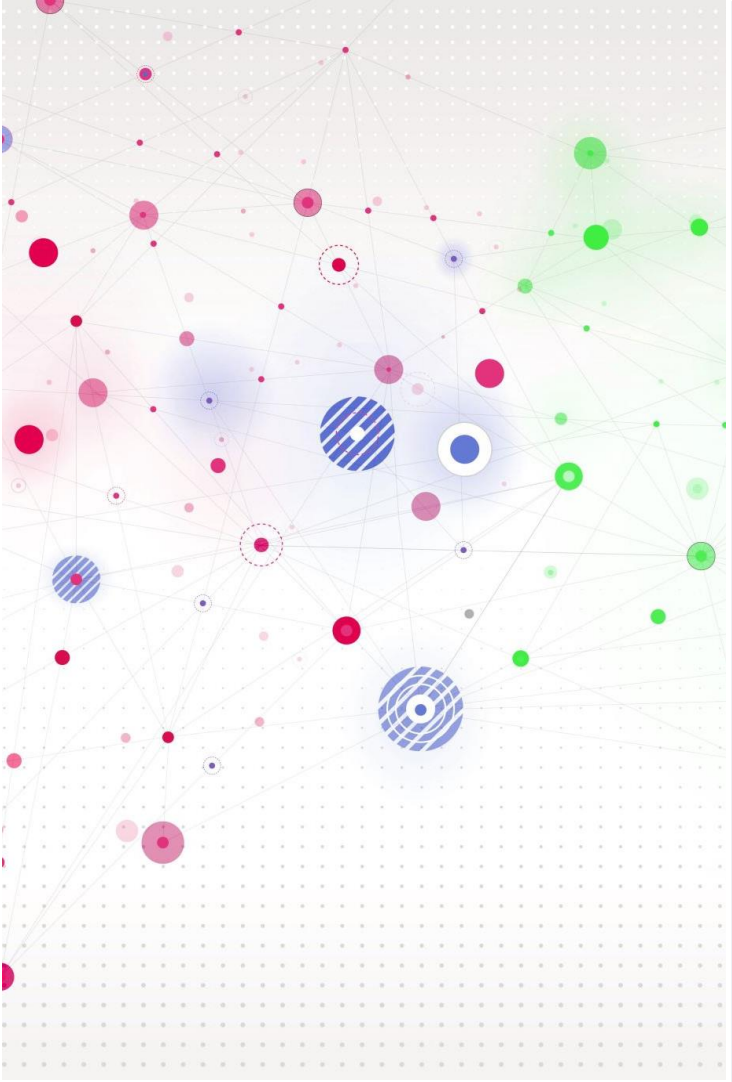


Estimands in Oncology Working Group: PRO Task Force

What is are the PRO objectives & how are they being analysed – applying the estimand framework

"Backfitting" – MMRM & Time to deterioration

V1_0

Dec 2021

PRO Task Force – Members/Contributors

Member	Affiliation
Rachael Lawrance	Adelphi Values
Libby Floden	Clinical Outcome Solutions
Konstantina Skaltsa	Iqvia
Jonathan Siegel	Bayer
Evgeny Degtyarev	Novartis
Antoine Regnault	Modus Outcomes
Xiangning Huang	AstraZeneca
Stacie Hudgens	Clinical Outcome Solutions
Johan Bring	Statisticon
Stephen Corson	Phastar
Monica Hadi	Evidera

Contents

Introduction

What are PRO endpoints?

- Patient-reported outcomes (PRO) endpoints are commonly included as secondary endpoints in oncology clinical trials to evaluate health-related quality of life (HRQoL)
  - This slide-deck uses an example of a 2-arm (active vs control) phase 3 clinical trial in a late-phase solid-tumour oncology setting, considering QoL as a secondary endpoint, where the primary endpoint is PFS or OS.

Variety of endpoints used

Key Challenges

- A key challenge is that there is not standard use, nor exact definitions of, QoL endpoints across healthcare industry, even when measuring the same concept
- There is generally a lack of clarity of the precise scientific question targeted for PRO analysis
- The variety of endpoints limits comparisons between trials
- Differences in protocols lead to different data points collected e.g. PRO data collected at clinical visits until disease progression, or whilst on-treatment?
- Examples:
  - "Change from baseline in QoL"
    - Does that mean At Week X, by Week X, over-time (until when?), on-treatment??
  - "Time to deterioration in QoL"
    - Deterioration in QoL score or progression or death?
    - Thresholds for deterioration – should it be confirmed? Limitations of the questionnaire
    - Exact censoring rules (missing timepoints) – rules really depend on the research question
- If the research question of interest is made clearer, then analysis methods used can be followed – the estimand framework can help to ensure clearer definition of the research question
- "Backfitting" highlights the importance of clarifying the actual research question of interest

Study example #1

Example clinical trial #1

- Phase 3 randomised clinical trial; 2-arm (active vs control)
- Late-phase solid-tumour oncology
- Primary endpoint: PFS
- Secondary endpoint: QoL

- PRO data collected at baseline and every 4 weeks until disease progression (or withdraw from study if prior to disease progression)
  - PRO data collected every 4 weeks **even if patients discontinue randomised treatment,** until disease progression
  - No PRO data collected post-progression

Study example #2

Example clinical trial #2

- Phase 3 randomised clinical trial; 2-arm (active vs control)
- Late-phase solid-tumour oncology
- Primary endpoint: PFS
- Secondary endpoint: QoL
- PRO data collected at baseline and every 4 weeks until disease progression or withdraw from study AND collected post progression
  - PRO data collected every 4 weeks **even if patients discontinue randomised treatment**, until disease progression
  - AND PRO data collected post progression every 6 weeks for up to 2 years

“Backfitting” - MMRM

“Backfitting” - MMRM

MMRM Model

- Change from baseline in PRO score – commonly analysed using a repeated measures mixed model (MMRM)
  - Accounts for multiple visits per subject
  - Unbiased if unobserved data are MAR
  - Treatment effect presented: Difference in LSmeans between active vs control - supported by directionality of the overall change from baseline in Lsmeans over the period of interest
- It has been recommended by SISAQoL (2020): *“Although the linear mixed model (time as continuous), pattern mixture model, and joint longitudinal model satisfy the set criteria,* ***the linear mixed model (time as discrete) was recommended because less assumptions were needed to be made a priori*** *(eg, regarding the relationship between time and outcome variable)”.*

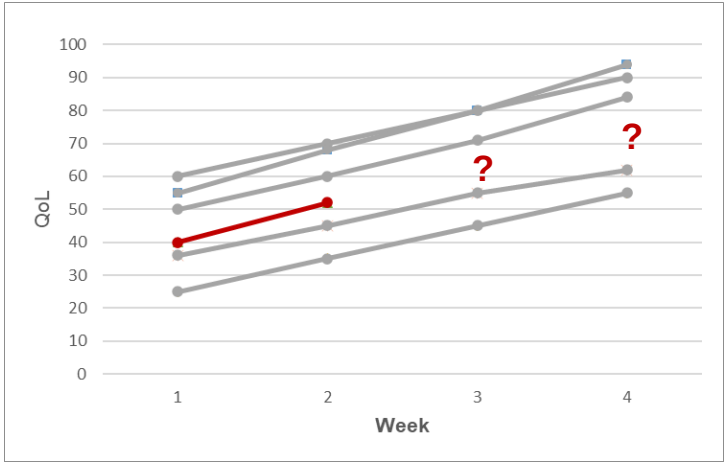
“Backfitting” - MMRM

MMRM model

- Fitting an MMRM – all unobserved data is treated the same – whether data is missing if patients discontinued treatment or died – the MMRM approach is a “hypothetical” approach

- Different data points may be included in MMRM analyses – often these decisions are not made clear. Is it:
  - all data points pre-progression;
  - all data until death; or
  - all data only while “on-treatment”?

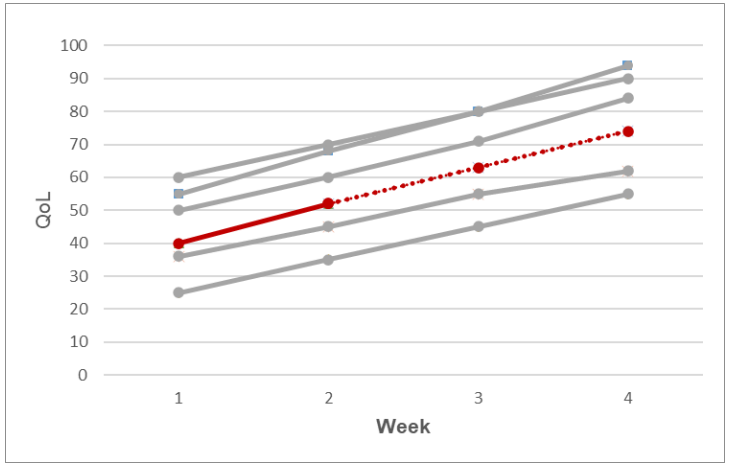
MMRM makes the assumption that the missing data are missing at random (MAR)



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

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: sMissing data in clinical trials – Past, present and future, 4th May 2021)

MMRM – all ICEs handled as hypothetical

Example objective in protocols, publications etc

"Change from baseline in QoL"

Timeframe

1. At a specific timepoint?
2. Over time, by timepoint X?

Non-exhaustive list of possible questions – including several unreasonable ones

Death	Discontinuation of randomized treatment	Receipt of other treatments
1. While alive?	1. While on randomized treatment?	1. Before switching to other antineoplastic treatments?
2. If patients had not died?	2. Regardless of treatment discontinuation?	2. Regardless of treatment received?
3. In survivors?	3. If patients had not stopped randomized treatment?	3. If patients had not switched to other therapies?

Estimands using MMRMs: Example Study 1

PRO data collected every 4 weeks until disease progression (even if discontinue treatment)

Population	Treatment	Variable	Variable - timepoint	Variable – data used	ICE	Missing data	Summary measure
Cancer patients	Active vs control (trt until disease progression)	Change from baseline in PRO score (be specific e.g. pain)	**Until month x (e.g. 6 months)**	PRO data until month 6	Prior to 6 months: • Disease progression - *hypothetical* • Death – *hypothetical*	Withdraw from study	LSmean difference between treatments

In cancer patients, what is the difference in mean change from baseline in PRO score treatment X compared to treatment Y, after 6-months from randomisation, if patients had not died or progressed?"

> *This is using a hypothetical approach for ICEs of death and disease progression*

> *THINK – is this the scientific question you are really interested in/relevant??*

Estimands using MMRMs for Example Study 2

PRO data collected pre and post progression for up to 2 years

Population	Treatment	Variable	Variable - timepoint	Variable – data used	ICE	Missing data	Summary measure
Cancer patients	Active vs control to disease progression **then any treatment**	Change from baseline in PRO score (be specific e.g. pain)	**2 years post randomisation (pre or post progression)**	**All PRO data (pre and post progression)**	*Death – hypothetical*	Discontinue treatment – tox Discontinue trt – other Withdraw from study	LSmean difference between treatments

"In cancer patients, what is the difference in mean change from baseline in PRO score treatment X compared to treatment Y (and any post-progression treatments), up to 2 years post-progression, if patients had not died?"

> *This is using a hypothetical approach for ICEs of death – other "events" now not formally ICEs*
> *THINK – is this the scientific question you are really interested in/relevant??*
> *In practice how can you collect this data – is it possible now/in the future?*

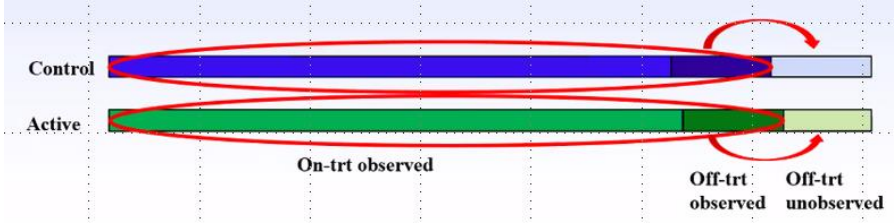
Including post-progression PRO data in a standard MMRM does *not* serve a treatment policy approach

1- Apparently

- The (presumably) poorer post-progression/off-treatment values will be used by the model to predict more pessimistic trajectories for similar patients that have missing data (although planned to be collected, these will probably occur at some point)

2- The reality#

- MMRM will use all observed data to infer unobserved data
- MAR assumes that off-txt unobserved patients are like all of the observed patients, conditional upon other patient characteristics and previous responses
- Observed data are primarily on-txt, e.g. if 90% observed data are on-txt, we would be implicitly *imputing the unobserved data as being 90% on-txt*
- Essentially, the issue is that such a model is **not making distinction between on- and off-txt assessments**, therefore we cannot claim it is estimating a treatment effect "regardless of txt status"



3- How do we do treatment policy then?#

A variation of the MMRM model with an introduction of **time-dependent off-treatment covariate** could be employed if **treatment policy** is desired#:

- Treatment/visit/off-trt interactions
- Combine estimates by observed on-/off-trt proportions and adjust variance

Schematic and explanation as in presentation by James Bell "The practicalities of treatment policy estimation" (PSI One-day meeting: Missing data in clinical trials – Past, present and future, 5th May 2021)

Key take-away messages

MMRM CAN

- MMRM, in its most commonly applied form, applies a **hypothetical** strategy for any ICE after which data are unobserved, e.g. *"Change from baseline in QoL as if patient is still taking randomized treatment"*
- If post-ICE data are collected, and an on- and off-ICE indicator is included, then **treatment policy** (for that ICE) is possible

MMRM CANNOT

- Answer questions like – "*Change from baseline in QoL…*"
  - **While on (randomized) treatment**
  - **Before progression**
  - **While alive**
  - **In survivors**
  - **In patients that would tolerate txt**
  - **Where death is a "bad" outcome**

While-on-treatment

Principal stratum

Composite

- When such questions are of interest, analysts should seek analytical solutions beyond the standard MMRM framework

“Backfitting” - MMRM

Consdierations

- Imperative that MMRM cannot be adequately interpreted alone
- What was the study design (PRO data collected until what point or until what ICE?)
- What timepoints are included in the MMRM
- What is the frequency of intercurrent events and impact of unobserved or missing data
- MMRM for change from baseline – only patients with baseline + post-baseline – generalisable to all randomised?
- Need to consider the variable – timeframe (end timepoint) (same for all patients, or vary by treatment?), and what data collected to be used – timepoint is critical – same for all patients (in MMRM)
- Helpful to identify missing data reasons to help identify what are ICEs and what are actually missing data
- Is the same strategy for all ICEs appropriate – same for death and treatment discontinuation?
- Is the analysis performed addressing a clear scientific question relating to quantifying the effect of treatment e.g. if collecting post progression data ?

“Backfitting” – Time to deterioration

“Backfitting” – Time to deterioration

Key concerns:

> Threshold for deterioration (individual decline in PRO score) – not focus here

> Many different definitions for time to deterioration used

> Also consider is deterioration in PRO score expected in the disease setting?

Possible definitions
Time to first PRO deterioration, disease progression or death
Time to first PRO deterioration or death
Time to first PRO deterioration
Time to definitive deterioration
Time to confirmed deterioration
Time to the first two consecutive deteriorations

“Backfitting”

Censoring

- Common to include a composite approach for some ICE (e.g. death or disease progression)
  - But not always clear if included or censored
- Censoring rules themselves not always clear
  - Censor those with missing PRO score at baseline at baseline?
  - Censor those with missing post-baseline PRO score at baseline?
  - Censor at last PRO assessment timepoints if deterioration not observed?
  - Censor at date of disease progression?
  - Rules for censoring for “missing” observations not always clear (e.g. missing scheduled PRO assessments – especially in survival follow-up)
  - Often radiological disease progression not considered to represent a patients reported deterioration in QoL – BUT study designs often do not collect PRO frequency post disease progression to assess, so is it reasonable to consider as event/timepoint for censoring?

Time to deterioration for Example Study 1 *#1*

PRO data collected every 4 weeks until disease progression (even if discontinue treatment)

Population	Treatment	Variable	Variable - timepoint	Variable – data used	ICE	Censoring rules	Summary measure
Cancer patients	Active vs control (trt until disease progression)	Time to 1st deterioration in **PRO score**	**Until patients disease progression**	All PRO data only	Disease progression – *hypothetical (censor last PRO)* Death – *hypothetical (censor last PRO)*	No baseline PRO – censor day 1 No post-baseline PRO – censor day 1 No event – censor last PRO assessment	HR

In cancer patients, what is the difference in time to PRO score deterioration between treatment X compared to treatment Y, if patients had not died or progressed"

> *This is a hypothetical approach for all ICEs – ie censoring*

> *THINK – is this the scientific question you are really interested in/relevant??*

Time to deterioration for Example Study 1 #2

PRO data collected every 4 weeks until disease progression (even if discontinue treatment)

Population	Treatment	Variable	Variable - timepoint	Variable – data used	ICE	Censoring rules	Summary measure
Cancer patients	Active vs control (trt until disease progression)	Time to 1st deterioration in **PRO score, or disease progression or treatment discontinuation or cross-over or death**	**Until patients disease progression or earlier event**	All PRO data, dates of disease progression, treatment discontinuation, start of new therapy, death	All included as events	No baseline PRO – censor day 1 No post-baseline PRO – censor day 1 No event – censor last PRO assessment	HR

In cancer patients, what is the difference in time to PRO score deterioration, disease progression, treatment discontinuation, start of new therapy or death between treatment X compared to treatment?"

> *This is a composite approach for all ICEs (censor only for when missing event data)*

> *THINK – is this the scientific question you are really interested in/relevant?? – maybe a combination of composite for some and hypothetical for other ICEs more relevant*

Time to deterioration for Example Study 2

PRO data collected pre and post progression for up to 2 years

Population	Treatment	Variable	Variable - timepoint	Variable – data used	ICE	Censoring rules	Summary measure
Cancer patients	Active vs control to disease progression **then any treatment**	Time to 1st deterioration in **PRO score, or disease progression or or death**	**Until patients die**	All PRO data, dates of disease progression, death	All included as events	No baseline PRO – censor day 1 No post-baseline PRO – censor day 1 No event – censor last PRO assessment	HR

In cancer patients, what is the difference in time to PRO score deterioration, disease progression, or death between treatment X compared to treatment Y (and any subsequent therapy)?"

> *This is a composite approach for all ICEs (censor only for when missing event data)*
> *THINK – is this the scientific question you are really interested in/relevant??*
> *In practice how can you collect this data till death – is it possible now/in the future?*

“Backfitting” – Time to deterioration

Considerations

- What is the definition of an event – is it decline in PRO score alone or a composite of PRO score decline and death? Is the decline in PRO score used appropriate?
  - What about composite include disease progression, treatment discontinuation, cross-over also?
- Censoring – censor at an event (like treatment discontinuation, or disease progression) or last PRO score? - are assumptions about censoring equally valid in the case of death or no data for other reasons?
- Ensure that the definition of an event is completely transparent – it may vary between studies/between treatments and makes comparing results across studies challenging
- Does a change in score have to be confirmed at a later timepoint – if so what about if no further PRO data available – due to death or due to other reasons?
- Is a deterioration expected in disease setting? Is it clinically meaningful to interpret? Are there enough timepoints for a comparative analysis?
- For certain PRO domains/scores there may be low QoL at baseline or a symptom score not impacted by treatment – therefore it is possible that not all patients will experience a decline in all PRO domains/scores – is a survival analysis the most appropriate approach for these PRO domains/scores?

Conclusions

- Not standard use, nor exact definitions of, QoL endpoints across healthcare industry
- Lack of clarity of the precise scientific question targeted for PRO analysis
- It is very difficult to “backfit” to an estimand statement for “typical” PRO analysis – generally much more detail is required
- Among the task force members basic “assumptions” differ – generally due to lack of clarity of the scientific question of interest – which makes “backfitting” challenging if not enough is known.
- MMRM – this is a hypothetical strategy for all ICEs – death or treatment discontinuation
- Time to deterioration – may be composite for death (if death or deterioration in score is an event) – otherwise other ICE usually lead to censoring (and therefore potentially informative censoring)
- How to handle death? – isn’t handled in any special way in MMRM; review definition use for time to deterioration
- Recommendation: be completely transparent in all aspects of analysis methods used to enable clarity of the exact question that is being addressed – don’t assume can compare to other studies easily