A practical toolkit with recommendations for analysing and visualising patient-reported outcomes in early phase dose-finding oncology trials: OPTIMISE-AR

Appendix

1 Literature review

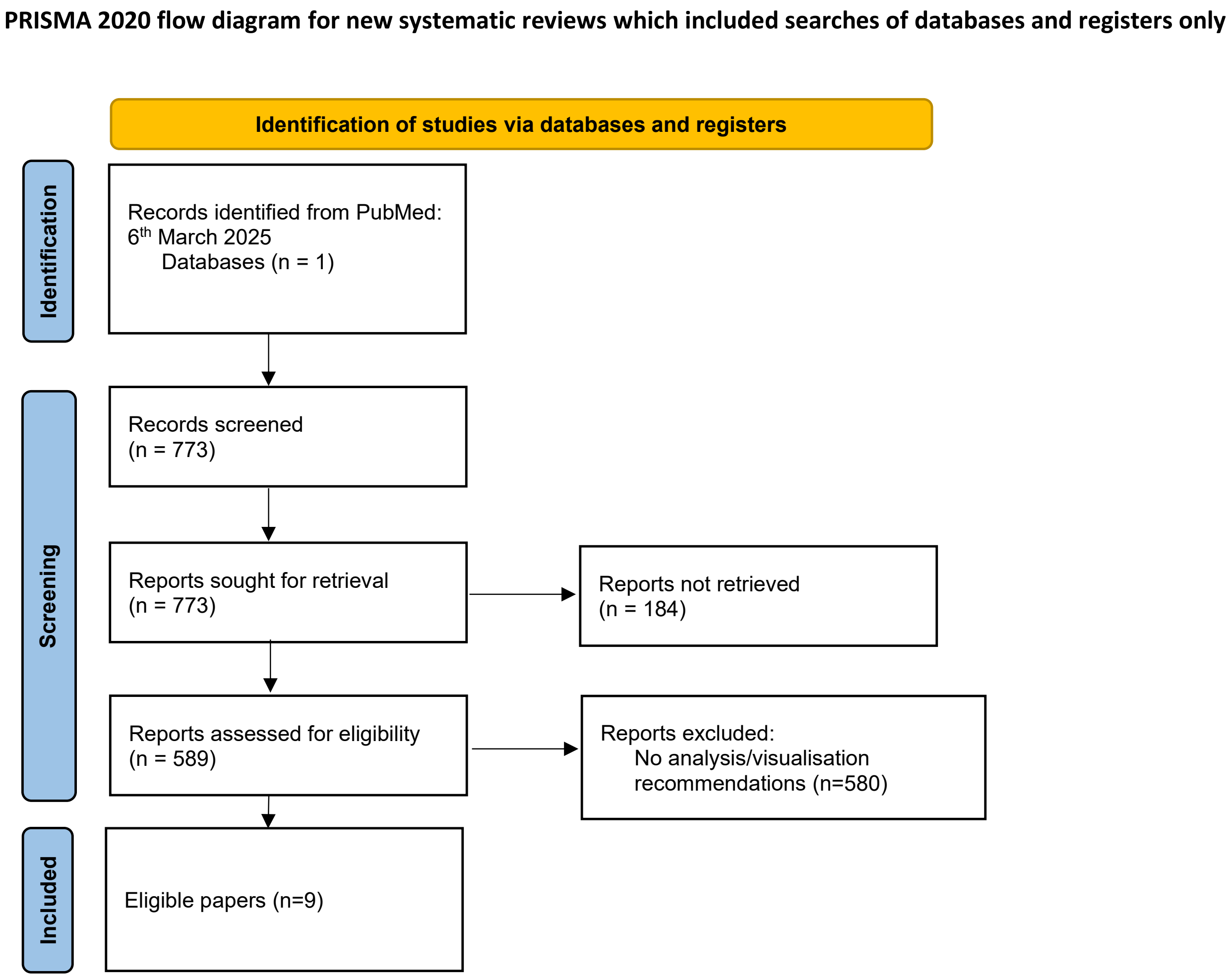

Figure S1: PRISMA study flow diagram illustrating selection of eligible studies.[1–9]

2 Data synthesis for presentation of results

Full details of the synthesis of data for the presentation of results in this paper are available in the GitHub respository: `https://github.com/alemily100/optimise-ar`.

2.1 Ordinal and continuous PRO data

This section details the simulation of PRO data relevant to the case study, (1) PRO endpoints across timepoints, (2) PRO endpoints between timepoints, and (3) Time-to-event PRO endpoints.
Data are simulated using a "while-on-treatment" approach. Both patient death and failure to report PROs are simulated in each dataset. Non-treatment related death is treated as an intercurrent event and thus represents an event after the initiation of treatment that affects the interpretation of the outcome. Patients experiencing this event are excluded from the analysis onwards of this timepoint. In contrast, missing data refers to unreported outcome scores, either due to incomplete patient reporting or data entry omissions, and is handled separately in the analysis.

Time to non-treatment related death

Time to non-treatment related death T is simulated using an exponential distribution,

$$T \sim \exp(\lambda). \tag{1}$$

$\lambda = \frac{\log(2)}{36}$ to ensure median survival is approximately 9 months.

Missing PRO report

Missing PRO report is simulated as per a Bernoulli distribution, such that a PRO report is missing with probability $p = 0.05$.

Synthesis of continuous PRO data

The PRO score $\tilde{y}_{i,f,j}$ for patient i, for functioning f at timepoint j is simulated such that,

$$\tilde{y}_{i,f,j} = b_{i,f,0} + b_{i,f,j} + b_{i,f}. \tag{2}$$

For functioning f, patient i's random effect $b_{i,f}$ is sampled as:

$$b_{i,f} \sim N(0, \sigma^2), \tag{3}$$

where σ^2 determines between-patient heterogeneity.
For functioning f, patient i's baseline score $b_{i,f,0}$ is sampled as:

$$b_{i,f,0} \sim N(\mu_{f,0}, \sigma^2_{f,0}) \text{ truncated between } [0, 100], \tag{4}$$

where $\mu_{f,0}$ and $\sigma^2_{f,0}$ represent baseline PRO score mean and variance for functioning f.
For functioning f at timepoint j, patient i's mean change from baseline is sampled as:

$$b_{i,f,j} \sim N(\mu_{f,j} + d_{[i]}, \sigma^2_{f,j}), \tag{5}$$

where $\mu_{f,j}$ and $\sigma^2_{f,j}$ represent mean and variance of change from baseline for functioning f at timepoint j and $d_{[i]}$ represents the effect of the dose on the mean change from baseline for patient i assigned to dose $d_{[i]}$.

Synthesis of ordinal PRO data

The ordinal patient score $h \in \{1, ..., H\}$ for each item at time j and dose k is sampled as per,

$$Y_{\text{item},j,k} \sim \text{ Categorical}(p_{\text{item},j,k,1}, ..., p_{\text{item},j,k,H}). \tag{6}$$

For $F(\cdot\,; \alpha, \beta)$ denoting a cumulative Beta distribution function, the probability of experiencing each grade h for each specific item at time j and dose k can be written as,

$$p_{\text{item},j,k,h} = \begin{cases} F(C_{\text{item},h}\,; \kappa_k, \tau_j), & \text{if } h = 1 \\ F(C_{\text{item},h}\,; \kappa_k, \tau_j) - F(C_{\text{item},h-1}\,; \kappa_k, \tau_j), & \text{if } 1 < h < H\ . \\ 1 - F(C_{\text{item},H-1}\,; \kappa_k, \tau_j), & \text{if } h = H \end{cases} \tag{7}$$

For $C^*_{\text{item},h} \sim$ Uniform$[0, 1]$ for $1 \leq h < H$ and $C^*_{\text{item}} = \{C^*_{\text{item},1}, ..., C^*_{\text{item},H}\}$, then $C_{\text{item},h}$ is sampled such that $C_{\text{item},h} = C_{\text{item},(h)}$ which is the hth-ordered sample of C^*_{item}.
Values of $\boldsymbol{\kappa}$ and $\boldsymbol{\tau}$ are chosen to ensure the marginal distribution of $Y_{\text{item},j,k}$ approximately align with aggregated Phase I clinical trial data collected by Princess Margaret Cancer Centre in Canada from May 1, 2017 to January 1, 2019.[10]
Exact parameter values used to simulate continuous and ordinal PRO data are indicated in `https://github.com/alemily100/optimise-ar`.

2.2 DLT data synthesis

This section details the simulation of PRO data relevant to (4) PRO endpoints embedded within a model-based dose-finding design for formal decision-making.
Correlated C-DLT and P-DLT binary outcomes are simulated utilising a Clayton copula where ϕ defines the correlation between the two outcomes with with $\phi \to 0$ inducing a strong positive correlation, and $\phi \to \infty$ inducing no correlation.
Suppose $S_C(t) = \exp(-\lambda_C t)$ and $S_P(t) = \exp(-\lambda_P t)$ are the survival functions for C-DLT and P-DLT respectively with λ_C and λ_P scalars. Time to Clinician-DLT (T_C) and Patient-DLT (T_P) are simulated as follows:

1. Simulate $u_1 \sim \text{Unif}(0, 1)$,
2. $T_C = \frac{\log(u_1)}{-\lambda_C}$,
3. Simulate $u_2 \sim \text{Unif}(0, 1)$,
4. $T_P = \frac{\phi}{\lambda_P} \log\left(a^{\frac{1}{-\phi-1}} - b\right)$, where $a = \frac{u_2}{u_1^{-\left(\frac{\phi+1}{\phi}\right)}}$ and $b = u_1^{-\frac{1}{\phi}} - 1$.

C-DLT and P-DLT outcomes are moderately correlated, with $\phi = 0.9$ (equivalent to Kendall's Tau of 0.31). For p_C and p_P the probability of C-DLT and P-DLT at a specific dose, $\lambda_C = -\log(1 - p_C)$ and $\lambda_P = -\log(1 - p_P)$ respectively.

3 Mathematical model for Case Study: Analysis of continuous PRO data at the final analysis

The mathematical model to analyse mean change from baseline of the EORTC QLQ-C30 score can be written as,

$$y_{i,j} = \beta_0 + \beta_1 t_j + \sum_{k=2}^{3} \beta_k \mathbb{1}(x_i = k) + \beta_4 b_i + r_{0,i} + \epsilon_{i,j}. \tag{8}$$

For patient i at timepoint j, change from baseline $y_{i,j}$ is modelled as a function of,

- intercept β_0,
- timepoint j (continuous),
- patient dose x_i (categorical with dose 1 as the reference dose),
- baseline PRO score b_i (continuous),
- normally distributed random intercept $r_{0,j}$,
- normally distributed error terms $\epsilon_{i,j}$.

4 Linear mixed-effects model results for the Case study

Covariate	Estimate	Standard Error	95% Confidence Interval
Intercept	-0.19	1.87	[-3.85, 3.47]
Timepoint	-0.50	0.12	[-0.73, -0.27]
Dose 2	-5.11	1.12	[-7.29, -2.92]
Dose 3	-10.80	1.11	[-13.00, -8.60]
Baseline score	0.07	0.02	[0.02, 0.11]

Table S1: Linear mixed-model analysis of mean change in EORTC QLQ-C30 Global health status from baseline until week 7 adjusted for dose, time, baseline PRO score and repeated patient measurements. Score range: 0-100, higher scores indicate improved global health. Dose discontinuation is an intercurrent event excluding patients from analysis. Missing data excludes patients with unreported scores due to incomplete patient reporting or data entry errors. Dose 1 is used as the reference category for comparisons in the analysis.

5 Statistical modelling approaches

5.1 PRO endpoints across timepoints

Statistical Modelling Approach	**Brief description**	**Key considerations for implementation and analysis in early-phase dose-finding trials**		**Exemplar use in published literature**	**Additional comments**
		Strengths	**Limitations**		
(Generalised) linear mixed-effects model (GLMMs)	Linear mixed effect models are likelihood-based models allowing for model estimation in the presence of fixed effects (e.g., dose) and random effects (e.g., patient-specific correlated observations over time) for continuous outcome data that are approximately normally distributed. Generalised linear mixed effects models extend this framework to handle various PRO data types and outcome distributions, including binary and ordinal outcomes.	(Generalised) linear mixed-effects models are a well-established[7]and flexible approach for PRO data analysis, accommodating binary, count, categorical and continuous data. Unlike some methods, these models can handle missing completely at random (MCAR) and missing at random (MAR) data and accommodate analysis for data collected at irregular timepoints within trials. They account for within-patient variation – considering the correlation of repeated measures provided by the same patient. These models allow for formal covariate adjustment.	This model is not suited for analysis in presence of missing not at random (MNAR) data. To ensure valid inference, careful consideration of the structure of the random-effect covariance matrix is needed. To ensure this parametric model fits the data, goodness-of-fit checks are advised.	Within a Phase I setting, linear mixed effect modelling has been utilised to evaluate on-treatment changes from baseline for patient-reported EQ-5D-3L QoL and WOMAC questionnaires within a dose-finding oncology trial.[11] Inferential analysis was provided solely for descriptive rather than confirmatory purposes.	
Area under the operating Curve (AUC)	AUC summarises longitudinal patient reported data into a cumulative summary of toxicity experienced by patients whilst on treatment.	Summarising longitudinal data as a cumulative toxicity for each dose is simple and may support clinical interpretability.[12]	It is not generally recommended to complete an AUC analysis in the presence of missing data.[12] The nature of the method requires missing data be imputed to ensure the AUC can be calculated for the same time interval for each patient. Summarising patient repeated measurements within a cumulative score reduces the dimensionality of data, perhaps obscuring potential longitudinal patterns.	Within the Phase II setting, AUC analysis has been utilised to quantify cumulative toxicity over time for two treatment arms for the PRO-CTCAE item constipation severity.[3]	

Continued on next page

Table S2 (continued)

Statistical Modelling Approach	**Brief description**	**Key considerations for implementation and analysis in early-phase dose-finding trials**		**Exemplar use in published literature**	**Additional comments**
		Strengths	**Limitations**		
Repeated-measures ANOVA	This method extends the traditional ANOVA to allow for correlated observations captured over multiple timepoints.[13]	It accounts for within-patient and between group (dose) variation. Unlike the traditional ANOVA, the repeated-measures ANOVA incorporates data from all timepoints in the analysis.	This method requires that PRO data be continuous and approximately normally distributed. Treating ordinal PRO data as continuous may violate this assumption. This analysis approach is also limited in handling missing data – it generally assumes data are MCAR and requires an equal number of repeated measurements for each dose. As time is considered a categorical factor, this method is less flexible for analyses involving irregular or unequally spaced timepoints.	There are extensions of ANOVA to better suit certain PRO research objectives, such as multivariate ANOVA for testing across dependent, multivariate toxicity outcomes. Additionally, ANCOVA may be considered for trialists wishing to adjust for patient covariates.	

Continued on next page

Table S2 (continued)

Statistical Modelling Approach	Brief description	Key considerations for implementation and analysis in early-phase dose-finding trials		Exemplar use in published literature	Additional comments
		Strengths	**Limitations**		
Generalised Estimating Equations (GEEs)	Generalised Estimating Equations are a non-likelihood-based approach for modelling correlated observations.	GEE can be utilised for ordinal ratings (e.g., PRO-CTCAE items). Similarly to linear mixed effect models (LMMs), GEEs can model within-patient variation. However, unlike LMMs, GEEs are robust to misspecification of the correlation structure, provided the mean model is correctly specified.[14]	GEEs provide a population average interpretation which provides less information about specific individual changes over time than LMMs. GEEs only provide unbiased estimates for MCAR data.	In a Phase II setting, a GEE model has been utilised to compare PRO-CTCAE ratings over time for two dose groups with point estimates and standard errors.[3]	Both (generalised) linear mixed effect modelling and GEEs may be extended to ordinal PRO data using proportional odds modelling, providing odds ratios for experiencing ordinal events between doses.[3] Proportional odds models,[15] including mixed-effects proportional odds models[16] have been implemented within dose-finding trials for clinician-assessed toxicity endpoints. Choosing between GLMMs and GEEs may depend on the research objectives of interest – GLMMs can provide subject-specific inference whilst GEEs provide population average effects. Whilst GLMMs are subject specific model and generally more efficient than GEEs, GEEs are more robust to the misspecification of the random-effect correlation structure.

Continued on next page

Table S2 (continued)

Statistical Modelling Approach	**Brief description**	**Key considerations for implementation and analysis in early-phase dose-finding trials**		**Exemplar use in published literature**	**Additional comments**
		Strengths	**Limitations**		
Ordinal log-linear models (OLLMs)	Ordinal log-linear models analyse the correlation of variables within contingency tables, providing odds-ratio estimates for comparisons.	OLLMs can be utilised for ordinal data where intervals between ordinal categories do not need to be equal. The odds-ratio output of the model aids the clinical interpretation of between-group differences.	Not all statistical software provides packages for OLLM analysis. For convergence, sufficient data and sample size is required for analysis.	Within the Phase II setting, a OLLM model has been utilised to compare maximum postbaseline PRO-CTCAE ratings adjusted for baseline.[3]	
Ordinal transition models	Ordinal transition models provide an alternative to proportional odds modelling for ordinal PRO data by modelling the transitions between states (ordinal PRO scores) over time.	Ordinal transition models can allow for formal covariate adjustment and may provide clinically interpretable estimands without need for a proportional odds assumption.	This analysis method may not be as well-known as other longitudinal analysis techniques, with limited real-world applications to trials.[17]	There are currently no examples of ordinal transition modelling for PROs, although the methods application to COVID-19 trials has been investigated.[17]	

Table S2: Strengths and limitations of statistical modelling methods for the analysis of PROs in DFOTs across timepoints. Methods are listed in order of most frequent use, followed by additional methods that may be appropriate depending on the trial setting.

5.2 PRO endpoints between timepoints

Statistical Modelling Approach	**Brief description**	**Key considerations for implementation and analysis in early-phase dose-finding trials**		**Exemplar use in published literature**
		Strengths	**Limitations**	
(Generalised) linear models (GLMs)	Linear models estimate the relationship between a normally distributed response variable and covariates. Linear models can be extended to GLMs to allow for PRO response data to be of other data types.	GLMs provide a simpler alternative to GLMMs when comparisons between tolerability are only required between two time-points.[1]	GLMs do not support analysis across timepoints and are best suited to cross-sectional or study-defined fixed timepoint endpoints.	Within the Phase I setting, GLMs have been utilised to compare baseline-adjusted PRO scores at month 3 for two doses.[18]
Regression trees	Regression trees predict outcomes using single decision trees – partitioning data at decision nodes.[19,20]	Regression trees are well suited to analyse non-linear relationships and dependencies within PRO data. They may also be suited to analysis for small sample sizes.[20]	Regression trees may yield models that are difficult to interpret. Whilst this modelling approach may have good predictive accuracy, it may be challenging to evaluate the relationship between the response variable and specific covariates.	Bayesian additive regression trees have been implemented within dose-optimisation designs to estimate the probability of patient-DLT using prognostic factors.[20] Bayesian additive regression trees can be considered a sum of weak-learning regression trees. Regression trees and other machine learning algorithms have been utilised to predict patient mortality using PROs.[19] Other methods such as random forests may be considered.

Table S3: Strengths and limitations of statistical modelling methods for the analysis of PROs in DFOTs between timepoints. Methods are listed in order of most frequent use, followed by additional methods that may be appropriate depending on the trial setting.

5.3 Time-to-event PRO endpoints

Statistical Modelling Approach	Brief description	Key considerations for implementation and analysis in early-phase dose-finding trials		Exemplar use in published literature
		Strengths	**Limitations**	
Kaplan-Meier	The Kaplan-Meier estimator provides a non-parametric approach to estimating a survival function.	Kaplan-Meier curves may provide the most benefit as a visual comparison between doses for a well-specified event of interest. Kaplan-Meier curves are a familiar survival analysis method and are simple to interpret.	Kaplan-Meiers curves do not provide formal covariate adjustment. Its utilisation for statistical inference is limited to hypothesis testing and estimation (please see section “PROs for exploratory analysis and hypothesis testing” for discussion of hypothesis testing in DFOTs). When censoring is informative (which may occur in the presence of intercurrent events), the Kaplan-Meier estimator may inadequately handle censored data – see Section 4 of the Supplementary Materials.	Within a Phase I setting, Anota et al present Kaplan-Meier graphs detailing time to health-related quality of life score deterioration with a 5-point minimal clinically important worsening for three dose levels across four dimensions of the EORTC QLQ-C30.[21]

Continued on next page

Table S4 (continued)

Statistical Modelling Approach	**Brief description**	**Key considerations for implementation and analysis in early-phase dose-finding trials**		**Exemplar use in published literature**
		Strengths	**Limitations**	
Cox proportional hazards model	Cox proportional hazards models provide a semi-parametric approach to estimating a survival function under a proportional hazards assumption.	This traditional survival analysis is well-known, aiding interpretability of results. The Cox proportional hazards model allows for formal covariate adjustment. This allows the inclusion of time-varying covariates in cases where patient dosage varies over the assessment window.	This class of models supposes a proportional hazards assumption which may not hold given the heterogeneous populations often analysed within early phase trials. The small sample sizes may also limit the opportunity to adequately test for the proportional hazards assumption. In instances of low event rates, robustness of dose-effect estimation may be compromised. Similarly to the Kaplan-Meier estimator, the Cox proportional hazards model may insufficiently handle informative censoring – see Section 4 of the Supplementary Materials.	Within a Phase I setting, Anota et al present the median time to HRQoL score deterioration and hazard ratio with associated confidence intervals at each dose under four definitions of time-to-deterioration.[21] These definitions provide a sensitivity analysis for the analysis. When the proportional hazards assumption does not hold, restricted mean survival time (RMST) method may be used.

Table S4: Strengths and limitations of statistical modelling methods for the analysis of PROs in DFOTs for time-to-event PRO endpoints. Methods are listed in order of most frequent use, followed by additional methods that may be appropriate depending on the trial setting.

6 Strengths and limitations of time-to event statistical modelling approaches

Patient censoring enables handling of missing data when it is either Missing Completely at Random (MCAR) or Missing at Random (MAR).[1] Time-to-event analysis techniques also accommodate varying patient follow-up duration.[3]

Intercurrent events such as death are considered a form of informative censoring and may require special methods to analyse appropriately. This is particularly relevant in DFOT settings, where deaths can be frequent.[22] When sample sizes permit, methods such as cumulative incidence functions and competing-risks methods may account for such intercurrent events.[5] Relevant considerations for applying these methods are addressed in the Discussion section.

Time-to-event PRO endpoints may be most suited for well-defined binary events such as Patient-DLTs as the dichotomising ordinal or continuous PRO data use as a time-to-event endpoints may result in some information loss.[3] However, if an appropriate clinically relevant threshold is available, such analysis may be acceptable given adequate justification.

When analysing a binary event, PRO data that do not define the event of interest (e.g., baseline-adjusted PRO-CTCAE scores smaller than the event of interest) are excluded from the analysis, thus limiting the capability of survival analysis to model longitudinal tolerability profiles of treatment. Methods should be consistent with SISAQOL recommendations,[23] including interval grading and interval censoring.

References

[1] Corneel Coens et al. “International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium”. In: *The Lancet Oncology* 21.2 (2020), e83–e96.

[2] Rachel Phillips et al. “Visualising harms in publications of randomised controlled trials: consensus and recommendations”. In: *bmj* 377 (2022).

[3] Antoine Regnault et al. “A toolbox of different approaches to analyze and present PRO-CTCAE data in oncology studies”. In: *JNCI: Journal of the National Cancer Institute* 115.5 (2023), pp. 586–596.

[4] Katherine E Francis et al. “Reporting the trajectories of adverse events over the entire treatment course in patients with recurrent platinum-sensitive ovarian cancer treated with platinum-based combination chemotherapy regimens: A graphical approach to trial adverse event reporting”. In: *European Journal of Cancer* 148 (2021), pp. 251–259.

[5] Elena Tassistro et al. “Adverse events in single-arm clinical trials with non-fatal time-to-event efficacy endpoint: from clinical questions to methods for statistical analysis”. In: *BMC Medical Research Methodology* 24.1 (2024), p. 3.

[6] Kentaro Sakamaki and Takuya Kawahara. “Statistical methods and graphical displays of quality of life with survival outcomes in oncology clinical trials for supporting the estimand framework”. In: *BMC Medical Research Methodology* 22.1 (2022), p. 259.

[7] Franck Bonnetain et al. *Statistical challenges in the analysis of health-related quality of life in cancer clinical trials.* 2016.

[8] David Collister et al. “Patient reported outcome measures in clinical trials should be initially analyzed as continuous outcomes for statistical significance and responder analyses should be reserved as secondary analyses”. In: *Journal of clinical epidemiology* 134 (2021), pp. 95–102.

[9] Ines Rombach et al. “Comparison of statistical approaches for analyzing incomplete longitudinal patient-reported outcome data in randomized controlled trials”. In: *Patient related outcome measures* (2018), pp. 197–209.

[10] Geoffrey A Watson et al. “Evaluation of the patient experience of symptomatic adverse events on Phase I clinical trials using PRO-CTCAE”. In: *British Journal of Cancer* 127.9 (2022), pp. 1629–1635.

[11] Philippe A Cassier et al. “Long-term clinical activity, safety and patient-reported quality of life for emactuzumab-treated patients with diffuse-type tenosynovial giant-cell tumour”. In: *European Journal of Cancer* 141 (2020), pp. 162–170.

[12] Melanie L Bell, Madeleine T King, and Diane L Fairclough. “Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data: summary measures versus summary statistics”. In: *Sage Open* 4.2 (2014), p. 2158244014534858.

[13] Lutfiyya N Muhammad et al. “Guidelines for repeated measures statistical analysis approaches with basic science research considerations”. In: *The Journal of clinical investigation* 133.11 (2023).

[14] Melanie L Bell et al. “Using generalized estimating equations and extensions in randomized trials with missing longitudinal patient reported outcome data”. In: *Psycho-oncology* 27.9 (2018), pp. 2125–2131.

[15] Emily M Van Meter, Elizabeth Garrett-Mayer, and Dipankar Bandyopadhyay. “Proportional odds model for dose-finding clinical trial designs with ordinal toxicity grading”. In: *Statistics in medicine* 30.17 (2011), pp. 2070–2080.

[16] Adélaıde Doussau, Rodolphe Thiébaut, and Xavier Paoletti. “Dose-finding design using mixed-effect proportional odds model for longitudinal graded toxicity data in phase I oncology clinical trials”. In: *Statistics in medicine* 32.30 (2013), pp. 5430–5447.

[17] Maximilian D Rohde et al. “Bayesian transition models for ordinal longitudinal outcomes”. In: *Statistics in Medicine* 43.18 (2024), pp. 3539–3561.

[18] Pavlos Msaouel et al. “A phase 1-2 trial of sitravatinib and nivolumab in clear cell renal cell carcinoma following progression on antiangiogenic therapy”. In: *Science translational medicine* 14.641 (2022), eabm6420.

[19] Chris J Sidey-Gibbons et al. “Predicting 180-day mortality for women with ovarian cancer using machine learning and patient-reported outcome data”. In: *Scientific reports* 12.1 (2022), p. 21269.

[20] Yu-Che Chung et al. “DOD-PRO-BART: Dose Optimization Design Incorporating Patient-Reported Outcomes via Machine Learning with Bayesian Additive Regression Trees”. In: *Statistics in Biopharmaceutical Research* 17.3 (2025), pp. 347–356. DOI: `10.1080/19466315.2025.2491596`. eprint: `https://doi.org/10.1080/19466315.2025.2491596`. URL: `https://doi.org/10.1080/19466315.2025.2491596`.

[21] Amélie Anota et al. “An explorative study to assess the association between health-related quality of life and the recommended phase II dose in a phase I trial: idarubicin-loaded beads for chemoembolisation of hepatocellular carcinoma”. In: *BMJ open* 6.6 (2016), e010696.

[22] Jennifer Wheler et al. “Survival of patients in a Phase 1 Clinic: the MD Anderson Cancer Center experience”. In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 115.5 (2009), pp. 1091–1099.

[23] Cecilie Delphin Amdal et al. “SISAQOL-IMI consensus-based guidelines to design, analyse, interpret and present patient-reported outcomes in cancer clinical trials”. In: *The Lancet Oncology* (2025).