AstraZeneca

Incorporating durability endpoints in decision making in early oncology clinical trials

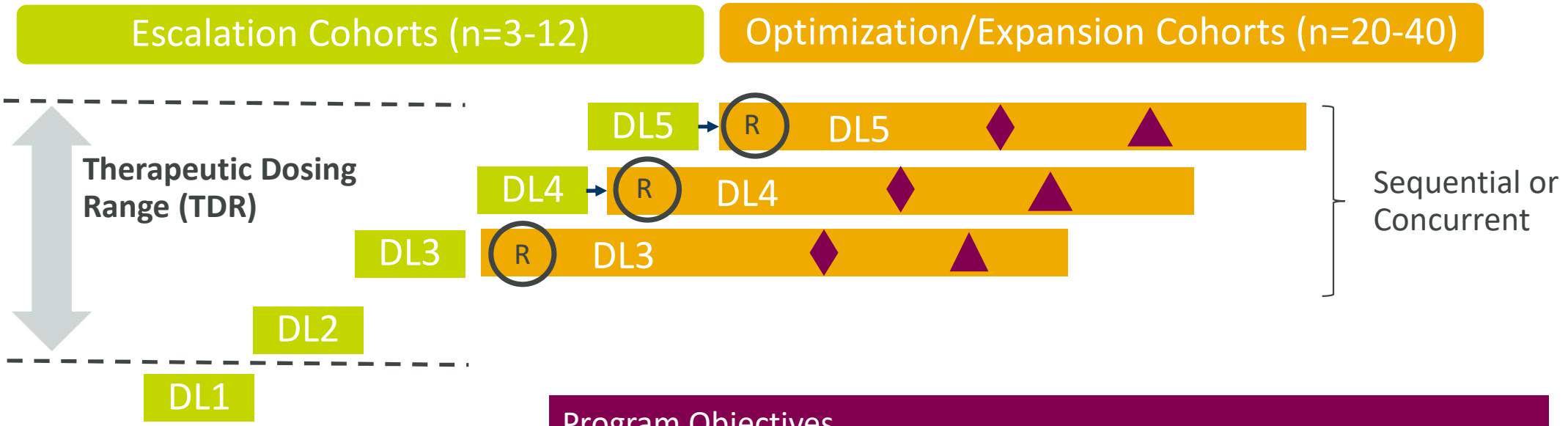
Presenter: Rosalind Hobson, Statistical Science Director, AstraZeneca, UK

Co-Authors: Wei Quan, Nairouz Elgeioushi

PSI conference, Beurs Van Berlage, Amsterdam,
18 June 2024

Introduction

Typical Phase I Oncology Study Design with Early Decision during Dose Expansion (N=20-40) - Overview



DL Dose Level

Early/multiple interim analyses

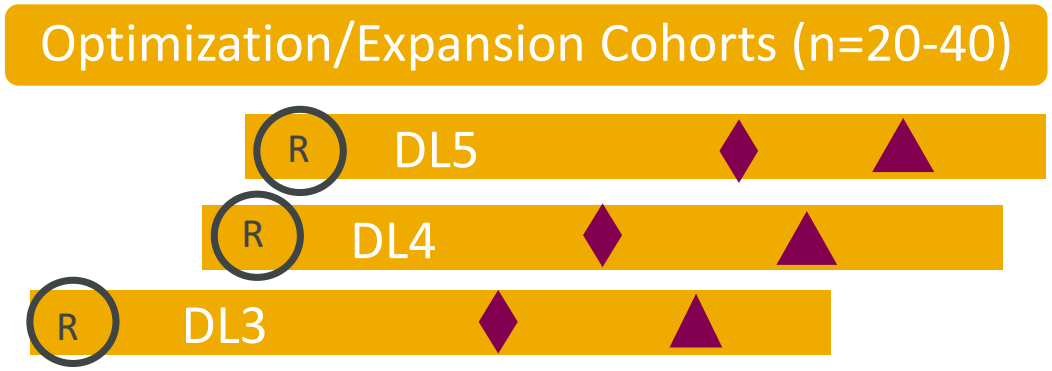
Primary analysis

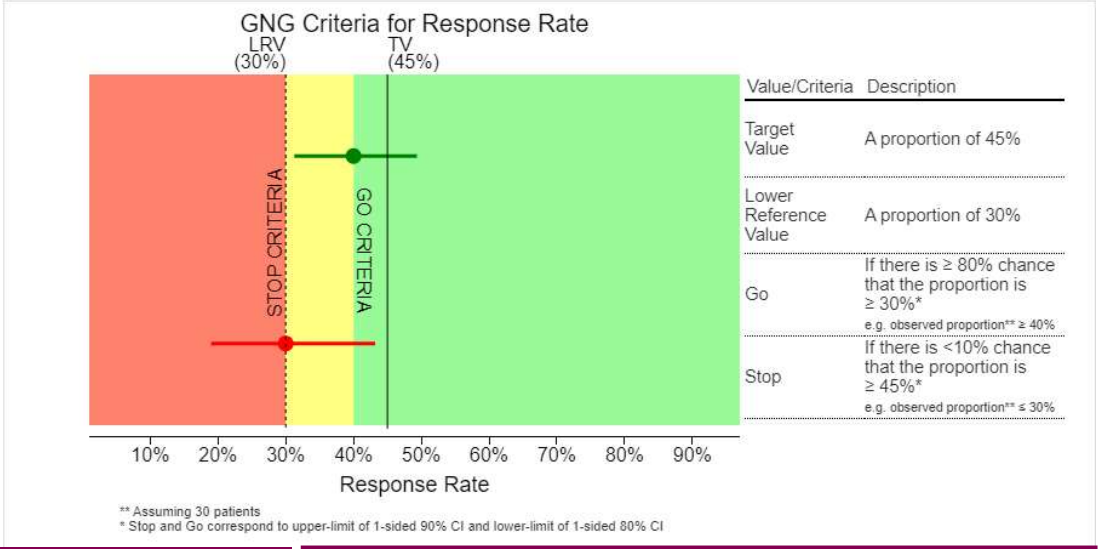
R Partial randomisation, when >1 DL in Exp.

Program Objectives

- **Escalation: Predict maximum tolerated dose, select doses within TDR and indications for further research**
- **Expansion: Recommended Phase 2 dose, Go/No Go for Phase 3 planning, Go/No Go for Indication**

Typical Phase I Oncology Study Design with Early Decision during Dose Expansion (N=20-40) - Lalonde framework





3 Outcome Lalonde Decision Framework (Stop/Consider/Go)

GO if there is ≥ 80% chance that the true outcome is ≥ LRV*

STOP if there is ≤ 10% chance that the true outcome is ≤ TV*

CONSIDER if neither STOP nor GO criteria are met.

Uses 1-sided Confidence Limits

In this example, a response rate of 40%, ie 12/30, has a lower 1-sided 80% CL which is above 30% (ie, LRV) so a response rate of 40% or above at ▲ would lead to a GO decision.*

*LRV: Lower reference value (smallest clinically meaningful effect) , TV: Target value

Lalonde framework: Frewer et al. 2016

Brief introduction to Early Oncology

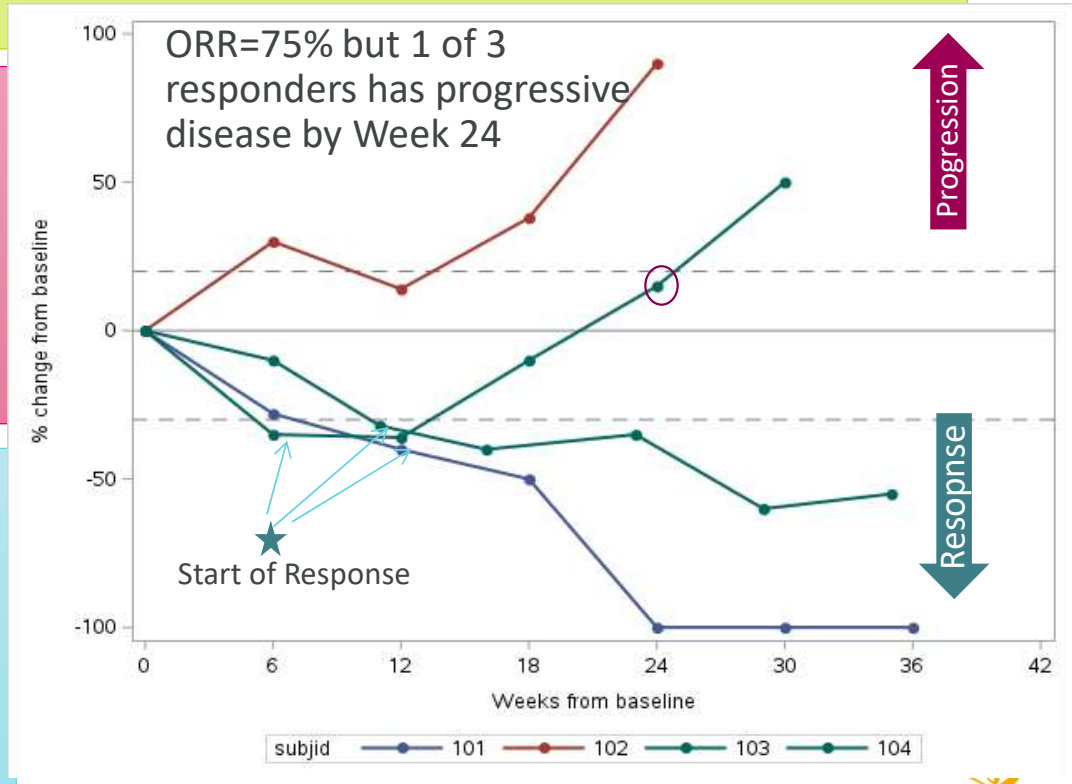
- Early Oncology trials are generally in PATIENTS (vs Healthy Volunteers)
- Potential for efficacy data at earliest stages of the development program
- Project Optimus – requirement to understand benefit/risk profile earlier than ever
- *Objective Response Rate (ORR) = binary variable where success is achieved if reduction in tumour size from baseline of 30% or more (much simplified definition of 'RECIST response')*
- *Progression Free Survival (PFS) = time from start of treatment to first sign of worsening (e.g. new lesion, growth of tumour >20%, death)*
- Registrational endpoints may include ORR or PFS but more typically OS (Overall Survival) which is often an unsuitable primary endpoint for Early Phase trials
- ORR is the preferred primary endpoint in early phase oncology trials

Advantages and disadvantages of ORR –Why add Durability?

- ✓ Outcome can be unequivocally attributed to drug (spontaneous shrinkage is not expected during natural course of disease) so ORR can be used without a control
- ✓ Fast (responses may occur from 1st scan, confirmed at second)
- ✓ Interpretable

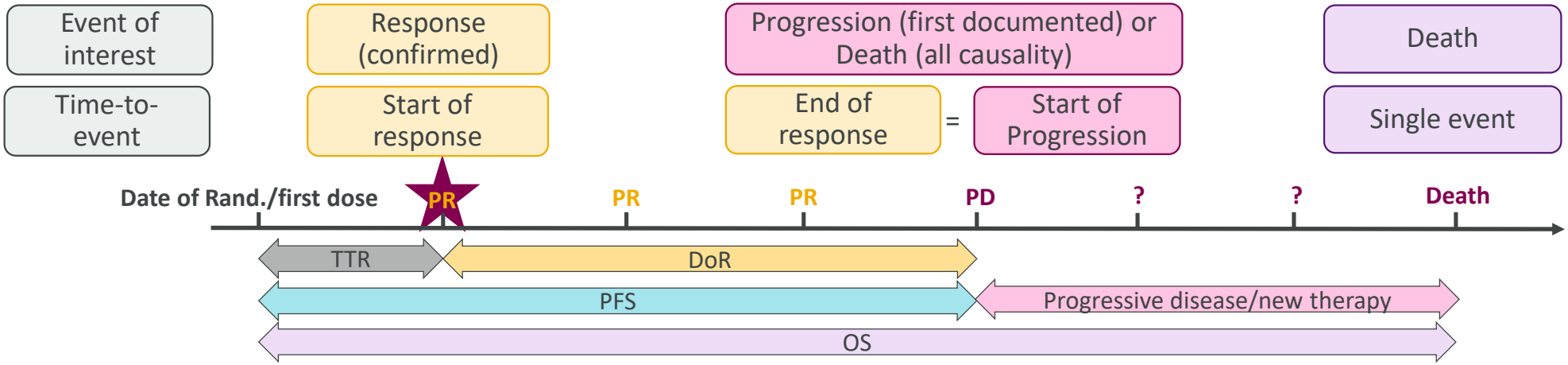
- × ORR doesn't always lead to tangible patient or clinical benefit
- × ORR and PFS don't always correlate with OS in comparative trials (Merino et al., 2023; Pasalic et al 202)

- ❖ Responses that last may improve QoL through delay of symptom worsening
- ❖ Durable responses may correlate with OS
- ❖ Measuring durability of response early in a project can de-risk later development



Deeper dive into durability

Durability Endpoints: Individual Patient Parameters



PR: Partial Response (e.g. tumour reduced from baseline by >30%)

PD: Progressive Disease (e.g. new lesions, or tumour increased by >20%)

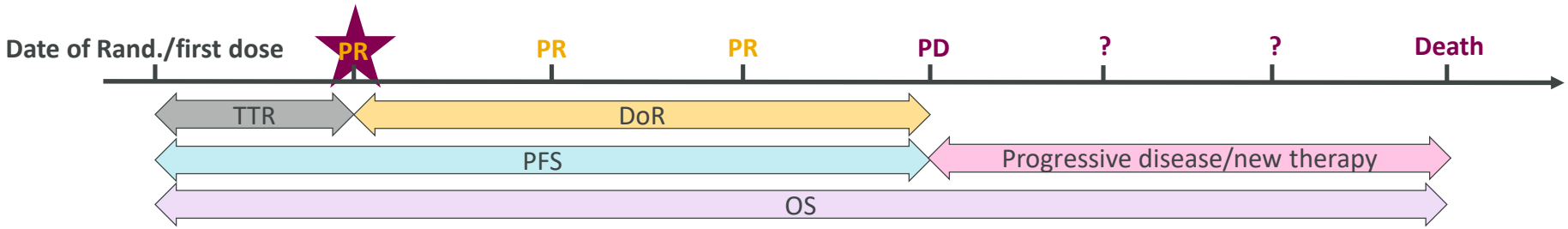
TTR: Time to Response

DoR: Duration of Response (Time to response until time to progression)

PFS: Progression-Free Survival (e.g. TTR + DoR)

OS: Overall Survival

Durability Endpoints: Summary Statistics



		Clinical Question
Conditional on Response	Median DoR	Among responders, what is then the DoR?
	Landmark DoR-6	What % of responders achieved a response that lasted at least 6 months?
All pts	Expected DoR/rmDoR	What is expected time in response for all patients?
	Durable Response DRR-X	What is proportion expected to have a response that lasts continuously for at least x months?
	PBIR-X	What is Probability of Being in Response (PBIR) at time X after start of treatment?
	Median PFS/Landmark PFS	How long is the progression-free survival time / Proportion progression free at time X ?

NB Definitions/Names are not all well-established. See references

Points to consider and case study

What to Consider when Choosing a Durability Endpoint for Go/No Go

How long will it take to get a meaningful read-out?

Time to response; Scan frequency; Expected duration of response

How relevant and interpretable is it?

Think about Clinical Question and apply an appropriate Estimand definition; Evidence of predictability of Phase 3 primary endpoints, e.g. correlation with OS; Availability of benchmarking; Confidence in shape of distribution

Are the operating characteristics reasonable?

If denominator is reduced to number of responders, then N may become too low to allow meaningful read-outs; Composite endpoints, e.g. DRR-6 may require TV & LRV for ORR to be wider than usual in order to maintain a reasonable consider zone in Lalonde; Overall risk of false stop/go should be considered when looking at a multi-factorial decision framework

Study AZ1: Case study – Choice of durability endpoint

A G/NG decision (s) is needed during Phase 2a expansion cohorts as to whether development of the compound in this specific disease should progress to Phase 2b/3, with potential for Accelerated Approval.

Target ORR is 65% (high) so it is statistically plausible to look at Duration of Response

How long will it take to get a meaningful read-out?	• Responses should happen by 12 weeks • Scans will take place every 9-12 weeks • Target median DoR is 12 months • Landmark DoR-6 would be ready 9 months post Day 0
How relevant and interpretable is it?	• PFS is expected to be very long (median > 2 years) as progression of disease is slow. Patients who do not respond will not necessarily worsen for a long time, but a response that endures will be clinically meaningful as it will delay further interventions
Are the operating characteristics reasonable?	• Minimum # responses for **GO** is 13 (max 21) which *may* be sufficient for DoR operating characteristics

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 1**

The landmark assessment of Durability of Response will be made 6 months after ORR.
DoR data *available at ORR* read-out will be assessed for **futility** against the landmark criteria.

Primary Analysis 12 weeks after initiation of therapy (ORR) N=21 TV=65%, LRV=50%

Go *(Unlock Phase 2)*: A ≥ 80% chance that the true rate is ≥ 50%,
e.g. observed rate is ≥ 62% (≥ 13 responders)

Consider: A >10% chance that the rate is ≥ 65% and <80% chance that the true rate is ≥50%, **e.g. observed rate between 48% and 62%, (i.e. 11,12 responders)**

Stop *(Do not proceed)*: A ≤ 10% chance that the rate is ≥ 65%,
e.g. observed rate ≤ 48% (≤ 10 responders)

Futility criteria for Duration of Response (at ORR timepoint) N<21, TV=mDoR 12 months

Stop *(Do not proceed)*: mDoR is calculable and upper 1-sided 90% CI is <12 months (TV), or already a clear STOP at Landmark DoR-6 analysis (see next slide,e.g. ≤6/13)

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 2**

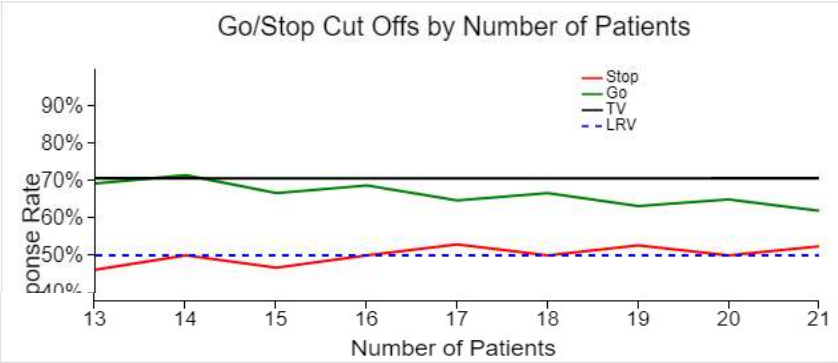
Standard Lalonde approach is used with LRV set to maximise probability of GO given TV, whilst also ensuring GO threshold corresponds as closely as possible to target mDoR.

Secondary Analysis Criteria
mDoR TV=12 mths, LRV=6 mths ⇔ DoR-6 TV=70.6%, LRV=50% (exponential) N=13

Go *(Unlock Phase 2)*: A ≥ 80% chance that the true rate is ≥ 50%, **e.g. observed rate still in response is ≥ 69.2% (≥ 9/13 responders)**

Consider: A >10% chance that the rate is ≥ 70.6% and <80% chance that the true rate is ≥ 50%, **e.g. observed rate between 46.1% and 69.2%, (i.e. 7,8 responders)**

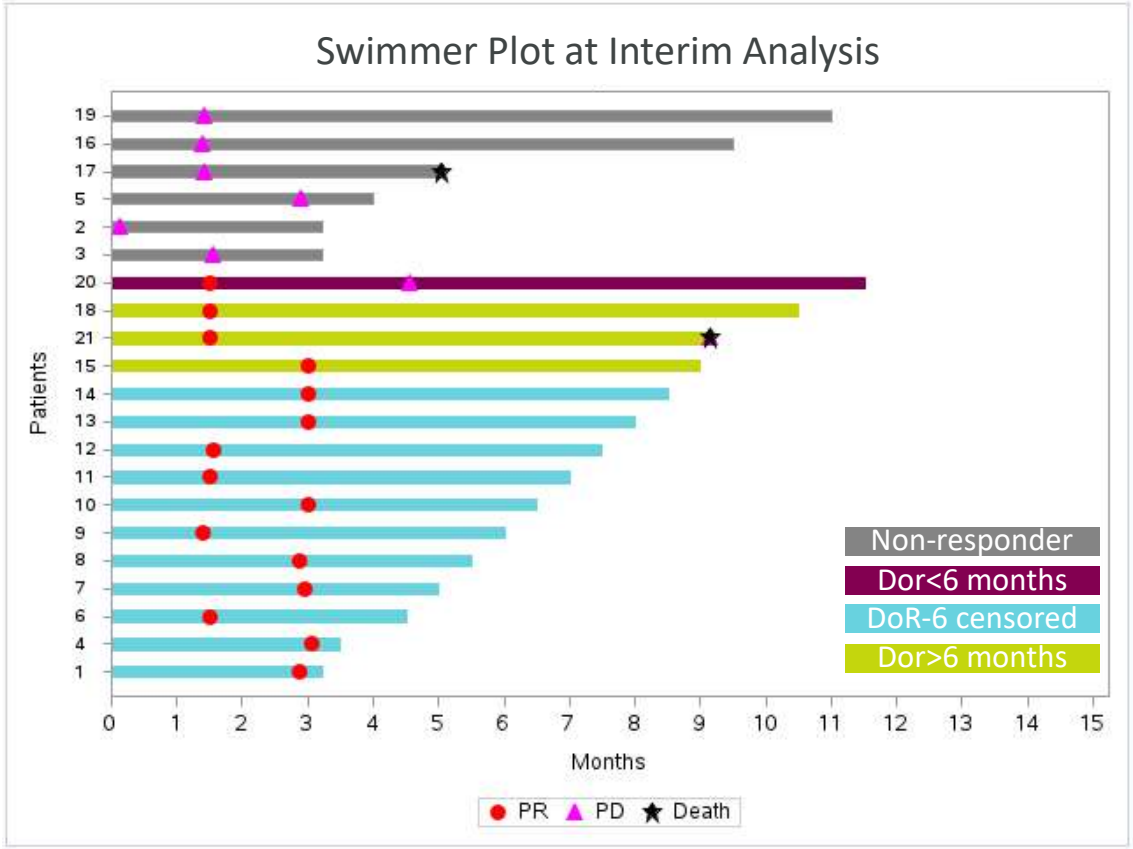
Stop *(Do not proceed)*: A ≤ 10% chance that the rate is ≥ 70.6%, **e.g. observed rate still in response ≤ 46.1% (≤ 6/13 responders)**



Given Step 1 decision was 'GO', N patients available for DoR-6 analysis could vary between 13 and 21

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 1**

Simulated Data – Good Duration of Response



ORR 15/21 (71.4%)

Lower 1-sided 80% CI is 60%

≥50% = **GO DECISION**

Duration of Response:

3/15 responses have lasted at least 6 months

11 are censored at 6 months

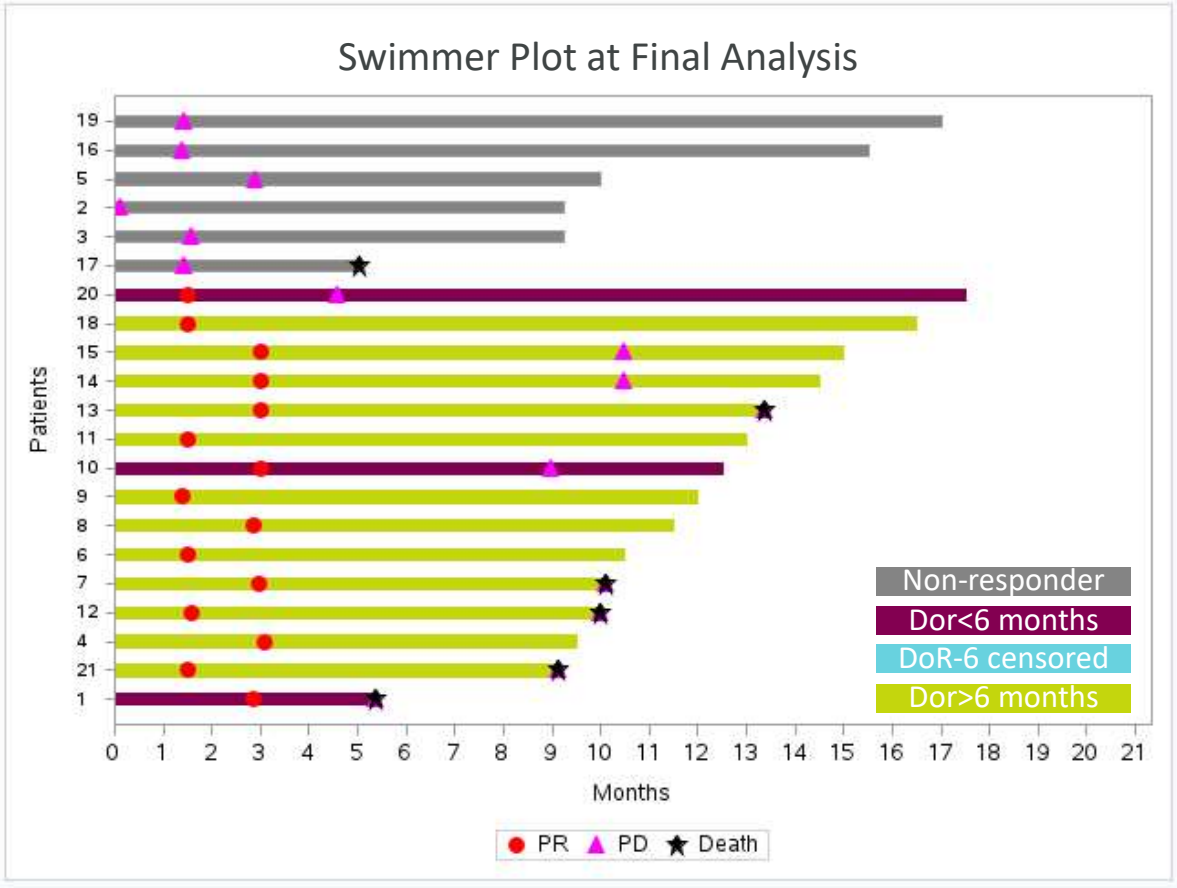
1 has DoR <6 months

Insufficient information to make a decision on DoR-6 at this point.

Wait 6 months

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 2**

Simulated Data – Good Duration of Response



Duration of Response

12/15 responses have lasted at least 6 months

3 had DoR <6 months

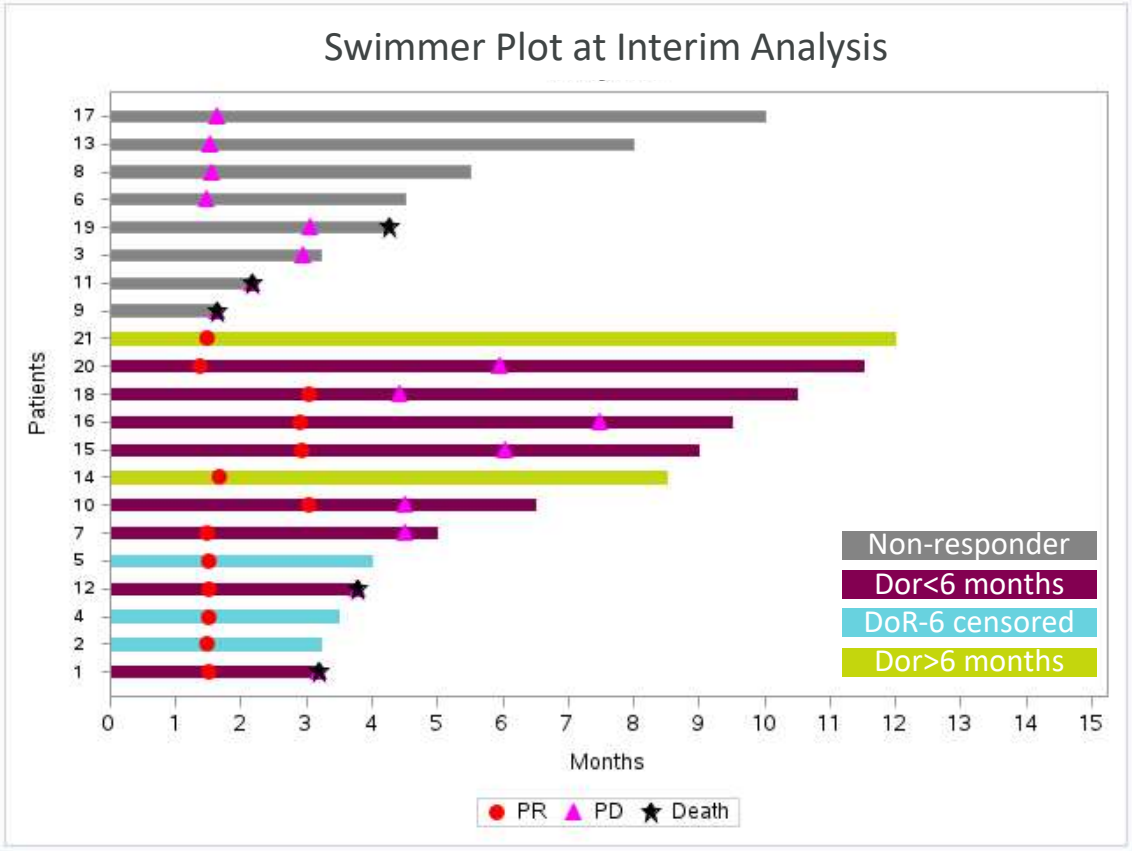
DoR-6 is estimated as 75%

Lower 1-sided 80% CL is 69.6% which leads to a GO decision

Median DoR is not yet mature, but is likely to be >12 months

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 1**

Simulated Data – Poor Duration of Response



ORR 13/21 (61.9%)

Lower 1-sided 80% CI is <u>50.3%</u>

≥50% = **GO DECISION**

Duration of Response- Futility

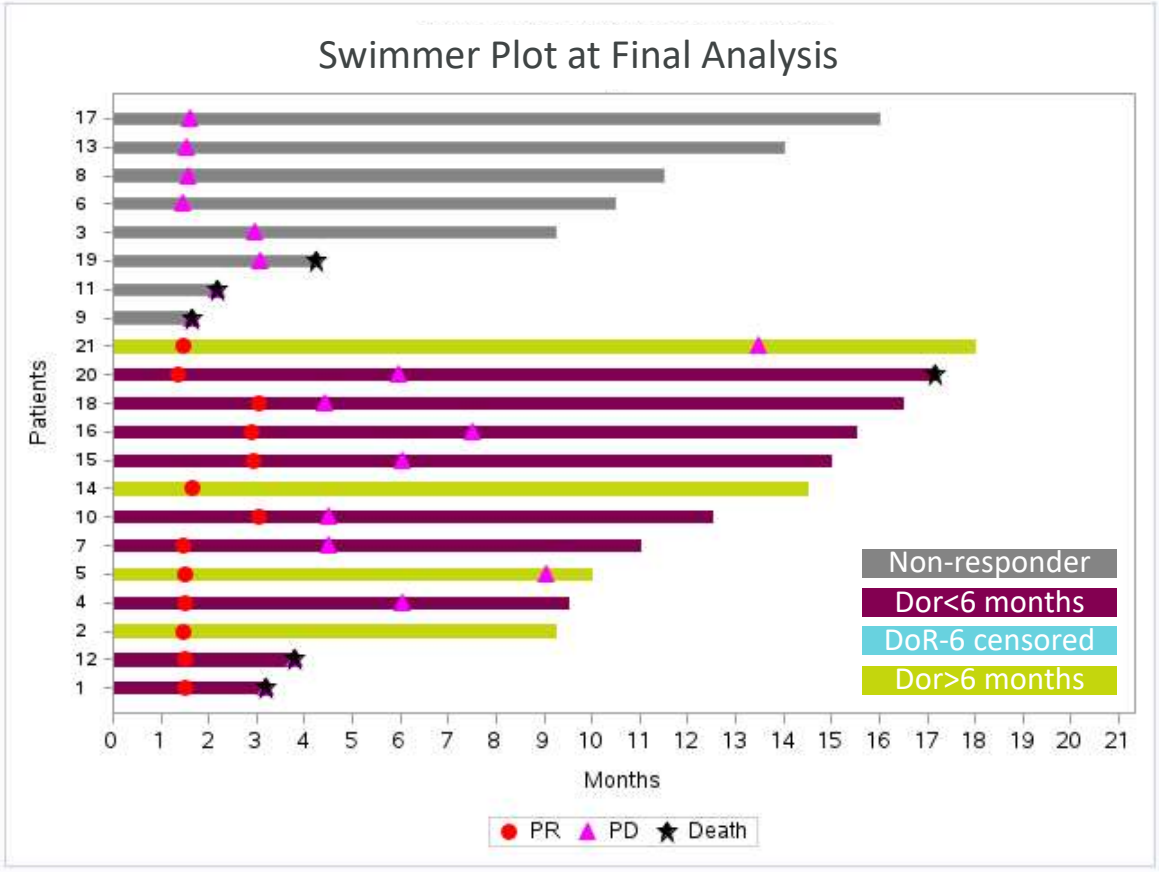
8 of 13 responses have already failed to last 6 months prior to interim analysis.

3 of 13 are censored, so best final outcome will be 5/13 which is **STOP**

mDoR is calculated as 3.1 months with upper one sided 90% CI 4.6 months. mDoR likely to be <12 months but dataset is not yet mature enough for a decision

Study AZ1:2-step decision process for ORR & Landmark DoR-6 **Step 2**

Simulated Data – Poor Duration of Response



Duration of Response – Futility Confirmed

2 of the pts with unknown outcome at interim now have a DoR>6 months

Patient 4 had progession of disease at 4.5 months, so DoR-6 is a failure.

Only 4/13 patients have DoR-6 success.

DoR-6 is 30.1%, with upper 1-sided 90% CI = 47.2%, which is a NO GO decision.

mDoR is calculated as 4.6 months with upper one sided 90% CI 4.6 months. Decision would be NO GO if we used median DoR instead of a landmark.

Study AZ1: Operating Characteristics for DoR-6 at ORR+ 6 months

The landmark assessment of Durability of Response will be made 6 months after ORR (N=21).

Decision criteria and operating characteristics under assumption of target mDoRs mDoR TV=12 mths, LRV=6 mths ≈ DoR-6 TV=70.6%, LRV=50%			
Number of responders (ORR)	Stop A ≤ 10% chance that the rate is ≥ 70.6%	Consider	Go A ≥ 80% chance that the true rate is ≥ 50%
13	<=6/13 **p(stop|TV)=5.6%** **p(stop|LRV)=50.0%**	7,8/13≈mDoR 6-10 mos **p(consider|TV)=27.1%** **p(consider|LRV)=36.7%**	>=9/13 **p(go|TV)=67.2%** **p(go|LRV)=13.3%**
21	<=11/21 **p(stop|TV)=6.0%** **p(stop|LRV)=66.8%**	12/21 ≈mDoR 7-8 mos **p(consider|TV)=7.4%** **p(consider|LRV)=14.0%**	>=13/21 **p(go|TV)=86.6%** **p(go|LRV)=19.2%**

Highlighted values show a Good Decision

Conclusion

Conclusions slide

- Early oncology trials provide unique challenges when basing decisions on multiple endpoints
- Early assessment of durability provides an additional measure on which to de-risk later development
- Careful choice of an appropriate endpoint should include consideration of how it may be interpreted

References

Frewer et al. Decision-making in early clinical drug development, Pharmaceutical Statistics 2016

Merino et al. Irreconcilable Differences: The Divorce Between Response Rates, Progression-Free Survival, and Overall Survival, Journal of Clinical Oncology. 2023

Weber et al. Duration of and time to response in oncology clinical trials from the perspective of the estimand framework. Pharmaceutical Statistics. 2023;1-16

Pasalic D et al. Progression-free survival is a suboptimal predictor for overall survival among metastatic sold tumour clinical trials. European Journal of Cancer. 2020;136:176-185

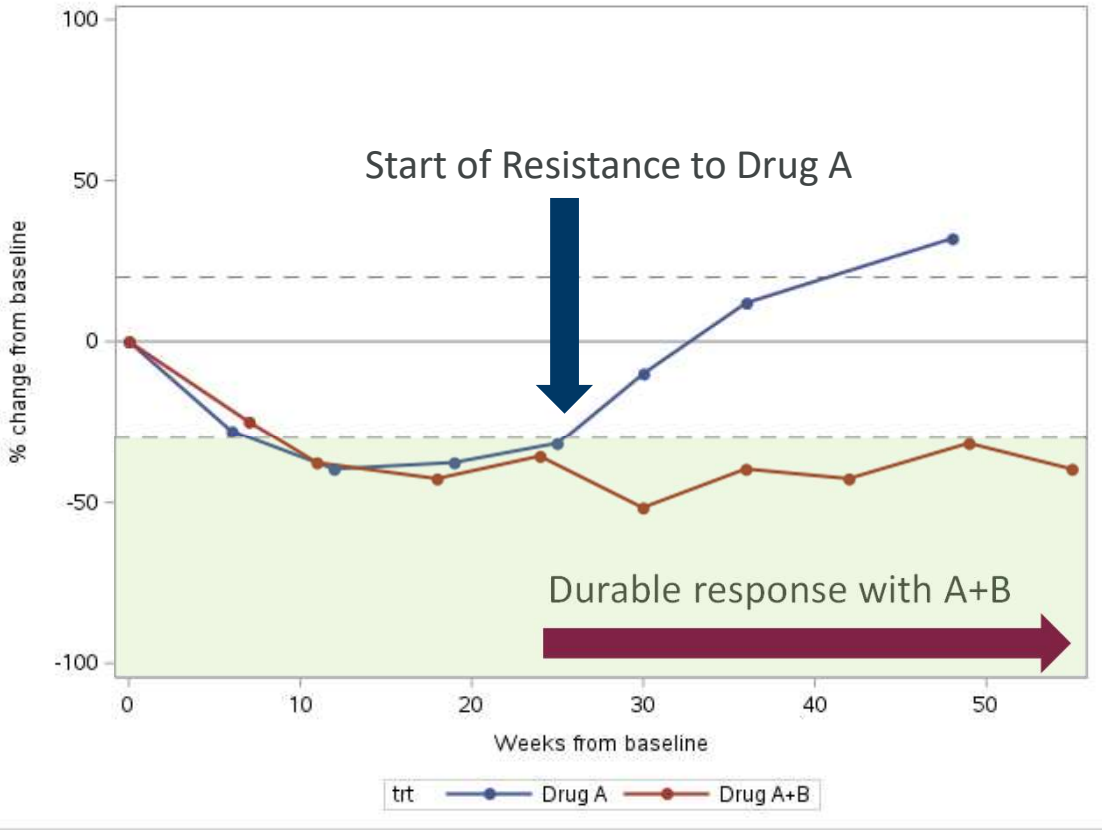
Ellis, Carroll, Pemberton. Analysis of duration of response in oncology trials, Contemporary Clinical trials 29, 2008 456-465

Huang and Tian. Utilizing restricted mean duration of response for efficacy evaluation of cancer treatments; Pharmaceutical Statistics. 2022;21:865-878

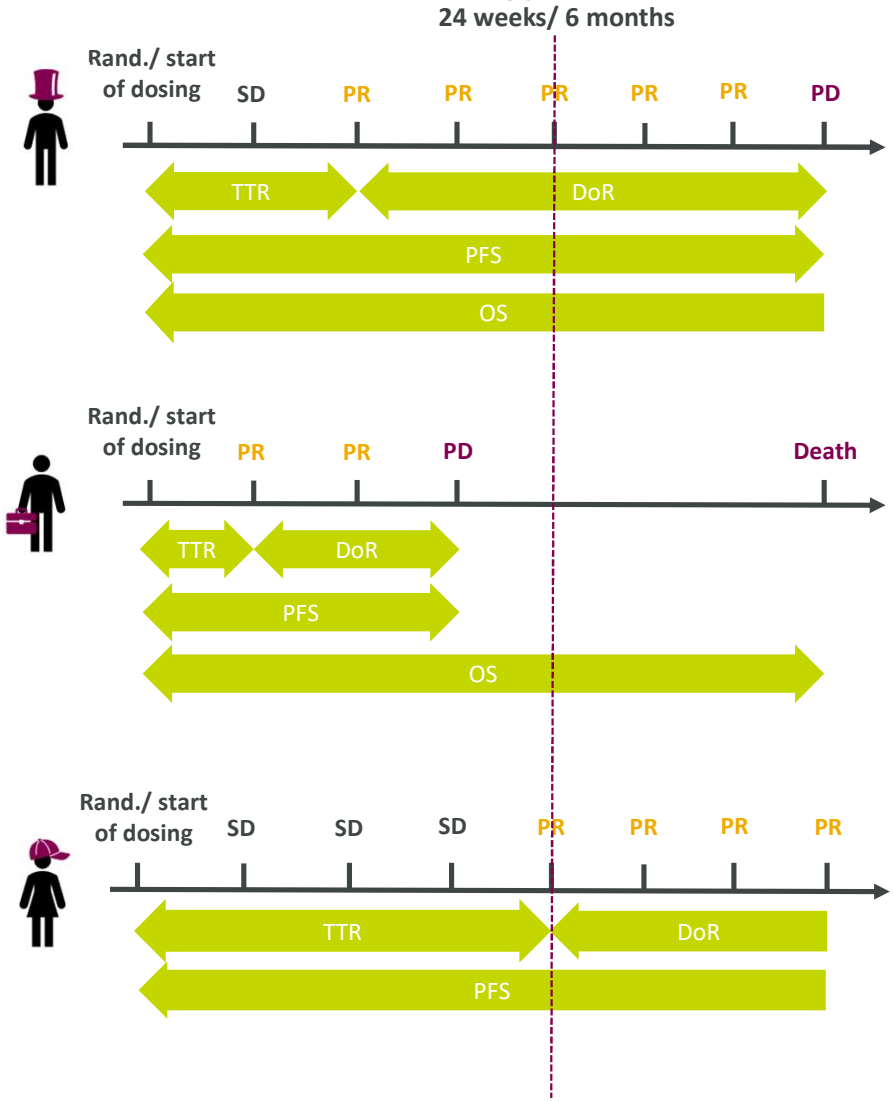
Back-ups

Why add Durability? – Combination therapies

- Better understanding of durability can help characterise resistance and anti-resistance mechanisms



Durability Endpoints 6 mos – Example of Responders



Example scan frequency 6 weeks,
ORR requires confirmed response according to RECIST 1.1 (single arm trial)

TTR=12 wks, DoR=30 wks, PFS=42wks, OS= censored 42wks.

- Landmark DoR-6='Yes'
- DRR-6='Yes'
- PBIR-6='Yes';
- Landmark PFS6='Yes';

Patient 1, responds at 12 weeks and response continues. Every durability endpoint shows a positive outcome.

TTR=6wks , DoR=12wks, PFS=18wks, OS=42 wks.

- Landmark DoR-6='No';
- DRR-6='No'
- PBIR-6=No'
- Landmark PFS6='No';

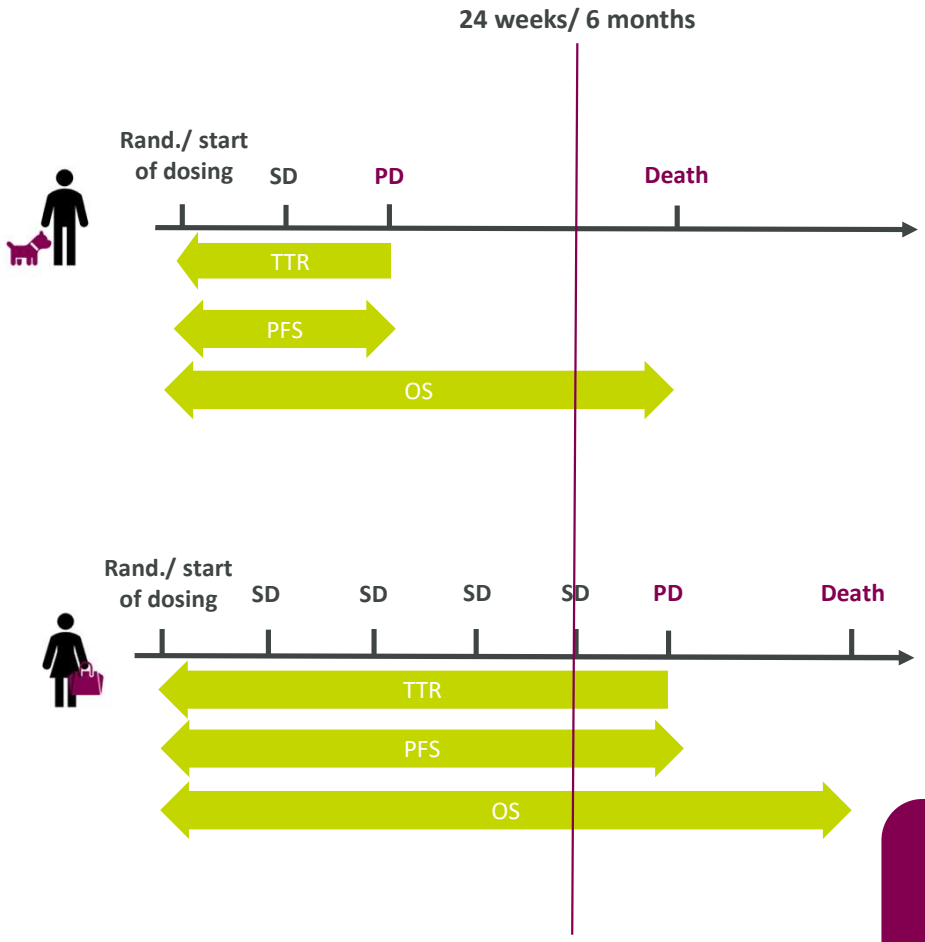
Patient 2, responds at first scan but response only lasts for 12 weeks. All durability endpoints are negative

TTR=24wks , DoR=censored 18wks, PFS/OS=censored 42 wks.

- Landmark DoR-6=Unknown
- DRR-6=Unknown
- PBIR-6='Yes'
- Landmark PFS6='Yes'

Patient 3, a late responder and response continues. PBIR-6 positive and landmark PFS-6 positive. Extend follow-up to see if DoR-6='yes' or DRR-6='yes'

Durability Endpoints 6 mos – Example of non-responders



Example scan frequency 6 weeks,
ORR requires confirmed response according to RECIST 1.1 (single arm trial)

TTR=censored 12wks, DoR=NA, PFS=12wks/3mos, OS=30wks

- Landmark DoR-6=NA
- DRR-6='No'
- PBIR-6='No'
- Landmark PFS6='No';

TTR= censored 30wks, DoR=NA, PFS=30wks, OS=42wks

- Landmark DoR-6=NA;
- DRR-6='No'
- PBIR-6='No'
- Landmark PFS6='Yes';

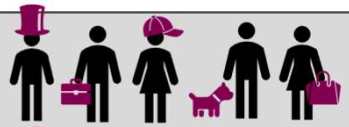
Patient 4 and 5, do not respond. These patients do not contribute to the denominator in a DoR endpoint*

Both patients are DRR-6 negative

Landmark PFS-6 may be preferable depending on clinical question

* Non-responders may or may not contribute to the denominator in a DoR endpoint depending on Estimand definition. The most commonly used Estimand uses # responders as denominator

Durability Endpoints – Summary

Population level summary		
FAS N	5	
Responders	3 (ORR=60%)	
Landmark DoR-6 Rate	(1 + 1 censored)/3	
DRR-6	(1 + 1 censored)/5	
PBIR-6	2/5	
Landmark PFS-6 Rate	3/5