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

Emily Alger BSc MMathStat[1]; **Antoine Regnault** PhD[2]; **Amylou C. Dueck** PhD[3]; **Madeline Pe** PhD[4]; **Michael J. Grayling** PhD[5]; **Melanie J. Calvert** PhD[6-10]; **Aaron R. Hansen** MBBS (Hons)[11]; **Olga Kholmanskikh** MD PhD[12]; **Julia Lai-Kwon** MBBS[13]; **J. Jack Lee** PhD[14]; **Anna Minchom** MD(res)[15]; **Yu Qiao** MSc[1]; **Khadija Rerhou Rantell** PhD[16]; **Jessica Roydhouse** PhD[17]; **Claire Snyder** PhD[18]; **Stefan N. Symeonides** PhD [19,20]; **Nolan A. Wages** PhD[21,22]; **Roger Wilson** CBE[23]; **Christina Yap** PhD[1]*.

1: Clinical Trial and Statistics Unit, Institute of Cancer Research, London, UK.
2: Modus Outcomes, A company of THREAD, Lyon, France.
3: Mayo Clinic, Scottsdale, Arizona, USA.
4: European Organisation for Research and Treatment of Cancer, Brussels, Belgium.
5: Statistics and Decision Sciences, Johnson & Johnson, High Wycombe, UK.
6: Centre for Patient-Reported Outcomes Research, University of Birmingham, Birmingham, UK.
7: National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK.
8: NIHR Blood and Transplant Research Unit in Precision Transplant and Cellular Therapeutics, University of Birmingham, Birmingham, UK.
9: Birmingham Health Partners Centre for Regulatory Science and Innovation, Birmingham, UK.
10: University Hospital Birmingham NHS Foundation Trust, Birmingham, UK.
Department of Applied Health Sciences, College of Medicine and Health, University of Birmingham, UK.
11: Division of Cancer Services, Princess Alexandra Hospital, Australia.
12: Federal Agency for Medicines and Health Products, Belgium.
13: Peter MacCallum Cancer Centre, Australia.
14: University of Texas MD Anderson Cancer Center, USA.
15: Royal Marsden Hospital/Institute of Cancer Research, UK.
16: Medicines and Healthcare products Regulatory Agency, UK.
17: Menzies Institute for Medical Research, University of Tasmania, Australia.
18: Johns Hopkins University School of Medicine and Bloomberg School of Public Health, USA.
19: Edinburgh Cancer Centre, NHS Lothian, Edinburgh, UK.
20: Edinburgh Experimental Cancer Medicine Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.
21: Virginia Commonwealth University, School of Public Health, Department of Biostatistics, USA.
22: Massey Comprehensive Cancer Center, USA.
23: Cancer Research Advocates Forum UK, UK.

*Corresponding author
Professor Christina Yap
Address: Clinical Trial and Statistics Unit, Institute of Cancer Research, London, UK.
Email: christina.yap@icr.ac.uk
Phone: 02087224022

The formal publication of this article is available at https://doi.org/10.1016/S1470-2045(26)00018-5

Summary

Patient-reported outcomes (PROs) are increasingly recognised for their role in assessing tolerability in dose-finding oncology trials (DFOTs). However, PRO analysis and reporting within DFOTs is often unclear and inconsistent. OPTIMISE-AR (incOrporating PaTIent-reported outcoMes In doSE-finding trials–Analysis Recommendations) establishes a practical toolkit supporting the statistical analysis, visualisation, and reporting of PRO data within DFOT publications.

International, multidisciplinary, cross-sector statistical analysis and data visualisation working groups identified analytical and visualisation approaches for PROs addressing key DFOT PRO research objectives. Informed by existing literature, case studies and recommendations are provided in this policy review for analysing binary, ordinal, and continuous PRO data to assess tolerability across dose levels and timepoints, and to integrate PROs into interim and final dose-decision processes. The OPTIMISE-AR toolkit is structured around four methodological domains aligned with key DFOT PRO research objectives, providing statistical analysis and data visualisation recommendations for (1) PRO endpoints across time points, (2) PRO endpoints between time points, (3) Time-to-event PRO endpoints, and (4) PRO endpoints for formal dose-decision making in model-based dose-finding designs.

As PROs play an ever-increasing role in tolerability assessment, this work promotes analysis and data visualisation of PRO data, facilitating robust, patient-centred tolerability conclusions and supporting the broader development of tolerable and effective treatments.

Introduction

DFOTs play an essential role in drug development, and are primarily focussed upon the assessment of safety, tolerability, and preliminary efficacy for new treatments with the aim of determining recommended dose(s) for further investigation. There is growing recognition among clinicians and policy makers of the value of patient-reported outcomes (PROs) within the assessment of tolerability in early phase dose-finding oncology trials (DFOTs).[1-3] Whilst a treatment's safety and tolerability has typically been assessed within DFOTs using clinician-reported measures such as Common Terminology Criteria for Adverse Events (CTCAE) grading,[4] the definition of tolerability has been further reconsidered to consider the direct measurement from the patient.[5]

The widening scientific interest in the use of PROs for tolerability assessment in DFOTs is driven by initiatives such as US FDA Project Optimus, Methodology for the Development of Innovative Cancer Therapies (MDICT) Taskforce, and Friends of Cancer Research.[3,6-8] Whilst such interest is reflected in the increasing use of PROs within DFOTs for tolerability assessment,[9] their adoption (and guidance to support their integration) remains scarce.

In recognition of the existing absence of well-defined PRO research objectives for DFOTs,[10] the OPTIMISE-ROR (incOrporating PaTIent-reported outcoMes In doSE-finding trials-Research Objective Recommendations) project[11,12] has recently established consensus-driven PRO research objective recommendations for tolerability assessment within the DFOT setting. However, there remains little guidance to support trialists wishing to analyse and report such research objectives alongside PRO data within their trial reports.

Recommendations have been provided for the analysis and reporting of PRO data within randomised controlled trial (RCT) and single-arm trial settings, including by the SISAQOL consortium (Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data)[13,14] and other groups including FDA,[15,16] EMA,[17] HTA Coordination Group,[18] and the PROTEUS consortium (Patient-Reported Outcomes Tools: Engaging Users and Stakeholders).[19] However, research indicates that when PROs are analysed and reported in DFOTs, statistical analysis approaches and data visualisation methods are commonly inconsistent and poorly reported - reducing the opportunities to provide reliable conclusions for treatment tolerability and compare study findings across trials.[10]

In response to this challenge, the OPTIMISE-AR project (OPTIMISE–Analysis Recommendations) was established to develop a toolkit of analysis and visualisation approaches for PRO data within DFOT reports. The project aims to support authors reporting PRO data in their DFOT publications, increasing transparency and facilitating interpretation by diverse readers, including (but not limited to) clinicians, statisticians, data ethics

committees and regulators. OPTIMISE-AR complements OPTIMISE-ROR guidance,[11,12] providing a PRO analysis toolkit aligned with PRO research objectives crucial to DFOTs. Whilst OPTIMISE-ROR identifies critical PRO research objectives for DFOTs, it does not provide guidance to support trialists in selecting rigorous statistical analysis and data visualisation approaches to report these objectives. OPTIMISE-AR extends beyond the scope of OPTIMISE-ROR to ensure that PRO research objectives implemented in line with OPTIMISE-ROR are analysed and reported rigorously and transparently within trial reports.

Methods

The OPTIMISE-ROR and OPTIMISE-AR projects were developed and registered together[20] with the Enhancing Quality and Transparency of Health Research (EQUATOR) Network's methodological framework for guideline development.[21] An international, cross-sector, multi-disciplinary team was assembled to contribute to the shaping of the OPTIMISE-AR toolkit. This team of 18 members included 6 statisticians and 5 PRO specialists from academia and industry, 4 clinicians, 2 regulators, and 1 patient partner. Based on each member's expertise, members were allocated to two working groups:

The Statistical Analysis group (primarily statisticians and PRO methodologists: ACD, AR, CY, EA, JJL, KRR, MJG, MP, and NAW) explored statistical methods for two key statistical analyses including analysis of PRO data over time and across doses, and estimation and inference using PRO data to find recommended dose(s).

The Data Visualisation group (including all stakeholder categories: ACD, AM, AR, ARH, CS, CY, EA, JL-K, JR, MJC, MJG, MP, OK, RW, SNS) discussed and refined figures and tables to be recommended for use in DFOTs to display PRO data, including descriptive summaries (e.g., boxplots and bar plots) or model-based/estimation outputs (e.g., Kaplan-Meier curves and fitted trends from mixed effects models).

Literature review of analysis and visualisation methods for PROs

A literature review was undertaken to identify relevant statistical analysis and data-visualisation techniques within published literature for the analysis of PRO data. This review informed the initial evidence base from which potential approaches for the toolkit were drawn.

Search strategy and selection criteria

Papers eligible for the literature review were English language guidelines, recommendations, or reviews providing guidance for data analysis or visualisation of tolerability data archived on PubMed between 28/02/2015 and 06/03/2025.

Literature was extracted by EA (Emily Alger) in XML format on 06/03/2025.

Eligible papers were extracted using the following search strategy: ("statistic*"[Title] OR "reporting"[Title] OR "visualising"[Title] OR "analyze"[Title] OR "review"[Title] OR "recommendation*"[Title] OR "guidance"[Title] OR "consensus"[Title] OR "guideline"[Title]) AND ("trial*"[Title] OR "studies"[Title]) AND ("quality of life"[Title] OR "patient reported outcome*"[Title] OR "patient-reported outcome*"[Title] OR "CTCAE"[Title] OR "adverse events"[Title] OR "harms"[Title]) AND ("2015/02/28 00:00":"3000/01/01 05:00"[Date - Publication] AND "English"[Language])

EA and Yu Qiao both assessed each entry for eligibility. All queries that arose during data extraction were discussed and differences of opinions between reviewers were resolved through discussion. 773 papers were assessed for eligibility, and 9 papers were eligible. Figure 1 of the supplementary materials presents the associated PRISMA flow diagram, citing eligible papers.

Extraction of relevant PRO visualisation and statistical modelling approaches

Relevant statistical analysis and data visualisation approaches were extracted from two sources: eligible recommendation papers identified through our literature review (as described above), and published DFOT reports analysing PROs, as collated in Alger and colleagues' methodological review.[10] Additional approaches

were then proposed by the two working groups (including grey literature[22]). Together, these methods formed the pool of candidate approaches for potential inclusion in the toolkit.

Development of analysis toolkit

Extracted statistical analysis and data visualisation approaches were evaluated with respect to the critical PRO research objectives identified for DFOTs as part of OPTIMISE-ROR.[11,12] With particular relevance to this toolkit, OPTIMISE-ROR identifies that (1) PRO concepts for the assessment of tolerability should include Overall Side Effect Impact, Symptomatic Adverse Events, and Overall health-related Quality of Life; (2) PROs should inform final dose-recommendations for dose-escalation and optimisation trials, regardless of trial design and be identified as exploratory/descriptive or inferential; and (3) PRO endpoints should be analysed over a defined time period and at each specific dose level.

These recommendations provided a reference for working group members to critically assess extracted analytical approaches for inclusion in the toolkit.

For (A), PRO measures used to assess PRO concepts for tolerability may include binary, ordinal, or continuous, data types and thus, analyses and visualisations for these data types are presented in this toolkit. Key PRO concepts of overall side effect impact, symptomatic adverse events, and overall health-related quality of life, can be assessed using validated instruments, such as Functional Assessment of Cancer Therapy (FACT)-GP5 item[23], PRO-CTCAE items,[24] and EORTC QLQ-C30 (summary scores)[25] respectively. More generally speaking, the PRO data type for each PRO concept within a trial depends on how the corresponding measure is scored. Furthermore, ordinal or continuous PRO data may be dichotomised and subsequently analysed as binary data using appropriate clinically relevant thresholds when applicable.

For (B), PRO analysis is important across trial design settings. Differences in typical sample sizes and number of investigated doses can shape whether PRO analyses are primarily descriptive or exploratory, or inferential in selecting the recommended dose or dose range. The clinical interpretability of inferences to guide decision making was also considered.

For (C), PRO data analysis and visualisation approaches should consider changes over time and differences at each dose level. Approaches were evaluated considering challenges in analysing DFOTs, which typically have small sample sizes, although this can vary. In two recent reviews, dose-escalation trials were typically planned for a median enrolment of 30 patients (IQR: 18-45)[26], whereas dose-expansion trials enrolled a median of 27 (IQR: 13-51), with the largest including 292 participants.[27]

Working group members assessed the relevance of each approach by considering its suitability and required revisions. They assessed the ability of the extracted approach for DFOTs to accommodate critical PRO tolerability concepts, DFOT sample sizes, and PRO analysis across time and dose levels as per (A), (B), and (C). Potential modifications or improvements to each approach to enhance clarity and use were subsequently considered before identifying additional approaches that could be valuable, or any extracted approaches which should be excluded.

The Statistical analysis working group discussed statistical analysis approaches for PRO data across four meetings between February 18th and July 4th 2025. The group critically assessed the strengths and limitations of each analysis approach for DFOTs, combining their considerations with evaluations from existing literature. Methods were assessed for their relevance within an exploratory or descriptive setting and for inferential statistical analysis, where PROs may be used in conjunction with other endpoints to determine a recommended phase II dose.[28]

The Data Visualisation working group held four meetings between February 18th and April 4th 2025, during which visualisation methods deemed unsuitable for PRO data in DFOTs were first excluded following group discussion and agreement. Visualisations deemed appropriate were standardised and modified in line with existing recommendations for PRO visualisations[22,29] and SISAQOL-IMI[14] guidance. These standardisations include highlighting directionality of PRO scores, descriptive labels (e.g., none/mild/moderate/severe) indicating score meaning, traffic light colors to indicate desirable or undesirable score ranges, and threshold lines across score bars to indicate whether scores are better or worse than threshold scores.

Visualisations were iteratively refined following each meeting and circulated to working group members for additional input. Pre-final recommendations in the OPTIMISE-AR toolkit were rigorously reviewed and refined over two rounds by all team members before full approval. Presented figures maximise interpretability for diverse stakeholders by presenting figures in familiar clinical formats (e.g., box plots, line trajectories) while maintaining methodological rigor in representing PRO change.

Role of the funding source

The funders had no role in study design, data collection, analysis, data interpretation, or writing of the report.

Results

Case study of the analysis of continuous PRO data at the final analysis

As a practical reference for the reader, we begin with a case study illustrating how a PRO research objective and its corresponding analysis can inform clinical decision-making before presenting the OPTIMISE-AR toolkit. Whilst applied to one example, this case study encourages reflection on the broader methods subsequently presented in the OPTIMISE-AR toolkit and the role they may play in the interpretation of PROs for a given trial.

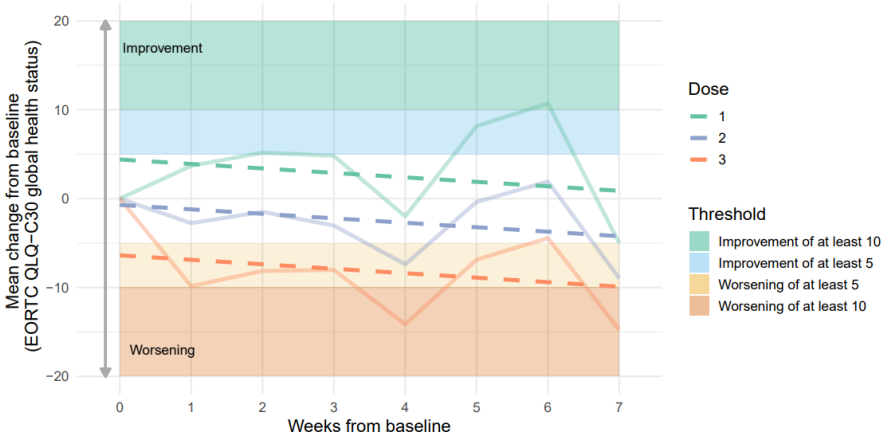
This case study illustrates the use of PROs to assess treatment tolerability based on overall health-related quality of life in a simulated DFOT with three dose levels. EORTC QLQ-C30 Global health status scores are collected over eight weeks (including baseline) for 30 patients per dose, with non-treatment related death and missing PRO data simulated.

Death-related dose discontinuation is one example of an intercurrent event – an event that occurs after treatment initiation and affects the interpretation or the existence of the outcome data.[30] In DFOTs, strategies to handle intercurrent events (e.g. while-on-treatment strategy) are increasingly relevant for endpoints such as PROs which are collected over extended time periods, with the analysis and reporting of PRO data under different strategies potentially influencing trial conclusions and decision-making.[31]

The objective in this study was to assess tolerability by comparing mean change from baseline in EORTC QLQ-C30 Global health status scores across three doses, using a while-on-treatment strategy up to week seven in the safety analysis set.

A linear mixed effects model is utilised to analyse mean change from baseline in the continuous PRO domain, adjusting for dose, time and baseline PRO score, while accounting for repeated patient measurements. The data being simulated and the associated mathematical model are presented in the appendix.

Figure 1 presents the descriptive observed mean changes from baseline at each timepoint, alongside PRO score trends modelled using a mixed effects model across the three dose levels, with a risk table indicating patients with intercurrent events or missing data. Table 1 of the appendix presents the corresponding statistical analysis of the linear mixed-effect model.



Sample size (n = 90)	Dose 1	28	29	25	27	26	24	23	22
	Dose 2	29	29	29	24	26	27	25	25
	Dose 3	29	28	28	27	26	26	24	26
Cumulative intercurrent events	Dose 1	0	1	2	2	2	4	4	6
	Dose 2	0	0	1	2	2	2	2	2
	Dose 3	0	1	2	2	2	2	3	4
Missing data at timepoint	Dose 1	2	0	3	1	1	2	3	2
	Dose 2	1	1	0	4	2	1	3	2
	Dose 3	1	1	0	1	2	2	3	0

Figure 1 Mean change in EORTC QLQ-C30 Global health status from baseline until week 7 (score range: 0-100; higher scores indicate improved global health. Solid lines indicate the observed mean change at each timepoint. Dotted lines display the modelled trend from the linear mixed effects model, adjusting for dose, time, baseline PRO score and accounting for repeated patient measurements. Dose discontinuation is treated as an intercurrent event, with patients excluded from the analysis at subsequent timepoints after discontinuation. Missing data excludes patients with unreported scores due to incomplete patient reporting or data entry errors.

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	-10.8	1.11	[-13.0, -8.60]
Baseline score	0.07	0.02	[0.02, 0.11]

Table 1 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.

In line with SISAQOL-IMI guidelines,[14] analysis inference should be accompanied alongside PRO score interpretation thresholds to indicate meaningful between-group difference of change in score across doses and, depending on the context, it may be appropriate to use thresholds of different magnitudes.[14]

Musoro and colleagues[32] suggest a between-group PRO score interpretation threshold of 5–10 points on the EORTC QLQ-C30 Global Health Status may be considered clinically important. For illustration, primary interpretation reporting of both 5- and 10-point thresholds are utilised for dose comparisons.[32]

The negative timepoint coefficient (-0.50, 95% CI: -0.73 to -0.27) indicates a statistically significant decline in global health status over time, with Dose 2 and Dose 3 showing larger reductions than Dose 1. Dose 1 vs Dose 2 shows a statistically significant difference (-5.11, 95% CI: -7.29 to -2.92), interpreted as clinically meaningful worsening only at the lower 5-point PRO score interpretation threshold. This difference suggests a potential tolerability concern that should be interpreted cautiously and evaluated alongside safety data, including dose-limiting toxicities (DLTs) and adverse events. Dose 1 vs Dose 3 difference is both statistically significant (-10.8

points. 95% CI: -13.0 to -8.60) and clinically meaningful at the 10-point threshold, indicating stronger evidence of tolerability concerns. As some studies have used lower interpretation thresholds for cancer types, careful justification of the chosen interpretation threshold is essential, as this threshold directly impacts the interpretation of results.[32]

OPTIMISE-AR Toolkit

The OPTIMISE-AR toolkit supports the analysis and visualisation of PRO data across relevant data types and is structured around four objectives: (1) PRO endpoints across time points, (2) PRO endpoints between time points, (3) Time-to-event PRO endpoints, and (4) PRO endpoints for formal dose-decision making in model-based dose-finding designs. After addressing objectives (1)-(3), we outline general inferential considerations for the introduced methods in both exploratory and formal decision-making contexts before discussing objective (4).

To ensure standardisation in toolkit presentation, synthesised data were generated to illustrate analysis and visualisation approaches. Simulated PRO-CTCAE data and EORTC QLQ-C30 data, informed by Watson et al[33] and CheckMate 066[34] (NCT01721772) respectively, demonstrate analysis for the case study and objectives (1) – (3). These simulated datasets were designed to reflect realistic patterns of treatment tolerability observed in phase I and advanced melanoma trials. Clinician- and patient-assessed DLT data are simulated for results presented for objective (4). Technical details on data synthesis are provided in the appendix (pp 2-3).

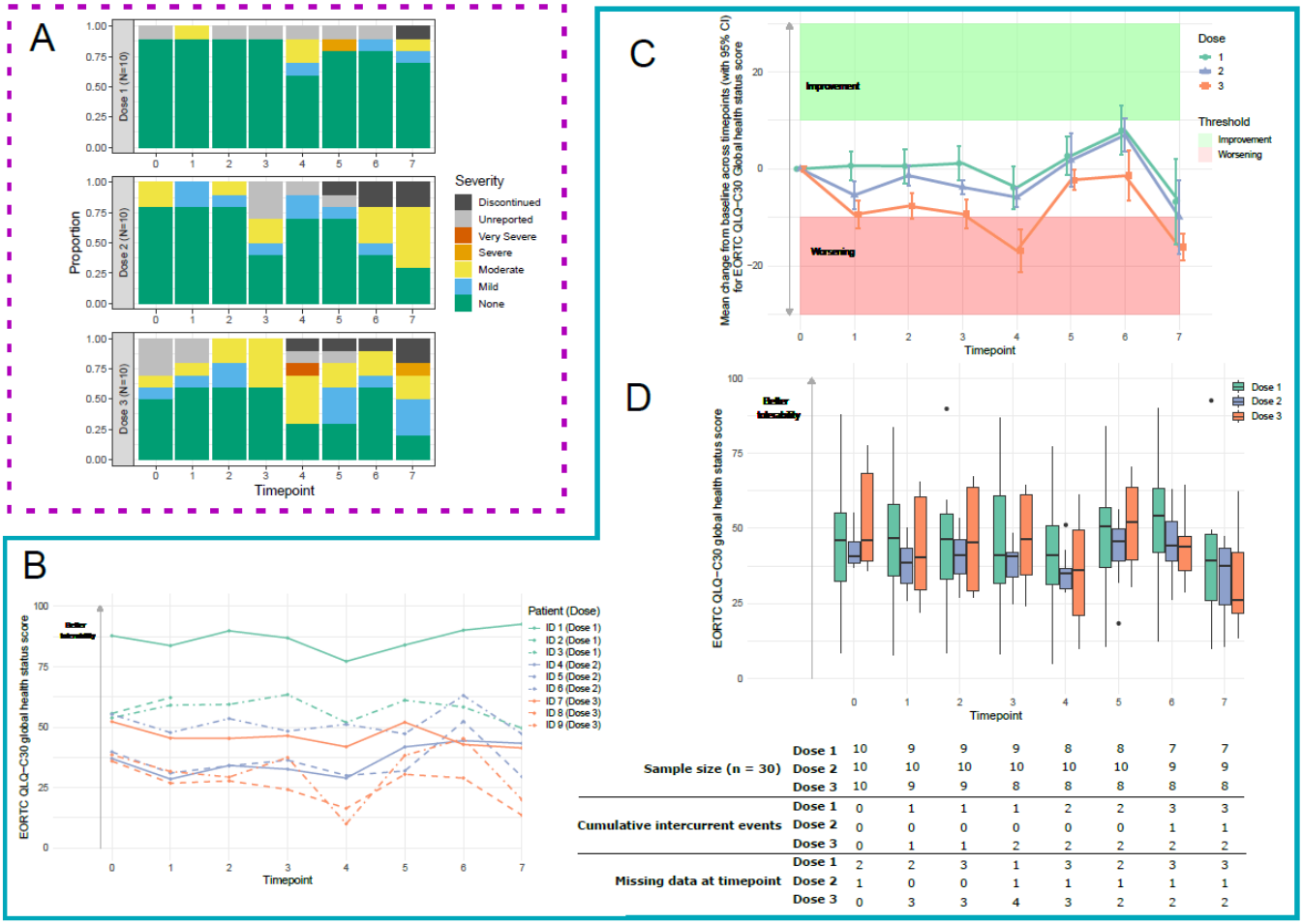
(1) PRO endpoints across timepoints

For DFOTs, a PRO objective could look to assess tolerability across dose levels by analysing the longitudinal tolerability profile of a treatment. Such endpoints may assess whether patients at higher doses experience worse tolerability than patients at lower doses at any timepoint within the trial.

These objectives support researchers to analyse PROs at all reported timepoints, including ordinal (eg. severity of nausea or overall side effect burden) or continuous PRO measures (eg. overall health-related quality of life). Figure 2 presents exemplar visualisations of PRO scores across multiple timepoints and dose levels. Figure 2A, 2C, and 2D focus on dose-level aggregate trends whilst Figure 2B considers individual heterogeneity. Figure 2A builds on proposals for ordinal PRO-CTCAE scores in later phase trials[35,36] to support the comparison across multiple dosages. Figures 2B-D focus on continuous PRO data, with Figure 2B showing individual patient trajectories.[37] Figure 2C evaluates mean change from baseline for PRO scores[35] with associated improvement and worsening thresholds indicated[13] though trialists may also consider visualising mean score rather than mean change in score.[38,39] Figure 2D summarises the distribution of PRO scores as a box plot[40] though violin plots may also be considered.[38]

Missing data and intercurrent events can be reported by dose and timepoint in a table, as shown in Figure 2D. Similar tables can be created for other visualisations (eg. Figure 2C) to transparently report sample sizes at each timepoint to support interpretation of PRO analyses and summaries.

The modelling of PROs across timepoints leverages their longitudinal profile to enable more efficient use of the data. However, effective longitudinal modeling requires careful consideration of the dependency structure inherent for repeated measurements. Statistical modelling approaches to analyse PROs across timepoints are discussed in the appendix (pp 5-8).



Sample size (n = 30)	Dose 1	10	9	9	9	8	8	7	7
	Dose 2	10	10	10	10	10	10	9	9
	Dose 3	10	9	9	8	8	8	8	8
Cumulative intercurrent events	Dose 1	0	1	1	1	2	2	3	3
	Dose 2	0	0	0	0	0	0	1	1
	Dose 3	0	1	1	2	2	2	2	2
Missing data at timepoint	Dose 1	2	2	3	1	3	2	3	3
	Dose 2	1	0	0	1	1	1	1	1
	Dose 3	0	3	3	4	3	2	2	2

Figure 2 Exemplar data visualisations for PROs across timepoints and dose levels, partitioned by whether PRO data is ordinal (purple dotted box, Figure A) or continuous (blue solid box, Figure B-D). (A) PRO-CTCAE symptom items severity scores at each timepoint for three doses with ten patients allocated to each dose. Category "Unreported" indicates patients with no PRO-CTCAE at baseline or post-baseline timepoints. (B) Patient-level trajectories of EORTC QLQ-C30 global health status scores at each assessment timepoint for three doses with three patients allocated to each dose. (C) Mean change from baseline across 7 timepoints of EORTC-QLQ-C30 global health status for three doses with ten patients allocated to each dose, with 95% confidence intervals and clinically important worsening and improvement thresholds indicated. (D) Box plots of EORTC QLQ-C30 global health status scores at each assessment timepoint for three doses with ten patients allocated to each dose, with risk tables indicating the number of patients who receive treatment and reported PROs at each timepoint. A higher score indicates better outcomes in (B)-(D).

(2) PRO endpoints between timepoints

PRO objectives for DFOTs might analyse tolerability to treatment at two clinically meaningful timepoints across dose levels. This objective might include endpoints to assess whether patients at higher doses experience worse tolerability at the final assessment timepoint than patients at lower doses when compared to baseline.

When visualising PROs between two clinically meaningful timepoints, we may consider maximum baseline-adjusted scores and changes from baseline among others. Figure 3 presents data visualisations for differences in PROs between timepoints. Collapsing multiple timepoints into a single summary measure enables figures to display more than one PRO score, whether that be to display multiple symptomatic adverse events or various functional domains. In any case, trialists should check that such comparisons between domains are reasonable. Figures 3A and 3B present differences between ordinal PRO scores as both a magnitude of worsening or improvement[41] and as a maximum baseline-adjusted difference[35] between baseline and a final assessment timepoint. Figure 3B can be readily extended to butterfly plots displaying clinician-reported adverse events alongside patient-reported symptoms, an example is provided within Supplemental Figure 4 in Watson et al.[33] Such figures can indicate concordance between outcomes by visualising insights for both clinician and patient reporting, providing a clear avenue to integrate PRO reporting within more traditional clinician-AE assessment. Figures 3C and 3D present visualisations for continuous PRO scores.[42] In each case, mean change in PRO score are presented with a corresponding measure of variability (95% CI or standard

error). The data presented in Figure 3C can equivalently be presented as a radar plot.[43] Statistical modelling approaches to analyse PROs between timepoints are discussed in the appendix (p9).

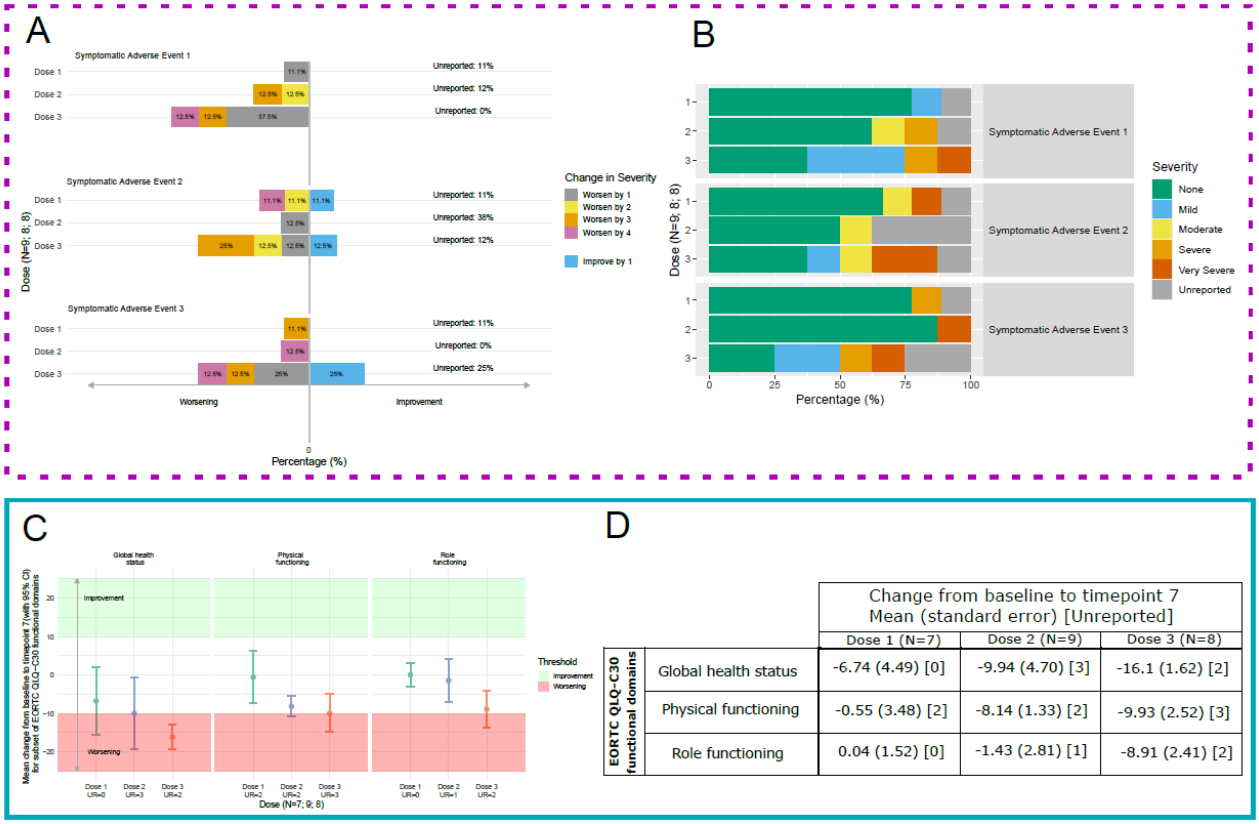


Figure 3 Exemplar data visualisations for PROs between timepoints partitioned by whether PRO data is ordinal (purple dotted box, Figures A and B) or continuous (blue solid box, Figures C and D). (A) Proportion of patients experiencing worsening or improvement for three PRO-CTCAE items between baseline and timepoint 7 for three doses with ten patients allocated to each dose. Category "Unreported" indicates patients with no PRO-CTCAE at the baseline or timepoint 7.Ten patients were allocated to each dose at baseline. At timepoint 7, nine patients in Dose 1, eight in Dose 2, and eight in Dose 3 received treatment. (B) Maximum baseline-adjusted scores for PRO-CTCAE symptom items between baseline and timepoint 7 for three doses with ten patients allocated to each dose. Category "Unreported" indicates patients with no PRO-CTCAE at the baseline or timepoint 7. At timepoint 7, nine patients in Dose 1, eight in Dose 2, and eight in Dose 3 received treatment. (C) Mean change from baseline to timepoint 7 (with 95% CI) for three EORTC QLQ-C30 functional domains for three doses with ten patients allocated to each dose. UR indicates the number of patients who did not report PROs at either timepoint. N indicates the number of patients who received treatment at that final timepoint. (D) Mean change from baseline to timepoint 7 for three EORTC QLQ-C30 functional domains for three doses with ten patients allocated to each dose, with positive change indicating improvement and negative change indicating worsening. Unreported (UR) indicates the number of patients who did not report PROs at either timepoint. N indicates the number of patients who received treatment at that timepoint. A meaningful clinical change in score is defined as an increase (improvement) or decrease (worsening) of at least 10 points.

(3) Time-to-event PRO endpoints

For DFOTs, a PRO objective could look to assess tolerability across dose levels by identifying whether patients at higher doses experience earlier undesirable PRO events (e.g., symptom worsening or dose-limiting events) compared to those at lower doses. Such objectives are typically described using time-to-event endpoints.

Time-to-event PRO endpoints can be evaluated for PRO data types including:
(i) Binary: Time to Patient-assessed DLT,[44]
(ii) Ordinal: Time to first occurrence of baseline-adjusted deterioration or worsening in ordinal PRO scores,
(iii) Continuous: Time to first deterioration or worsening in continuous PRO measures.[45]

For time to first event, trialists might want to analyse PRO endpoints as the expected time to first event occurring or not occuring, or the probability the event (does or does not) occur by a pre-specified timepoint.

Defining a time-to-event PRO endpoint requires pre-specification of the event of interest and analysis timepoint, both of which should be clinically relevant. Trialists should carefully define the analysis population. As some patients may experience events at baseline, endpoints evaluating change from baseline may be more appropriate for survival analysis.

Figure 4 presents an illustrative example of a Kaplan-Meier graph for time-to-event endpoint: no first occurrence of a baseline-adjusted severe PRO-CTCAE score for patients reporting 15 PRO-CTCAE items. The time to first occurrence is stratified by dose and includes 95% confidence intervals for the estimator. Risk tables indicate the number of patients at risk and the cumulative number of patients who experience the event of interest for each dose level. Such tables facilitate clear reporting of intercurrent events which may be handled with patient censoring.

Time-to-event statistical modelling approaches for PRO endpoints are discussed in the appendix (pp 10-11). In settings with informative censoring and competing risks (potentially arising in light of intercurrent events), methods such as the cumulative incidence function may support analysis. Further discussion of intercurrent event handling for time-to-event PRO endpoints, alongside common strengths and limitations of methods is presented in the appendix (p 11).

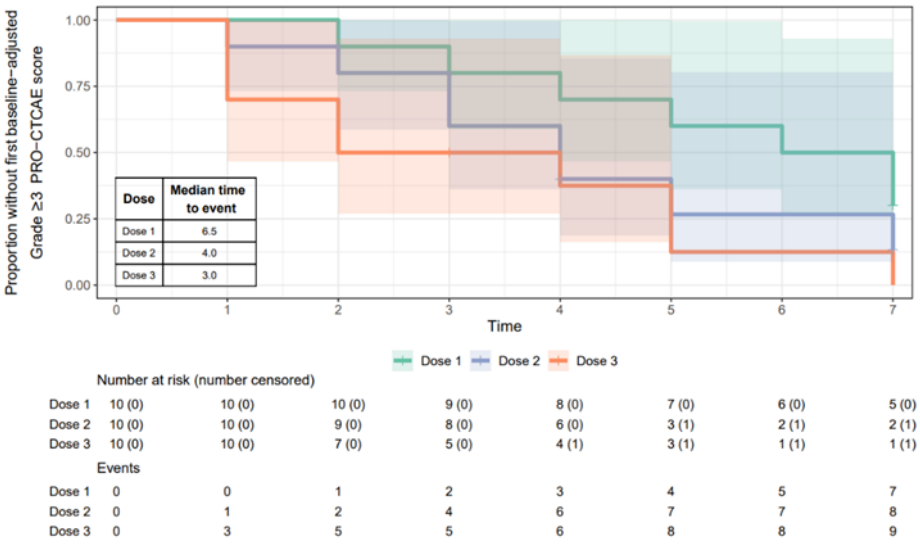


Figure 4 Kaplan-Meier survival curves of time to first occurrence of a baseline-adjusted Grade >=3 PRO-CTCAE score for three doses with ten patients allocated to each dose, with 95% confidence intervals at eight timepoints.

General considerations for Statistical Inference in DFOTs

For formal decision making, the goal of PRO analysis is to support the determination of the maximum tolerated dose (MTD) or optimal biological dose (OBD) - expanding on existing decision-making to consider patient tolerability as a component to the optimal dose. Such doses may be considered the Optimal Tolerable Biological Dose or the PRO-informed tolerable MTD or OBD. A trialist should consider how PRO data is used jointly with existing data to inform decision making for the recommendation of a Phase II dose, including exploratory or descriptive analysis of PROs for clinical interpretation, or the formal embedding of PRO statistical modelling within formal decision-making.

PROs for exploratory analysis, estimation, and hypothesis testing

The use of complex, model-based approaches for the analysis of PROs may provide rigorous statistical inference on patient-assessed tolerability within DFOTs, but feasibility, sample size, and model assumptions should be carefully considered. Whilst rigorous statistical methods are essential for analysis, it is equally important that clinical teams can readily interpret results to support clinical decision-making. To achieve this balance, more complex models could be utilised to assess patient tolerability, while simpler, more interpretable models be utilised to communicate clinically actionable inferences. Exploratory analysis can summarise key trial data and estimation of MTD, OBD, and tolerability rates can provide statistical or clinical inference.

Hypothesis testing is sometimes utilised in DFOTs to analyse PROs for statistical or clinical inference.[10] Trialists should consider whether a summary measure or p-value (or both) is more meaningful within their PRO reporting. While hypothesis testing may support the determination of a recommended Phase II dose, the traditionally small sample sizes in DFOTs often limit hypothesis testing to exploratory, rather than confirmatory

purposes. For each case, trialists should clearly describe where hypothesis testing is undertaken for exploratory or confirmatory purposes. A trialist should clearly identify the research objective they deem most appropriate to ensure estimation and/or hypothesis testing may contribute to the determination of a recommended Phase II dose.

Given the limited sample sizes in DFOTs, we may expect PRO statistical analysis to have wide confidence intervals. It is generally advisable to not only present p-values or point estimates, but also appropriate measures of uncertainty such as confidence intervals or credible intervals (for Bayesian approaches) in line with good reporting standards.[46-48] Such measures provide valuable context on the level of uncertainty within the specific trial setting – even if imprecise.

Trials have utilised exploratory PRO analysis for final dose decision-making, both to confirm identified MTDs as tolerable[45,49,50] and within efficacy-integrated DFOTs when two doses are equally desirable in terms of toxicity and response.[51]

PROs embedded within dose-finding design for formal decision-making

PROs may be utilised formally within interim or final decision-making within DFOTs. Designs such as PRO-CRM (PRO-Continual Reassessment Method),[28] extensions,[44,52] and others[53] introduce PROs as co-primary endpoints within dose-escalation trials, whilst research also indicates PROs' potential role and strengths within dose-optimisation trials.[54,55] In such instances PROs may be introduced alongside other endpoints within one or two-stage designs with benefit-risk trade-off approaches (such as utility frameworks) to determine optimal doses.[56-58]

(4) PRO endpoints for formal dose-decision making in model-based dose-finding designs

We present a simulated trial using a model-based PRO-CRM design demonstrating how PROs can be incorporated into formal interim and final dose-decision making, alongside clinician-reported DLTs.

Similarly to the single outcome CRM design,[59] the PRO-CRM model-based design identifies a clinician-MTD and patient-MTD using clinician-assessed DLTs (C-DLTs) and patient-assessed DLTs (P-DLT) respectively before recommending the minimum of both MTDs. We present statistical analyses and data visualisations for a simulated PRO-CRM trial.

This section is motivated by a phase I study of radiotherapy in endometrial cancer which utilised the PRO-CRM design.[60] A C-DLT was defined as an acute grade 3 or higher gastrointestinal or genitourinary adverse event as per CTCAE. A P-DLT was defined as a severe or very severe gastrointestinal patient-reported symptom per PRO-CTCAE across a subset of clinically relevant adverse events and attributes identified by the trial team.[60]

Patients are enrolled in cohorts of three and observed for DLTs within the first month of treatment. The sample size for this case study is 15 patients, with a target clinician-DLT rate of 0.25 and patient-DLT rate of 0.35. The simulated true C-DLT and P-DLT rates for the three investigated doses are (0.10, 0.16, 0.25) and (0.20, 0.35, 0.50) respectively. The true MTD under this simulation scenario is Dose 2 as it is constrained by the patient-assessed MTD.

In Figure 5A we illustrate estimation of C-DLT and P-DLT rate at final analysis alongside the reporting of C-DLT and P-DLT observations within the trial. This figure can be presented at any interim or final analysis to illustrate PRO-CRM decision boundaries and estimates for decision making. Dose 3 has an estimated C-DLT rate closest to the C-DLT target. Dose 2 has an estimated P-DLT rate closest to the P-DLT target, thus the MTD is identified as Dose 2, the smaller of the two doses closest to each target boundary.

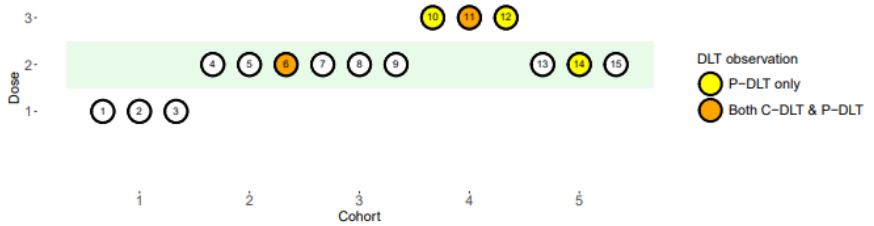
At the final analysis, C-DLT and P-DLT observations are utilised to determine the MTD recommendation. Existing tabular visualisations for DLT estimates for designs such as the CRM[61] can be extended for trial designs incorporating PROs, see Figure 5B. For trialists wishing to use PROs with a benefit risk trade-off, providing tables that present PRO results alongside other key endpoints can be helpful in demonstrating how PROs contribute to the overall benefit-risk evaluation.[51]

Flow diagrams are recommended by CONSORT (CONsolidated Standards Of Reporting Trials) for RCTs to provide an overview of the trial to support readers' critical appraisal of trial findings.[62] Proposals have been made to optimise the illustration of flow diagrams for DFOTs.[47,63] Figure 5C presents a flow diagram for the PRO-CRM trial indicating the number of observed C-DLTs and P-DLTs within each cohort of patients. This figure is relevant and useful for any design which utilises PROs as a key endpoint for dose-decision making.

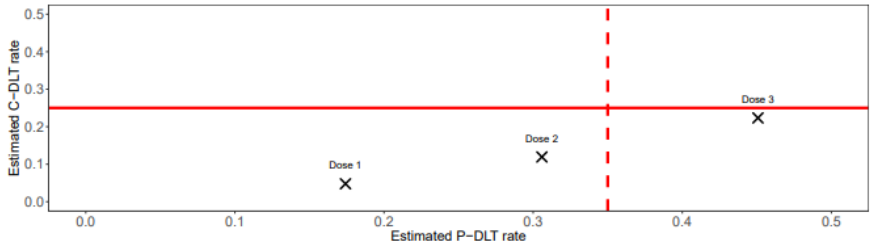
Swimmer plots have been used to illustrate individual patient trajectories on trial and can include key information including dosage, efficacy[64] and DLT information[65]. Figure 5D presents an exemplar swimmer plot for the PRO-CRM trial, with the time of C-DLTs and P-DLTs indicated during each patient's follow-up. Trialists may wish to include other patient outcomes to their swimmer plot, including response indicators.

In this example, PRO-CRM identifies Dose 2 as the MTD. Utilising C-DLT data alone would identify Dose 3 as the MTD, however this dose is associated with a P-DLT rate above the pre-specified target. The inclusion of P-DLT ensures that the MTD is determined considering patient tolerability, ensuring the MTD is both safe and tolerable. Further scenarios demonstrating the role of C-DLT data and P-DLT data within decision making for the PRO-CRM design are presented in more detail in Lee et al.[28]

A



(a) Patient dose assignment in PRO–CRM trial case study with observed Clinician–DLT and Patient–DLT outcomes.

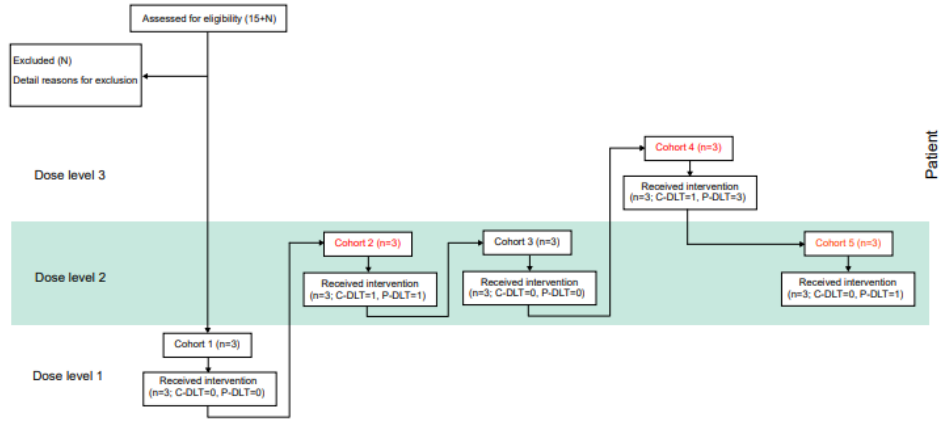


(b) Estimated P–DLT and C–DLT rate at final analysis.

B

	Clinician-DLTs				Patient-DLTs			
Dose	Prior C-DLT Rate	No. of Evaluable Patients	No. of C-DLTs	Estimated DLT rate (90% probability interval)	Prior P-DLT Rate	No. of Evaluable Patients	No. of P-DLTs	Estimated DLT rate (90% probability interval)
Dose 1	0.06	3	0	0.05 (0.01, 0.18)	0.10	3	0	0.17 (0.04, 0.39)
Dose 2	**0.14**	**9**	**1**	**0.12 (0.02, 0.30)**	**0.21**	**9**	**2**	**0.31 (0.11, 0.53)**
Dose 3	0.25	3	1	0.22 (0.07, 0.43)	0.35	3	3	0.45 (0.22, 0.65)

C



D

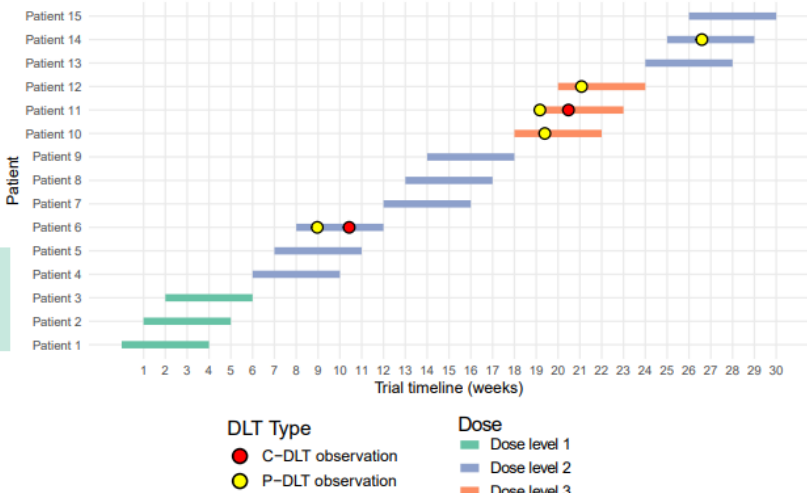


Figure 5 **A***: Exemplar figure detailing (a) observed C-DLTs and P-DLTs and (b) estimation of Clinician-reported DLTs (C-DLTs) and Patient-reported DLTs (P-DLTS) at final analysis using PRO-CRM trial design. The sample size for this case study is 15 patients, with a target clinician-DLT rate of 0.25 (red solid line) and patient-DLT rate of 0.35 (red dotted line).* **B***: Prior and estimated probability of Clinician-DLT and Patient-DLT rate for each dose level at the end of the trial using the PRO-CRM trial design. The final estimated Maximum tolerated dose highlighted in bold. The sample size for this case study is 15 patients, with a target clinician-DLT rate of 0.25 and patient-DLT rate of 0.35. C-DLTs are defined as an acute grade 3 or higher gastrointestinal or genitourinary severity per CTCAE and P-DLTs are defined as a severe or very severe gastrointestinal side effect as per PRO-CTCAE.* **C***: Flow diagram of dose escalation trial design using PRO-CRM. Cohorts where either clinician-reported DLTs (C-DLTs) or patient-reported DLTs (P-DLTs) occur are highlighted in red. The recommended maximum tolerated dose is indicated in green. The sample size for this case study is 15 patients, with a target C-DLT rate of 0.25 and P-DLT rate of 0.35.* **D***: Swimmer plot by dose level for 15 patients enrolled in cohorts of 3 across doses with a dose-escalation study using PRO-CRM.*

Discussion

Building upon OPTIMISE-ROR guidance for critical PRO research objectives,[11,12] OPTIMISE-AR provides an analytical toolkit to support trialists wanting to analyse and visualise PRO data for published DFOTs reports together with setting-specific implementation considerations for DFOTs.

Designed for a multidisciplinary readership, OPTIMISE-AR balances statistical rigor with clarity, with methods offering insights at both patient and dose level whilst remaining accessible and consistent with existing PRO guidance.

As the role of PROs in DFOTs continues to grow, OPTIMISE-AR offers a practical, fit-for-purpose resource that promotes transparent and rigorous PRO analysis and reporting. Drawing on existing evidence in DFOTs and other trial settings, OPTIMISE-AR provides foundational recommendations designed to be adaptable rather than exhaustive. Further consensus-driven guidance may be needed to support trialists as DFOT designs and PRO integration evolve.

Building upon case studies demonstrating PRO analysis and reporting, the OPTIMISE-AR toolkit is positioned for application to additional real-world case studies to further refine its recommendations. To support practical implementation, the toolkit is accompanied by R code (https://github.com/alemily100/optimise-ar). A user-friendly Shiny app is in development to support multidisciplinary trialists apply recommended methods. Although designed primarily to support PRO data analysis in publications, future developments could broaden OPTIMISE-AR scope to facilitate PRO reporting at other critical stages, including interim decision-making and review by safety and dose-decision oversight committees, whilst ensuring analyses are accessible and meaningful for patient advocates.[29]

Though not deemed critical by OPTIMISE-ROR,[11,12] the application of statistical analysis in the estimand framework is gaining increasing relevance for DFOTs. For trialists wishing to utilise the estimand framework, existing literature outlines the attributes of estimands relevant to many analysis methods discussed in this paper.[66] In the presence of intercurrent events, trialists may wish to utilise more complex statistical analysis approaches including weighted average survival and inverse probability of censoring weighting for time-to-event endpoints.[67] Joint-modelling approaches may also be considered for the longitudinal analysis of PROs in the presence of censoring.[68] The utility of such advanced statistical methods remains dependent on sample size, with methods adapted to small samples particularly relevant. The figures presented in OPTIMISE-AR use a while-on-treatment approach.[31] To utilise other strategies for the handling of intercurrent events, demonstrated approaches suggested here can be adjusted appropriately.

Trialists analysing PROs in DFOTs can utilise OPTIMISE-AR alongside other resources developed for PRO integration in other trial settings including SISAQOL-IMI,[14] providing guidance for handling missing data and applications to the estimand framework. Further recommendations are provided by the PROTEUS Consortium,[19] and SPIRIT-PRO (Standard Protocol Items: Recommendations for Interventional Trials-PRO)[69] and CONSORT-PRO[70] reporting guidelines for trial protocols and RCT reports. However, such guidance is not designed specifically for the DFOT setting. Trialists should carefully assess its relevance considering DFOT-specific challenges such as the presentation of analysis across doses, typically small sample sizes, and the need to support both clinical and statistical assessment of tolerability and dose recommendations.

Conclusion

By supporting trialists analyse PRO data within DFOTs, OPTIMISE-AR facilitates the adoption of transparent, rigorous analytic and reporting practices, thereby enabling patient-centred tolerability conclusions to inform dose-decision making. This work strengthens the methodological rigour applied to PRO data, facilitating robust inferences and conclusions, and the development of both tolerable and effective treatments. OPTIMISE-AR promotes interdisciplinary collaboration among statisticians, clinicians, and patient representatives, promoting shared understanding and communication to integrate PRO findings meaningfully in DFOTs. Together, these efforts aim to advance the development of more patient-focused and scientifically robust trial outcomes.

Authors' contributions

EA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualisation, Writing – original draft, Writing – review & editing. **AR**: Investigation, Methodology, Writing – review & editing. **ACD**: Investigation, Methodology, Writing – review & editing. **MP**: Investigation, Methodology, Writing – review & editing. **MJG**: Investigation, Methodology, Writing – review & editing. **MJC**: Investigation, Methodology, Writing – review & editing. **ARH**: Investigation, Methodology, Writing – review & editing. **OK**: Investigation, Methodology, Writing – review & editing. **JLK**: Investigation, Methodology, Writing – review & editing. **JJL**: Investigation, Methodology, Writing – review & editing. **AM**: Investigation, Methodology, Writing – review & editing. **YQ**: Data curation, Investigation, Methodology, Software, Validation, Writing – review & editing. **KRR**: Investigation, Methodology, Writing – review & editing. **JR**: Investigation, Methodology, Writing – review & editing. **CS**: Investigation, Methodology, Writing – review & editing. **SNS**: Investigation, Methodology, Writing – review & editing. **NAW**: Investigation, Methodology, Writing – review & editing. **RW**: Investigation, Methodology, Writing – review & editing. **CY**: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

EA has been supported to undertake this work as part of a PhD studentship from the Institute of Cancer Research within the MRC/NIHR Trials Methodology Research Partnership. **JJL**'s research was supported in part by the grants P30CA016672from the National Cancer Institute in USA. **JR** receives funding outside the submitted work to her institution from the Patient-Centered Outcomes Research Institute, Prostate Cancer Foundation of Australia, Cancer Australia, EuroQol Foundation, Cancer Council of Tasmania, Movember Foundation, Royal Hobart Hospital Research Foundation, Gilead Sciences, and Pfizer Inc. **NAW** is supported by the National Institute of Health grant R01CA247932. **MJC** is funded by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. **CY** receives programmatic infrastructure funding from Cancer Research UK (CTUQQR-Dec22/100004) which supported this work.

The other authors declared no conflicts of interest.

Declaration of interests

AR is an employee of Modus Outcomes, a patient-centered research consultancy providing services to multiple pharmaceutical industry companies. **MJC** is Director of the Centre for the Centre for Patient Reported Outcomes Research, Deputy Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation and is a National Institute for Health and Care Research (NIHR) Senior Investigator. Grants: NIHR Applied Research Collaboration, Health Data Research UK, Innovate UK, Macmillan Cancer Support, GSK, Research England, European Commission and EFPIA, Brain Tumour Charity, Gilead, Janssen, Merck, NIHR, UKRI. Licenses: Symptom Burden Questionnaire-Long COVID. Consulting fees: Aparito Ltd, Boehringer Ingelheim, CIS Oncology, Merck, Genentech, GSK, Pfizer, Vertex Pharmaceuticals, ICON, Halfloop, University of Michigan, Northwestern University, EuroQoL, Shionogi, Cell and Gene Therapy Catapult outside the submitted work. Payment or honoraria for lectures: University of Maastricht, Sourh-Eastern Norway Regional Health Authority, Cochrane Portugal, Singapore National Medical Research Council. Leadership: PROTEUS consortium. Stock: Family member has stock in GSK. **ARH** Grants or contracts: Advancell, Aveo, BMS, Full-Life, Fusion, Janssen, Macrogenics, MSD, Roche, Seagen, Tyra Biosciences, MOMA Therapeutics, Astellas, Bayer, Eisai. Consulting fees: Astellas, Bayer, Full-Life Technologies, Eisai, Janssen and MSD. Honoraria for lectures: Astellas, Janssen and MSD. Support for meeting attendance: Astellas, Bayer. **AM** grants: Astex, Merck, MSD. Consulting fees: Merck, Janssen, Pfizer, AstraZeneca, Immutep, MSD. Speaker: AstraZeneca, Janssen, Takeda, Seagen, Merck, GSK, Faron. Support for meeting attendance: Janssen. **JR** is an Editor-in-Chief of

Quality of Life Research. Grants or contracts: Prostate Cancer Foundation of Australia, Patient-Centered Outcomes Research Institute, Food and Drug Administration, Movemeber Foundation, EuroQol Foundation, Cancer Australia, Pfizer Inc, Cancer Council Tasmania, Royal Hobart Hospital Research Foundation. Consulting fees: Gilead Sciences. Support for meeting attendance: Australian Clinical Trials Alliance. **CS** has received research funding from Pfizer and Genentech through her institution, personal fees from Shionogi and Movember, and travel support to present at conferences from Shionogi and Executive Insights Healthcare Consultants. **SNS** reports institutional grants: MSD, Verastem. Payment or honoraria for lectures: Cancer Drug Development Forum (NPO), Ipsen. Support for meeting attendance: Cancer Research UK (charity). Advisory boards: Duke Street Bio, EISAI, Ellipses, MSD, Roche. Data Safety Monitoring board: Exscientia/Recursion, Grey Wolf Therapeutics, WCG/Valley Hospital. Unpaid board member: Cancer Drug Development Forum (NPO). **NAW** reports institutional grants: NIH/NCI P30CA016059, NIH/NCI P01CA275740, NIH/NCI R01CA285391, NIH/NCI U54CA283762, US Department of Defense DoD HT94252310742. Payment or honoraria for lectures: Bristol-Myers Squibb Foundation)-AACR Design and Implementation of Clinical Trials Workshop, The Robert A. Winn Excellence in Clinical Trials Award Program. **CY** reports personal fees from Faron Pharmaceuticals, Bayer, Trogenix and Merck, outside the submitted work.

All other authors have no disclosures to declare.

The personal views and opinions expressed in this publication are those of the individual authors and may not be understood or quoted as being made on behalf of or reflecting the position of any organization or working group with which the authors are affiliated, the European Medicines Agency, or any of its scientific committees or working parties. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

1. Basch E, Yap C. Patient-reported outcomes for tolerability assessment in phase i cancer clinical trials. *JNCI: Journal of the National Cancer Institute*. 2021;113(8):943-944. doi:10.1093/jnci/djab017
2. Yap C, Lee Aiyegbusi O, Alger E, et al. Advancing patient-centric care: Integrating patient reported outcomes for tolerability assessment in early phase clinical trials - insights from an expert virtual roundtable. *eClinicalMedicine*. 2024;76doi:10.1016/j.eclinm.2024.102838
3. Friends of Cancer Research. Supporting a patient-centric approach to dose optimization in oncology: The essential role of patient-reported outcomes (pros). Accessed 19th January 2024, https://friendsofcancerresearch.org/wp-content/uploads/Supporting_Patient-Centric_Approach_Dose_Optimization_Oncology-PROs.pdf
4. Lai-Kwon J, Vanderbeek AM, Minchom A, et al. Using patient-reported outcomes in dose-finding oncology trials: Surveys of key stakeholders and the national cancer research institute consumer forum. *The Oncologist*. 2022;27(9):768-777. doi:10.1093/oncolo/oyac117
5. Basch E, Campbell A, Hudgens S, et al. *Broadening the definition of tolerability in cancer clinical trials to capture the patient experience*. 2020. https://friendsofcancerresearch.org/wp-content/uploads/Comparative-Tolerability-Whitepaper_FINAL.pdf
6. United States Food and Drug Administration. Project optimus: Reforming the dose optimization and dose selection paradigm in oncology. Accessed 17th April 2024, https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus
7. United States Food and Drug Administration. Optimizing the dosage of human prescription drugs and biological products for the treatment of oncologic diseases guidance for industry. 2023. Accessed 9th June 2025. https://www.fda.gov/media/164555/download
8. Araujo D, Greystoke A, Bates S, et al. Oncology phase i trial design and conduct: Time for a change - mdict guidelines 2022. *Annals of Oncology*. 2023;34(1):48-60. doi:10.1016/j.annonc.2022.09.158
9. Lai-Kwon J, Yin Z, Minchom A, Yap C. Trends in patient-reported outcome use in early phase dose-finding oncology trials - an analysis of clinicaltrials.Gov. *Cancer Med*. Nov 2021;10(22):7943-7957. doi:10.1002/cam4.4307
10. Alger E, Minchom A, Lee Aiyegbusi O, Schipper M, Yap C. Statistical methods and data visualisation of patient-reported outcomes in early phase dose-finding oncology trials: A methodological review. *eClinicalMedicine*. Oct 1 2023 2023;64:102228. doi:https://doi.org/10.1016/j.eclinm.2023.102228

11. Alger E, Aiyegbusi OL, Dueck A, et al. 87o international consensus-driven recommendations for patient-reported outcome research objectives in early phase dose-finding oncology clinical trials: Optimise-ror. *ESMO Open*. 2025;10doi:10.1016/j.esmoop.2025.105438
12. Alger E, Aiyegbusi OL, Dueck AC, et al. International consensus-driven recommendations for patient-reported outcome research objectives in early phase dose-finding oncology trials: Optimise-ror. *Journal of Clinical Oncology (in press)*.
13. Coens C, Pe M, Dueck AC, 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. *The Lancet Oncology*. Feb 1 2020 2020;21(2):e83-e96. doi:https://doi.org/10.1016/S1470-2045(19)30790-9
14. Amdal CD, Falk RS, Alanya A, et al. Sisaqol-imi consensus-based guidelines to design, analyse, interpret and present patient-reported outcomes in cancer clinical trials. *The Lancet Oncology*. 2025;
15. United States Food and Drug Administration. Fda patient-focused drug development guidance series for enhancing the incorporation of the patient's voice in medical product development and regulatory decision making. Accessed 17th April 2024, https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical
16. United States Food and Drug Administration. Core patient-reported outcomes in cancer clinical trials guidance for industry. Accessed 17th April 2024, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials
17. European Medicines Agency. *Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man - the use of patient-reported outcome (pro) measures in oncology studies - scientific guideline*. 2016. Accessed 26th June 2025. https://www.ema.europa.eu/en/documents/other/appendix-2-guideline-evaluation-anticancer-medicinal-products-man_en.pdf
18. Health Technology Assessment Coordination Group. Guidance on outcomes for joint clinical assessments. Accessed 26th June 2025, https://health.ec.europa.eu/document/download/a70a62c7-325c-401e-ba42-66174b656ab8_en?filename=hta_outcomes_jca_guidance_en.pdf
19. Proteus Consortium. Accessed 2nd June 2025, https://theproteusconsortium.org/
20. Optimise-ar: Incorporating patient-reported outcomes in dose-finding oncology trials – analysis recommendations Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network,. Accessed 2nd June 2025, https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-clinical-trials/#OPTIMISE-AR
21. Altman DG, Simera I, Hoey J, Moher D, Schulz K. Equator: Reporting guidelines for health research. *Open Med*. 2008;2(2):e49-50.
22. U.S. Food and Drug Administration. Project patient voice. Updated 17/03/2023. Accessed 12th March 2024, 2024. https://www.fda.gov/about-fda/oncology-center-excellence/project-patient-voice
23. Facit.Org. Facit searchable library and custom form developer (build-a-pro). Accessed 15th April 2025, https://wizard.facit.org
24. Basch E, Reeve BB, Mitchell SA, et al. Development of the national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (pro-ctcae). *J Natl Cancer Inst*. Sep 2014;106(9)doi:10.1093/jnci/dju244
25. European Organisation for Research and Treatment of Cancer. Eortc item library. Accessed 15th April 2025, https://itemlibrary.eortc.org
26. Villacampa G, Patel D, Zheng H, et al. Assessing the reporting quality of early phase dose-finding trial protocols: A methodological review. *eClinicalMedicine*. 2023;60doi:10.1016/j.eclinm.2023.102020
27. Herrero Colomina J, Hu X, Dinizulu H, et al. Expansion cohorts in phase 1 oncology trials: A systematic review of their design, implementation and outcomes. *British Journal of Cancer*. 2026/01/10 2026;doi:10.1038/s41416-025-03334-5
28. Lee SM, Lu X, Cheng B. Incorporating patient-reported outcomes in dose-finding clinical trials. *Stat Med*. Feb 10 2020;39(3):310-325. doi:10.1002/sim.8402
29. Snyder C, Smith K, Holzner B, et al. Making a picture worth a thousand numbers: Recommendations for graphically displaying patient-reported outcomes data. *Quality of Life Research*. 2019/02/01 2019;28(2):345-356. doi:10.1007/s11136-018-2020-3
30. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: A primer on the ich e9(r1) addendum. *BMJ*. 2024;384:e076316. doi:10.1136/bmj-2023-076316
31. Mercier F, Homer V, Geng J, et al. Estimands in oncology early clinical development: Assessing the impact of intercurrent events on the dose-toxicity relationship. *Statistics in Biopharmaceutical Research*. 2025/01/02 2025;17(1):78-86. doi:10.1080/19466315.2023.2296648

32. Musoro JZ, Coens C, Sprangers MaG, et al. Minimally important differences for interpreting eortc qlq-c30 change scores over time: A synthesis across 21 clinical trials involving nine different cancer types. *European Journal of Cancer*. 2023/07/01/ 2023;188:171-182. doi:https://doi.org/10.1016/j.ejca.2023.04.027
33. Watson GA, Veitch ZW, Shepshelovich D, et al. Evaluation of the patient experience of symptomatic adverse events on phase i clinical trials using pro-ctcae. *British Journal of Cancer*. 2022/11/01 2022;127(9):1629-1635. doi:10.1038/s41416-022-01926-z
34. Long GV, Atkinson V, Ascierto PA, et al. Effect of nivolumab on health-related quality of life in patients with treatment-naïve advanced melanoma: Results from the phase iii checkmate 066 study. *Annals of Oncology*. 2016/10/01/ 2016;27(10):1940-1946. doi:https://doi.org/10.1093/annonc/mdw265
35. Regnault A, Loubert A, Gorsh B, et al. A toolbox of different approaches to analyze and present pro-ctcae data in oncology studies. *JNCI: Journal of the National Cancer Institute*. 2023;115(5):586-596. doi:10.1093/jnci/djad018
36. U.S. Food and Drug Administration. Aura3: Numbness or tingling in hands or feet. Accessed 25th July 2025, https://www.fda.gov/about-fda/aura3/aura3-numbness-or-tingling-hands-or-feet
37. Schmidt M, Lenhard H, Hoenig A, et al. Tumor suppression, dose-limiting toxicity and wellbeing with the fetal estrogen estetrol in patients with advanced breast cancer. *J Cancer Res Clin Oncol*. Jun 2021;147(6):1833-1842. doi:10.1007/s00432-020-03472-8
38. Phillips R, Cro S, Wheeler G, et al. Visualising harms in publications of randomised controlled trials: Consensus and recommendations. *BMJ*. 2022;377:e068983. doi:10.1136/bmj-2021-068983
39. Ballas LK, Luo C, Chung E, et al. Phase 1 trial of sbrt to the prostate fossa after prostatectomy. *Int J Radiat Oncol Biol Phys*. May 1 2019;104(1):50-60. doi:10.1016/j.ijrobp.2018.12.047
40. Cassier PA, Italiano A, Gomez-Roca C, et al. Long-term clinical activity, safety and patient-reported quality of life for emactuzumab-treated patients with diffuse-type tenosynovial giant-cell tumour. *Eur J Cancer*. Dec 2020;141:162-170. doi:10.1016/j.ejca.2020.09.038
41. Zhou X, Eid D, Gnanasakthy A. Methods for reporting the patient-reported outcomes version of the common terminology criteria for adverse events (pro-ctcae) data in cancer clinical trials. *Value in Health*. 2018;21:S226. doi:10.1016/j.jval.2018.04.1528
42. Chawla SP, Goel S, Chow W, et al. A phase 1b dose escalation trial of nc-6300 (nanoparticle epirubicin) in patients with advanced solid tumors or advanced, metastatic, or unresectable soft-tissue sarcoma. *Clinical Cancer Research*. 2020;26(16):4225-4232. doi:10.1158/1078-0432.Ccr-20-0591
43. Raman B, Cassar MP, Tunnicliffe EM, et al. Medium-term effects of sars-cov-2 infection on multiple vital organs, exercise capacity, cognition, quality of life and mental health, post-hospital discharge. *EClinicalMedicine*. 2021/01/01/ 2021;31:100683. doi:https://doi.org/10.1016/j.eclinm.2020.100683
44. Andrillon A, Biard L, Lee SM. Incorporating patient-reported outcomes in dose-finding clinical trials with continuous patient enrollment. *J Biopharm Stat*. Jul 26 2023:1-12. doi:10.1080/10543406.2023.2236216
45. Anota A, Boulin M, Dabakuyo-Yonli S, 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. *BMJ Open*. Jun 24 2016;6(6):e010696. doi:10.1136/bmjopen-2015-010696
46. Yap C, Solovyeva O, De Bono J, et al. Enhancing reporting quality and impact of early phase dose-finding clinical trials: Consort dose-finding extension (consort-define) guidance. *BMJ*. 2023;383:e076387. doi:10.1136/bmj-2023-076387
47. Rekowski J, Guo C, Solovyeva O, et al. Consort-define explanation and elaboration: Recommendations for enhancing reporting quality and impact of early phase dose-finding clinical trials. *eClinicalMedicine*. 2025;79doi:10.1016/j.eclinm.2024.102987
48. Homer V, Yap C, Bond S, et al. Early phase clinical trials extension to guidelines for the content of statistical analysis plans. *BMJ*. 2022;376:e068177. doi:10.1136/bmj-2021-068177
49. Goody RB, Mackay H, Pitcher B, et al. Phase 1/2 study of the addition of cisplatin to adjuvant chemotherapy with image guided high-precision radiation therapy for completely resected gastric cancer. *Int J Radiat Oncol Biol Phys*. Dec 1 2016;96(5):994-1002. doi:10.1016/j.ijrobp.2016.08.034
50. Guiu B, Jouve J-L, Schmitt A, et al. Intra-arterial idarubicin_lipiodol without embolisation in hepatocellular carcinoma: The lida-b phase i trial. *Journal of Hepatology*. 1 Jun 2018 2018;68(6):1163-1171. doi:https://doi.org/10.1016/j.jhep.2018.01.022
51. Msaouel P, Goswami S, Thall PF, et al. A phase 1-2 trial of sitravatinib and nivolumab in clear cell renal cell carcinoma following progression on antiangiogenic therapy. *Sci Transl Med*. Apr 20 2022;14(641):eabm6420. doi:10.1126/scitranslmed.abm6420

52. Alger E, Lee SM, Cheung YK, Yap C. U-pro-crm: Designing patient-centred dose-finding trials with patient-reported outcomes. *ESMO Open*. Jul 1 2024;9(7):103626. doi:https://doi.org/10.1016/j.esmoop.2024.103626
53. Wages NA, Lin R. Isotonic phase i cancer clinical trial design utilizing patient-reported outcomes. *Statistics in Biopharmaceutical Research*. 2024:1-19. doi:10.1080/19466315.2023.2288013
54. Chung Y-C, Zhao Y, Liu M, Lin J, Liu R. Dod-pro-bart: Dose optimization design incorporating patient-reported outcomes via machine learning with bayesian additive regression trees. *Statistics in Biopharmaceutical Research*.1-10. doi:10.1080/19466315.2025.2491596
55. Alger E, Van Zyl M, Aiyegbusi OL, et al. Patient and public involvement and engagement in the development of innovative patient-centric early phase dose-finding trial designs. *Research Involvement and Engagement*. 2024/06/19 2024;10(1):63. doi:10.1186/s40900-024-00599-7
56. Yuan Y, Zhou H, Liu S. Statistical and practical considerations in planning and conduct of dose-optimization trials. *Clinical Trials*. 2024/06/01 2024;21(3):273-286. doi:10.1177/17407745231207085
57. Zhou Y, Lee JJ, Yuan Y. A utility-based bayesian optimal interval (u-boin) phase i/ii design to identify the optimal biological dose for targeted and immune therapies. *Statistics in Medicine*. 2019/12/10 2019;38(28):S5299-S5316. doi:https://doi.org/10.1002/sim.8361
58. Lin R, Zhou Y, Yan F, Li D, Yuan Y. Boin12: Bayesian optimal interval phase i/ii trial design for utility-based dose finding in immunotherapy and targeted therapies. *JCO Precision Oncology*. 2020/11/01 2020;(4):1393-1402. doi:10.1200/PO.20.00257
59. O'quigley J, Pepe M, Fisher L. Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics*. 1990;46(1):33-48. doi:10.2307/2531628
60. Wages NA, Nelson B, Kharofa J, Meier T. Application of the patient-reported outcomes continual reassessment method to a phase i study of radiotherapy in endometrial cancer. *Int J Biostat*. Nov 17 2022;doi:10.1515/ijb-2022-0023
61. Craddock C, Slade D, De Santo C, et al. Combination lenalidomide and azacitidine: A novel salvage therapy in patients who relapse after allogeneic stem-cell transplantation for acute myeloid leukemia. *Journal of Clinical Oncology*. 2019/03/01 2019;37(7):580-588. doi:10.1200/JCO.18.00889
62. Hopewell S, Hirst A, Collins GS, Mallett S, Yu L-M, Altman DG. Reporting of participant flow diagrams in published reports of randomized trials. *Trials*. 2011/12/05 2011;12(1):253. doi:10.1186/1745-6215-12-253
63. Alger E, Zhang Y, Yap C. Reporting quality of consort flow diagrams in published early phase dose-finding clinical trial reports: Improvement is needed. *Contemporary Clinical Trials*. 2023/08/01/ 2023;131:107277. doi:https://doi.org/10.1016/j.cct.2023.107277
64. Fukuoka S, Hara H, Takahashi N, et al. Regorafenib plus nivolumab in patients with advanced gastric or colorectal cancer: An open-label, dose-escalation, and dose-expansion phase ib trial (regonivo, epoc1603). *Journal of Clinical Oncology*. 2020/06/20 2020;38(18):2053-2061. doi:10.1200/JCO.19.03296
65. Yin Z, Mander AP, De Bono JS, Zheng H, Yap C. Handling incomplete or late-onset toxicities in early-phase dose-finding clinical trials: Current practice and future prospects. *JCO Precision Oncology*. 2024/01/01 2024;(8):e2300441. doi:10.1200/PO.23.00441
66. Sakamaki K, Kawahara T. Statistical methods and graphical displays of quality of life with survival outcomes in oncology clinical trials for supporting the estimand framework. *BMC Medical Research Methodology*. 2022/10/04 2022;22(1):259. doi:10.1186/s12874-022-01735-1
67. Tassistro E, Bernasconi DP, Valsecchi MG, Antolini L. Adverse events in single-arm clinical trials with non-fatal time-to-event efficacy endpoint: From clinical questions to methods for statistical analysis. *BMC Medical Research Methodology*. 2024/01/03 2024;24(1):3. doi:10.1186/s12874-023-02123-z
68. Daza JF, Mitani AA, Alibhai SMH, et al. Joint models inform the longitudinal assessment of patient-reported outcomes in clinical trials: A simulation study and secondary analysis of the restrictive vs. Liberal fluid therapy for major abdominal surgery (relief) randomized controlled trial. *Journal of Clinical Epidemiology*. 2024/12/01/ 2024;176:111553. doi:https://doi.org/10.1016/j.jclinepi.2024.111553
69. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: The spirit-pro extension. *JAMA*. 2018;319(5):483-494. doi:10.1001/jama.2017.21903
70. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: The consort pro extension. *JAMA*. 2013;309(8):814-822. doi:10.1001/jama.2013.879