypothetical language ☺

Slide presented at ISOQoL 2021 "An estimand perspective on the Mixed Model Repeated Measures (MMRM) for the analysis of longitudinal PRO data in clinical trials". Oral presentation by Konstantina Skaltsa.

Quote by James Roger. https://www.psiweb.org/docs/default-source/resources/psi-subgroups/scientific/2015/estimands-28-09-2015/jamesroger.pdf?sfvrsn=bba3d0db_2
Graph inspired by presentation by Jiawei Wei "On the role of hypothetical estimand in clinical trials and its estimation" (PSI One-day meeting: Missing data in clinical trials – Past, present and future, 4th May 2021)

ICE: Intercurrent Event; MAR: Missing At Random; MMRM: Mixed Model Repeated Measures

Q3: What do regulators say about MMRM when death happens?

A couple of FDA responses from the Oncology division

*FDA has major concerns regarding the statistical analyses as proposed: In general, PROs for superiority and non-inferiority **may not be interpretable for efficacy due to mortality**. The mixed model repeated measures (MMRM) relies on the assumption that data are missing at random (MAR). **If a patient is missing due to death, the MAR assumption is likely not a reasonable assumption**, which can lead to bias in the estimated treatment effect.*

FDA Oncology Division 2021

*We are concerned about the interpretability of Physical functioning/Global health status/QOL for efficacy **due to the observed mortality on this trial**. Mixed Model Repeated Measures (MMRM) relies on the assumption that data are missing at random (MAR), therefore **if a patient is missing due to death, the MAR assumption is likely not a reasonable assumption**. This could lead to bias in the estimated treatment effect.*

FDA Oncology Division 2021

How should death be dealt with when estimating the treatment effect in repeatedly collected COAs? Konstantina Skaltsa, IQVIA, 7th November 2022, ISPOR EU Conference

FDA: Food and Drug Administration; MAR: Missing At Random

Q4: Shall we be thinking of the estimand first?

Brief reminder of the recent past and the present/future

Old era

- Protocol objective: *"assess the QoL/symptoms in patients with …"*
- Protocol endpoint: *"change from baseline in QoL/symptoms"*
- Trial begins and clinical SAP is drafted
- Statisticians to decide how to best estimate the treatment effect corresponding to the protocol endpoint; i.e.

What to do with missing data (e.g. due to death) usually an issue for Statisticians to solve

Consequences

1. Treatment effects presented in CSR generating more questions than answers
2. Post-hoc analyses looking for (post-hoc) more plausible solutions

Estimands era

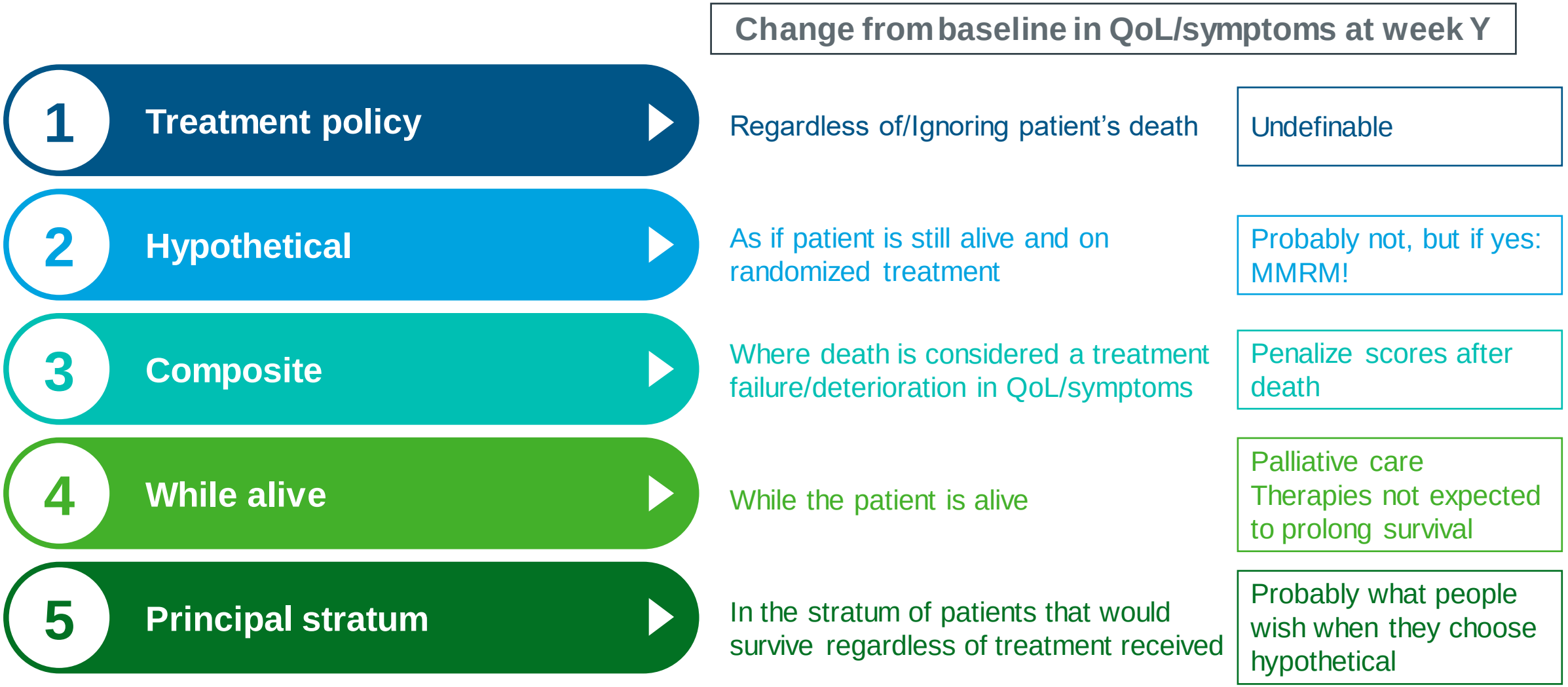
- Objectives, endpoints and analyses framed using Estimand language
- Death is an *Intercurrent Event* (ICE)…*"precluding observation of the outcome of interest"*
- *Strategies* for dealing with ICE should be described in the protocol (i.e. not only in the SAP)
- The ICH E9 (R1) dictates the following 2 steps (…grossly…)

The clinical question of interest is discussed and decided first, including a multi-disciplinary team; ideally at the time of the study design

Methods for estimating the treatment effect of interest should follow naturally if Step 1 was carefully executed

CSR: Clinical Study Report; Intercurrent Event = ICE; MMRM: Mixed Model Repeated Measures; QoL: Quality of Life; SAP: Statistical Analysis Plan
ICH E9(R1) Final version. Feb 2020

So what is the treatment effect we are interested in?



How should death be dealt with when estimating the treatment effect in repeatedly collected COAs? Konstantina Skaltsa, IQVIA, 7th November 2022, ISPOR EU Conference

QoL: Quality of Life; MMRM: Mixed Model Repeated Measures

Q5: Are there options in the current statistical literature?

Some potential estimators for dealing with the ICE of death

Estimand	Estimator	Advantages	Disadvantages
Hypothetical	MMRM	• Well-known method, easy to implement	• Missing At Random assumption generally implausible ("if patient were still alive")
While on treatment (i.e., while alive)	Area Under the Curve (AUC)[1]	• Well-known patient-level endpoint, straightforward analysis • Standardized and unstandardized version can be considered if patients on one drug tend to die earlier (**unstandardized AUC can also be categorized as Composite*)	• It may not discriminate between long survival/poor health-related quality of life (HRQoL) and shorter survival/great HRQoL • Some implementations (e.g. standardized AUC) ignore death
Composite	Responder analysis[2]	• Easy to interpret • Conservative	• Loss of power • Does not preserve the continuous nature of the COA variable • Treats all non-responders equally (e.g., earlier deaths or deaths due to other causes)
	MMRM with worst score imputation[3]	• Worst score may be appropriate for short-range scales (e.g., 0-3)	• Selection of post-mortem value for COAs challenging • Variance of outcome post-death distorted
	Rank-based Analysis of Covariance (ANCOVA)[2]	• Based on ranks, rather than scores	• Provides p-value only, no estimate of treatment effect
	Quantile regression[4]	• Provides treatment effect estimate on original scale	• May not work if too many deaths
	Win Ratio/Win Odds[5]	• Based on ranks • Provides interpretable treatment effect	• Treatment effect is not on original scale, therefore harder to communicate to clinicians/patients
Principal stratum	Any standard analysis estimating a treatment effect in stratum of interest (stratum of those who would survive irrespective of treatment)[6]	• Multiple imputation (MI) can be used to allow appropriate uncertainty with regard to stratum.	• Inference is on stratum, not on Intention-To-Treat population • Strong assumptions when predicting belonging to stratum

How should death be dealt with when estimating the treatment effect in repeatedly collected COAs? Konstantina Skaltsa, IQVIA, 7th November 2022, ISPOR EU Conference

Table presented at ISOQoL 2022 poster presentation by Skaltsa K et al "How should death be handled when estimating the treatment effect in repeatedly collected COAs?"

[1]Feldstein ML. Cancer. 1991 Feb 1;67(3 Suppl):851-4; [2]Stokes M. et al Categorical Data Analysis 1995; [3]Mallinckrodt C. Drug Inf. Journal 2008, 42 308-319; [4]Mehrotra DV, Liu F, Permutt T. Pharmaceutical Statistics. 2017;16:378–392; [5]Wang D et al. Pharmaceut. Statist. 2016, 15 238–245; [6]Rubin D. Statistical Science 2006, 21,(3), 299–309