

PRESENTED BY ELS GOETGHEBEUR

SENSITIVITY ANALYSES FOR MISSING DATA IN OBSERVATIONAL STUDIES: A PRACTICAL GUIDE UNDER MAR & BEYOND

for planning, conducting and reporting

THE (CO)AUTHORS

- Dries Reynders (UGent)
- Rheanna M. Mainzer (UMelbourne)
- Saskia le Cessie (LUMC)
- James R. Carpenter (MRC CT and LSHTM, London)
- Els Goetghebeur (UGent)
- Katherine J. Lee (UMelbourne)

on behalf of TG1 and TG7 of the STRATOS initiative

TG1-TG7 COLLABORATION

Missing data (TG1) and Causal effect estimation (TG7)

2 highly prevalent conditions

- TG1 several papers* on MNAR the general outcome and estimand
- TG7 -> specific causal estimand (MAR so far**), typically a contrast between treatment specific standardized means or survival (curves, restricted means, possibly hazard ratios)
 - IMI SISAQOL: missingness post disease progression
 - EORTC asks for reasons for missingness in RCTs

* Cro et al. '20,'20; Jackson et al. '14; Atkinson et al. '19, *Lee et al. '21, Mainzer et al. '26?*

** Thomassen et al. '25; Reynders et al.'26

KATE LEE AND TG1

Practical guide for sensitivity analyses

1. Planning the analysis

1. Which deviations from the assumptions made for the primary analysis need to be considered? i.e. what sensitivity analyses are required?
2. Which method of estimation?
3. How to choose the missing values/sensitivity parameter
4. Writing the analysis plan

2. Conducting the analysis – follow the analysis plan

3. Reporting the sensitivity analysis

1. Describe the missing data
2. Presenting the results (graphically)
3. Interpreting the results – how results differ to the primary results, which is the most believable

WHEN INTERESTED IN TREATMENT CONTRAST:

- **Treatment specific expected potential outcomes** are key components and of interest (in their own right)

We focus on missing *outcomes* in a case study with

- Outcomes:
 - Time to death (PFS) and
 - Repeated QoL measures while alive
- Estimands (difference in treatment specific):
 - Potential survival curves and
 - Expected QoL while alive over time

MAR MEANS

A1: $P(y \mid \text{'observed data' and } y \text{ missing}) = P(y \mid \text{'observed data' and } y \text{ not missing})$

A2: $P(y \text{ missing} \mid \text{'observed data' and } y) = P(y \text{ missing} \mid \text{'observed data'})$

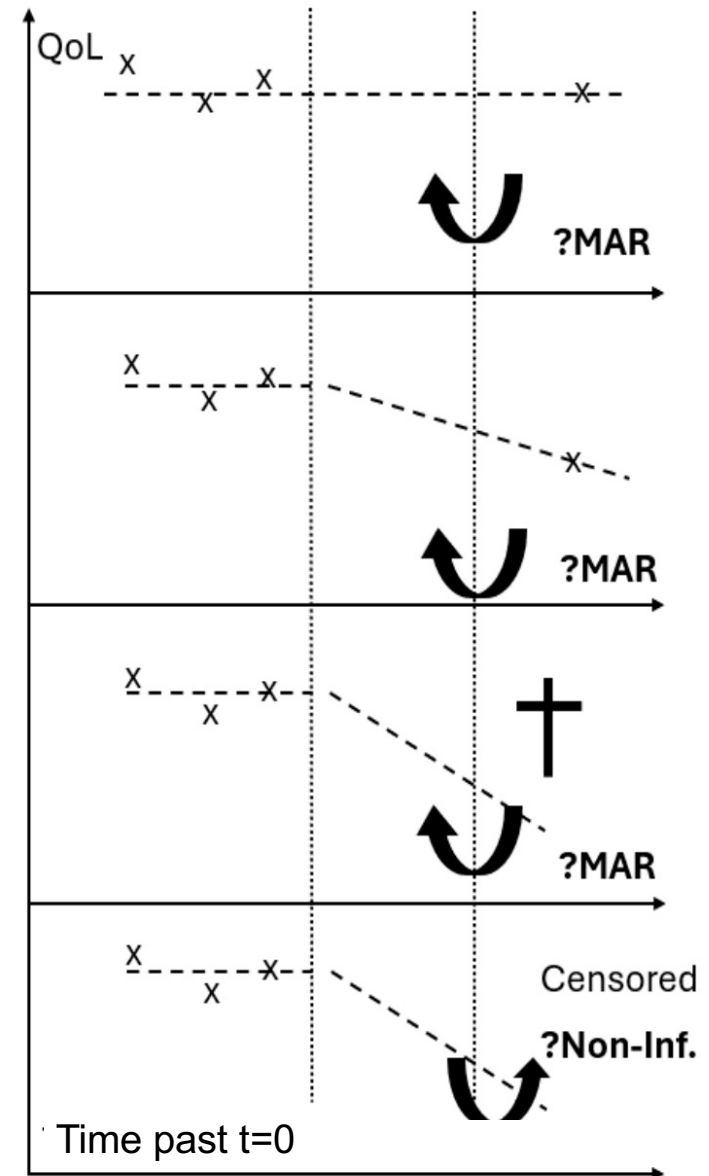
Imputation model and analysis model can be separated (Cro '20)

(With attrition/censoring, observed data come before the current data as with models built to account for confounding)

MAR is untestable, hopefully plausible, the observed pattern
(with extra insight) is suggestive

SUGGESTIVE PATTERNS

- Intermittent (stationary setting)
- Intermittent (changing setting)
- Censoring with (some) clinical FUP
- Censoring without further FUP



POSSIBLE LEAD-IN PATTERNS FOR CENSORING STRUCTURALLY OR OTHERWISE

- Downward cognitive trend
- Downward trend of QoL
- Disease progression
- Treatment nonadherence/switching
- ‘When more is measured’ - > MAR more likely

... BUT POSITIVITY NEEDED FOR BORROWING

If $\{QoL(t) < b\}$ does *not enforce* censoring time, and we have MAR given QoL(t) history,
We can solve the estimation problem under **positivity** with:

A propensity for missingness

– $P(M(t) | QoL(t)) = \pi_{(c)t}$ -> estimate for inverse probability of missing/censoring weight

and/or a conditional outcome regression model given not missing

- $P(BP(t) | QoL(t), M(t)) = P(BP(t) | QoL(t), !M(t))$ -> estimate and impute from

MAR + non-missing conditional alternatives to upweight or **impute from**

MISSINGNESS OF 'QOL' OFTEN FOLLOWS

- ✓ Qol trend
- ✓ (treatment) nonadherence
- ✓ treatment switching (esp. in RCT)
- ✓ other intercurrent events
 - ✓ Disease progression
 - ✓ Death?
- ✓ Hormone cycle adapted

- ✓ Recorded Reasons for missingness (too ill, travel...)
 - per definition **no positivity**
 - But informs MNAR direction for sensitivity analysis

MISSINGNESS ASSUMPTIONS BEYOND MAR

- ❑ Due to
 - ✓ treatment inefficacy or withdrawal of consent
 - > jump to `reference' approach (Cro et al. 2020)
 - ✓ for other reasons: **neutral**, **good** or **bad** news
 - > **MAR**, and **plus** Δ or **minus** Δ
- ❑ Reference-based imputation strategies (Cro et al. 2024)
 - `follow' post-trt-dropout circumstances
 - `Aligns' with what caused the missingness...

RECORDED REASONS FOR MISSING

Meaning of 'Too ill'

- not different from other missing
 - ⇒ same M(N)AR for all
- indicator of acute condition
 - ⇒ 'Too Ill' worse at that time
- indicator of broader condition
 - ⇒ 'Ever Too Ill' worse

In our Case Study, After 'Too ill' we observe:

- More missingness in the next QoL
- Higher hazard of (PF)S

THE DELTA METHOD FOR

- QoL: impute MAR and subtract Δ from (Cro et al., 2020)

$$E(Y_{ij,miss}) = \underbrace{E(Y_{ij} | \text{Observed}_i)}_{MAR} - \delta_{Qij}$$

- Survival: impute residual time assuming shift in hazard after censoring (Jackson et al. 2014)

$$h(t | C_i < t, Z_i, S_i) = h_{0,S_i}(t) \exp(\beta \cdot Z_i) \exp(\delta_{S_i})$$

DIFFERENTIAL M(N)AR OVER TREATMENTS

- ❑ If the treatment shows more toxicity, possibly correlated with general sensitivity to the drug, leading to more missing data; differential DELTA may be needed between treatments
- ❑ If the causal estimand compares **standardized means**, for correct x-specific means in each treatment arm -> x-specific delta per arm; or if 'old' people quicker to lapse...
- ❑ Bivariate outcome focus may require linked Delta's

EORTC-26981 (TAPHOORN, 2005)

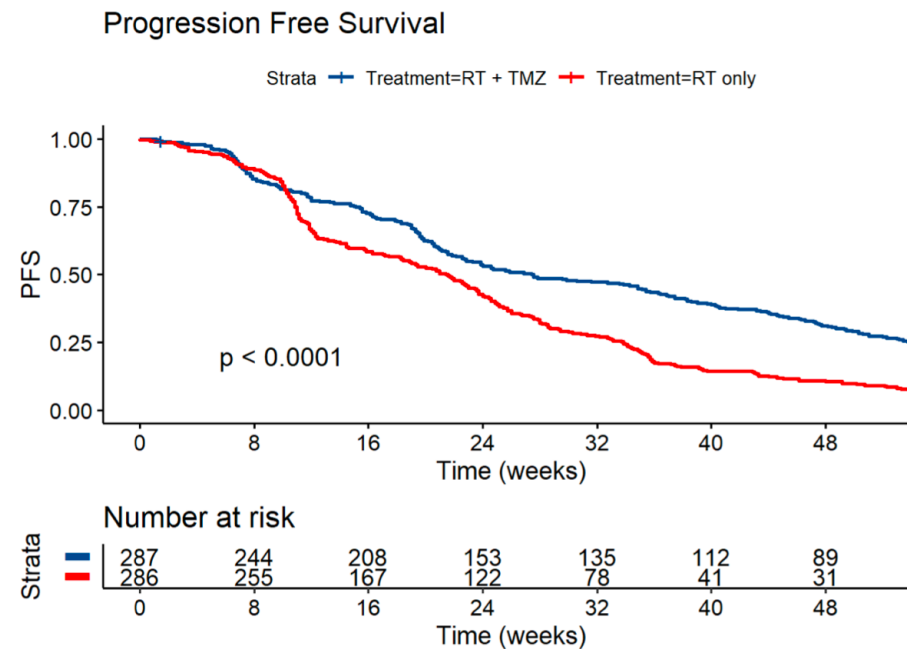
- HRQoL was measured at baseline (week 0), during radiotherapy (week 4), 4 weeks after the end of radiotherapy (week 10) and then at 12-week intervals until progression.
(picture..?)
- Reasons of missingness from the list:
 - ‘Too ill’
 - ‘Inconvenient’
 - ‘Didn’t understand’
 - ‘Administrative Failure’, ‘Unknown’, ‘Other’,...

ESTIMANDS – OUR AIM

- The estimand of interest is
 - overall survival and
 - HRQoL while-alive, contrasting mean HRQoL among survivors at each visit
- **Protocol:** stop QOL-measurement at progression
- Still: some provided it ... once (by design)
- For while-alive estimand this is not the best design, but often unavoidable loss of FUP there
- we target this most relevant endpoint

EORTC GLIOBLASTOMA RCT (N=573)

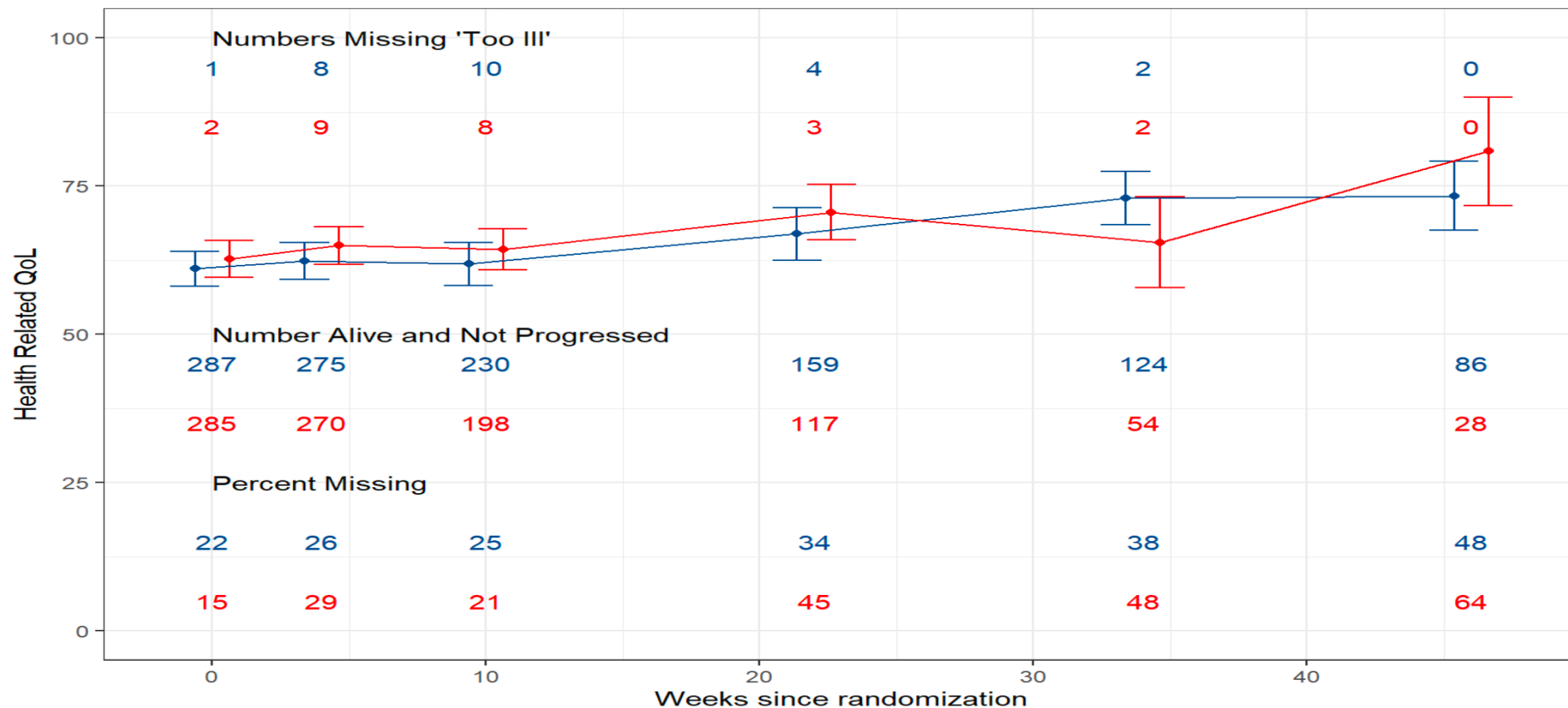
Radiotherapy vs. Radiotherapy + Temozolomide¹



- By design no QoL measured after Progression

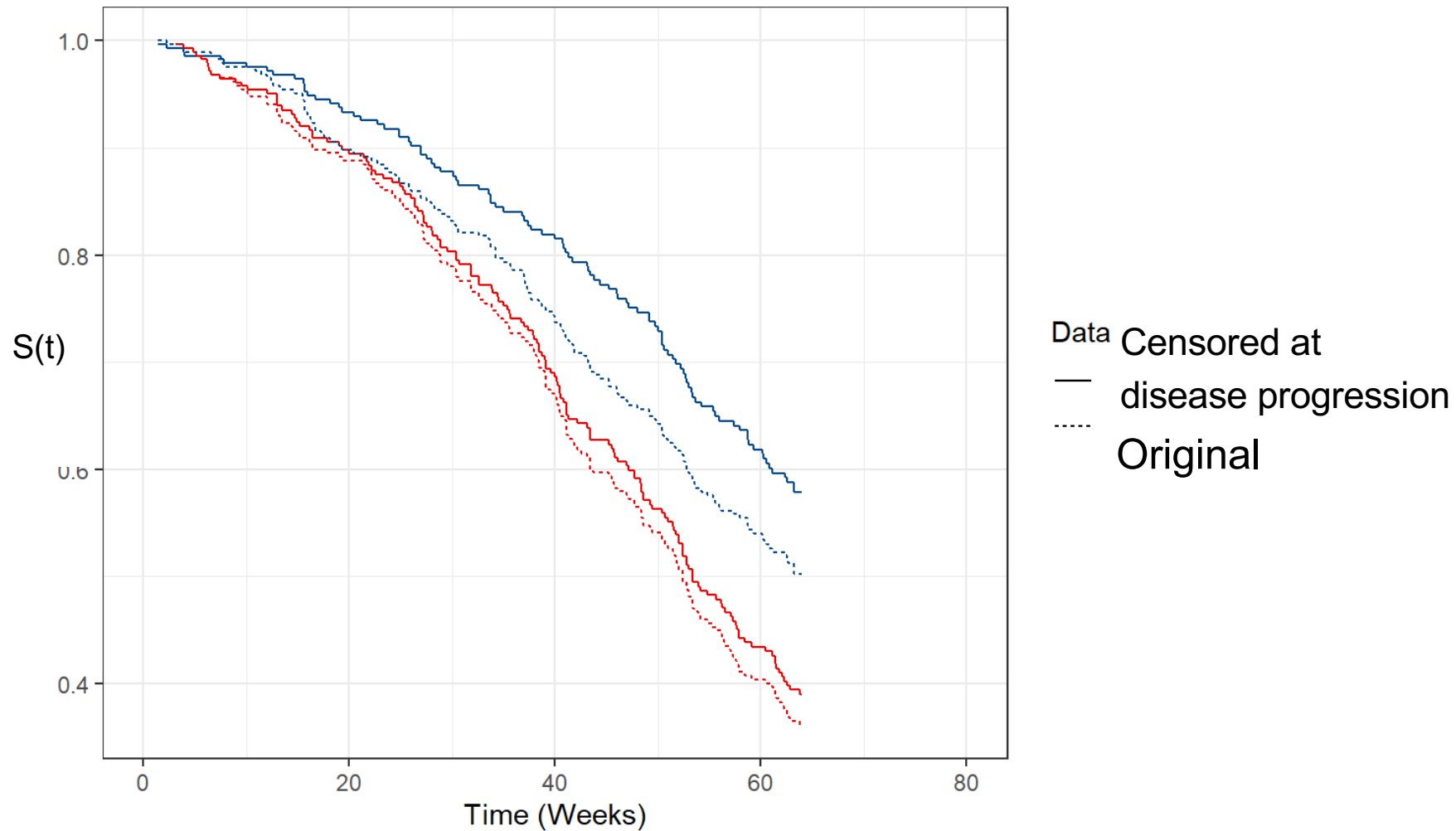
¹Stupp et al. (2005) NEJM - Taphoorn et al. (2005) Lancet Oncology

Health Related Quality of Life RT + Temo vs RT only

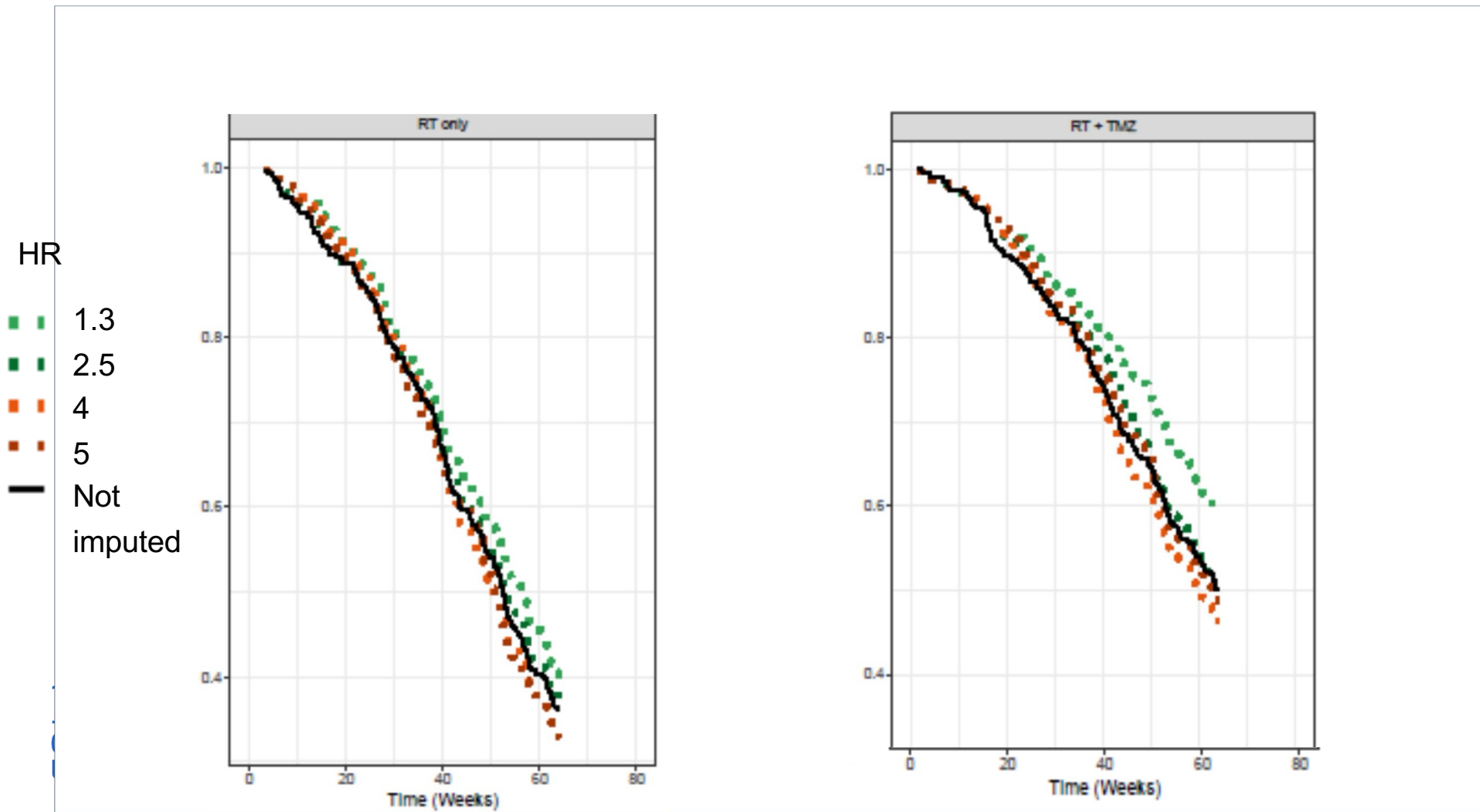


OS with/without censoring at disease progression

RT + Temo vs RT only



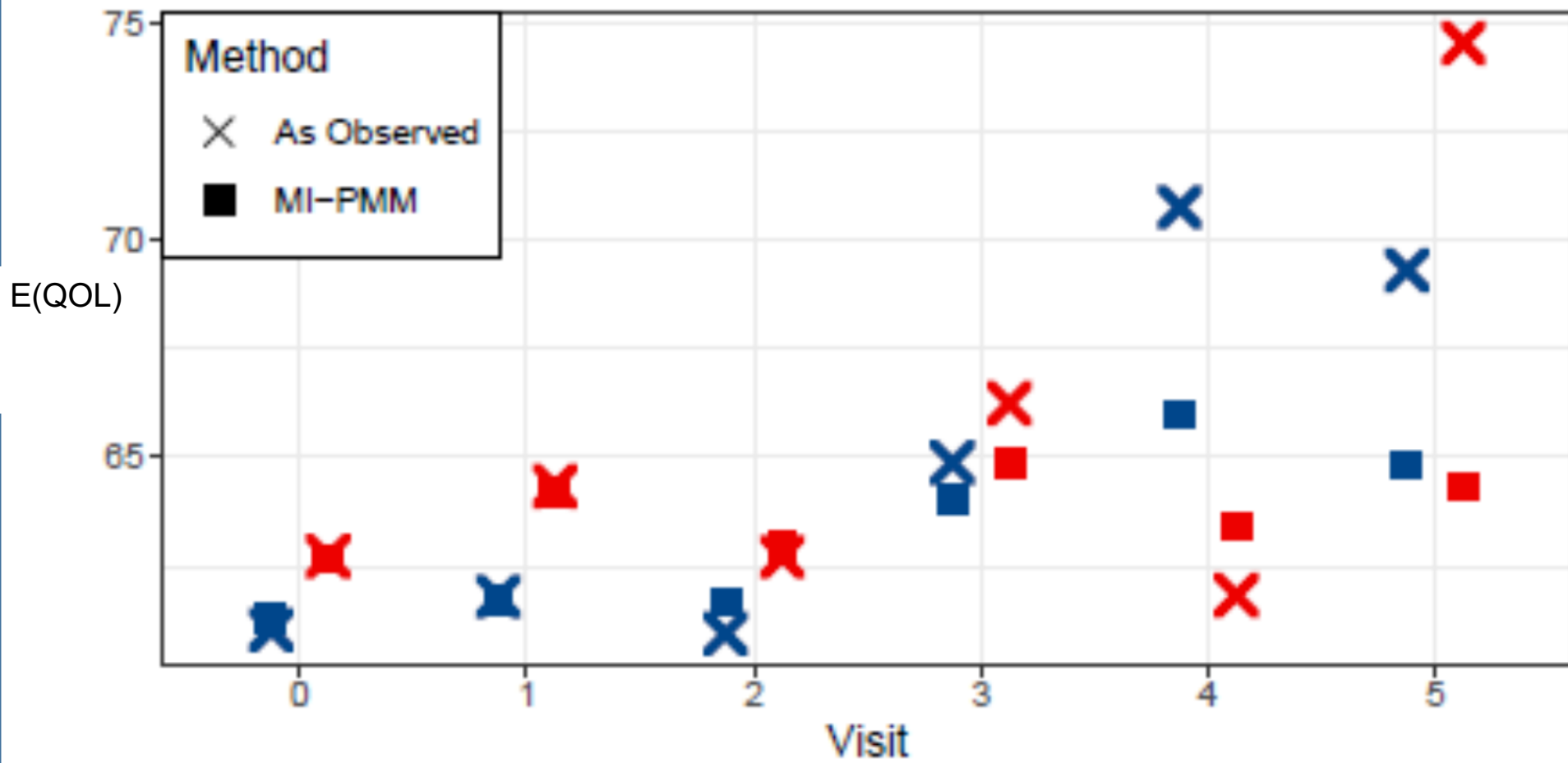
K-M AFTER Δ IMPUTATION OF ALL CENSORED



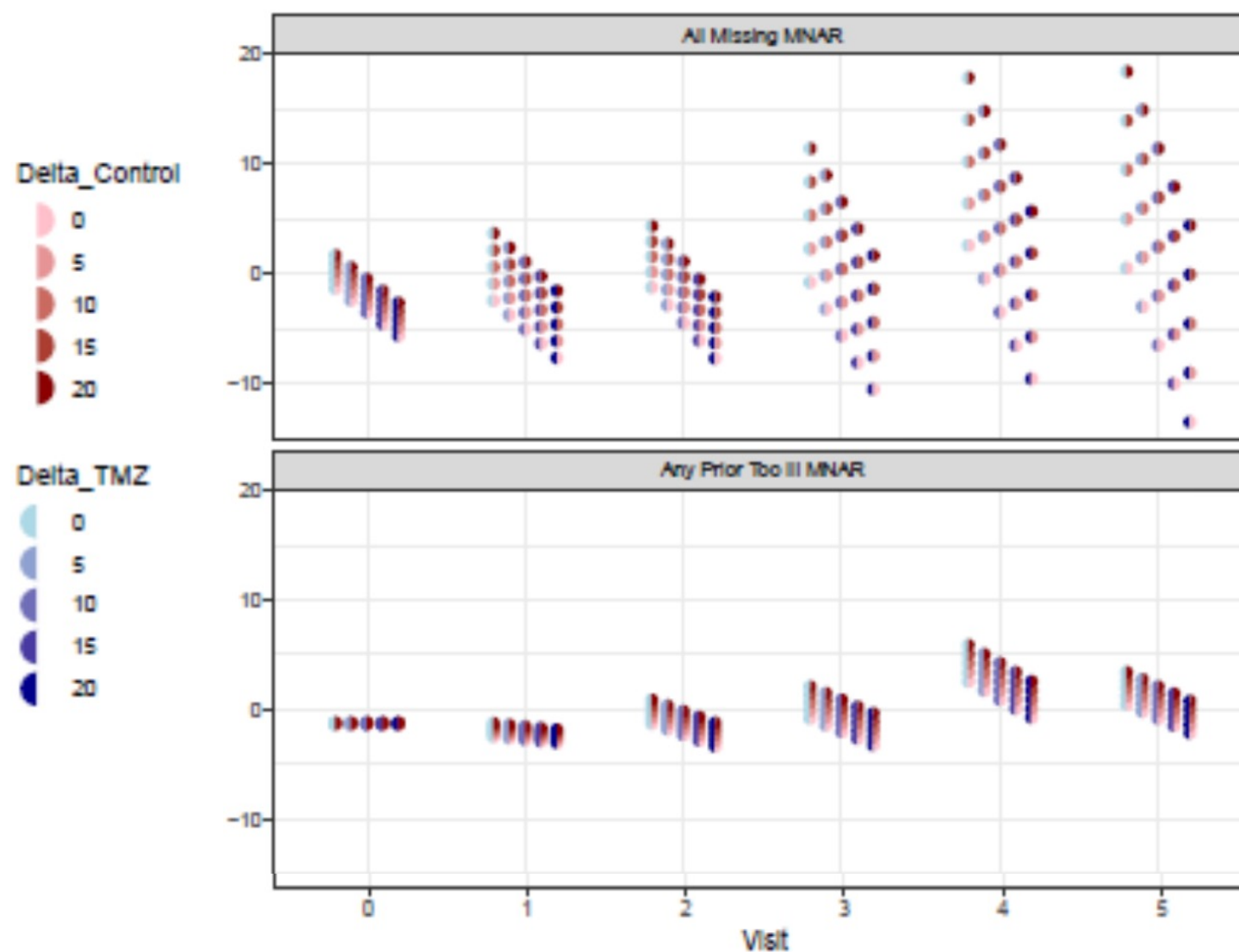
Average QoL While not Progressed

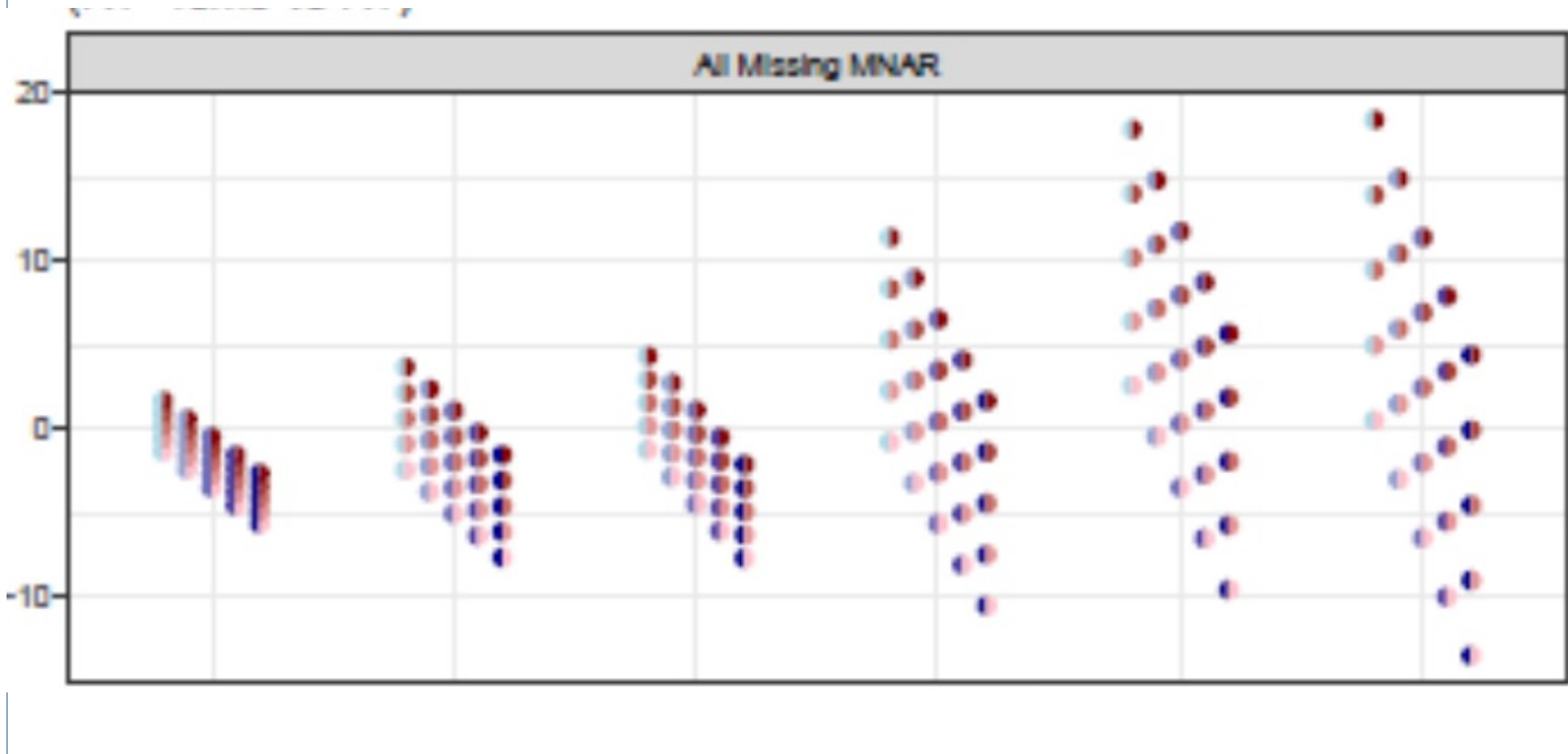
Observed/MAR Imputed

RT + Temo vs RT only

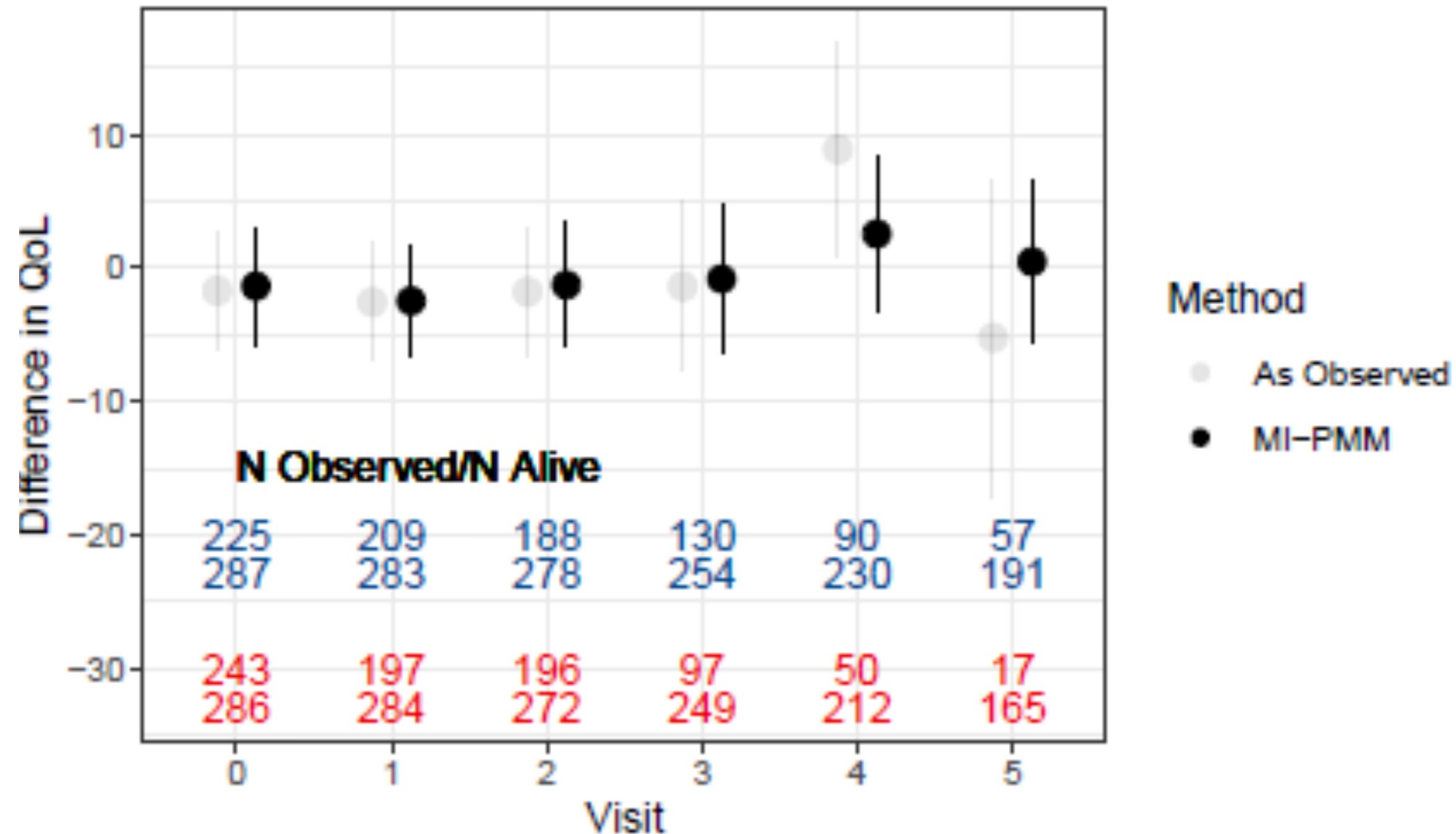


DELTA -ADJUSTED MI FOR QOL WHILE ALIVE



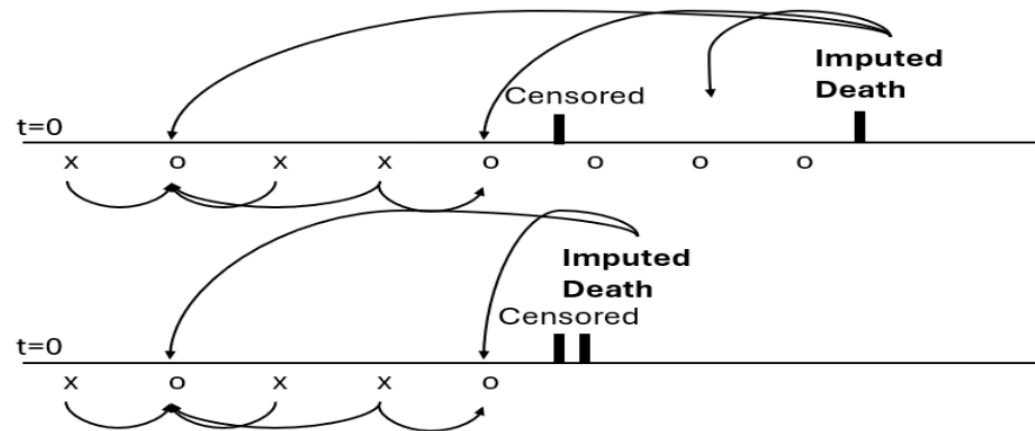


MEAN QOL(CI) DIFFERENCE (BEFORE PFS)



TWO-DIMENSIONAL OUTCOME & BIVARIATE Δ

- QoL (While Alive) as two-dimensional outcome
 - QoL and Survival inherently tied together
 - QoL only exists while Alive
- Imputation of one component impacts the other
 - Joint Imputation
 - Joint Bivariate Delta (? Relevant ranges)



DISCUSSION

- ‘All missing MNAR’ versus ‘Too Ill MNAR’: big change
- ‘Reasons informed’ scenarios can bring important input to the sensitivity analyses.
- Bigger impact on treatment-specific summaries than on treatment effects -> both matter
 - Eliciting delta/reference is work (Lee et al.)
 - Information-anchored analysis matters (Carpenter et al)

DISCUSSION

- Sensitivity analysis is critical part of the causal roadmap
- Similar efforts for confounding assumption and censoring (Informative Censoring IPCW with time-varying covariates)
- E-value – ignorance and uncertainty region?
- To do: Summary and Visualization acknowledging jointly **Ignorance** (in the Delta's) and **Uncertainty** (sampling)*
 - MI software and bootstrap can help there
 - Mainzer et al. upcoming

* Stijn Vansteelandt et al. (2006); S. Vansteelandt and Goetghebeur (2005)

THERE IS LOTS MORE TO DISCOVER

Good luck to the TG1-TG7 collaboration!

SOME REFERENCES

