Operating characteristics of tumor growth inhibition-overall survival models to support early Phase Ib decisions:

An evaluation in first-line metastatic non-small cell lung cancer patients treated with atezolizumab plus chemotherapy based on the Phase III study IMpower150

R Bruno, M Marchand[1], K Yoshida, H Li, W Zou, F Mercier, P Chanu, B Wu, A Lee, C Li, JY Jin

Genentech-Roche Clinical Pharmacology, Biostatistics and Product Development; [1]Certara Strategic Consulting

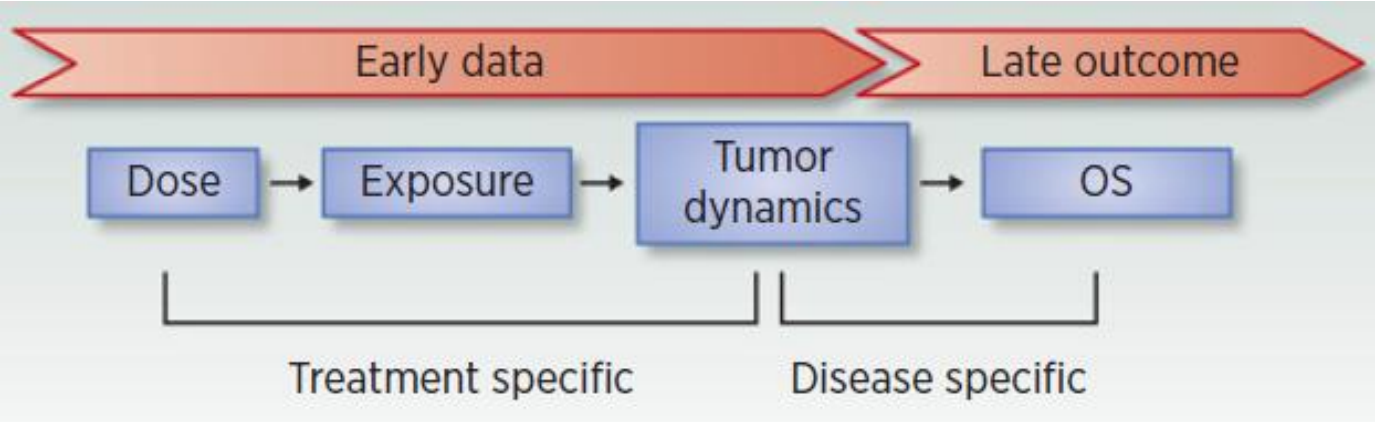
Dedicated to Laurent Claret

29th Virtual PAGE Meeting, Sept 6, 2021

CLINICAL CANCER RESEARCH | REVIEW Clin Cancer Res 2020;26:1787–95

Progress and Opportunities to Advance Clinical Cancer Therapeutics Using Tumor Dynamic Models

René Bruno[1], Dean Bottino[2], Dinesh P. de Alwis[3], Antonio T. Fojo[4], Jérémie Guedj[5], Chao Liu[6], Kristin R. Swanson[7], Jenny Zheng[8], Yanan Zheng[9], and Jin Y. Jin[10]

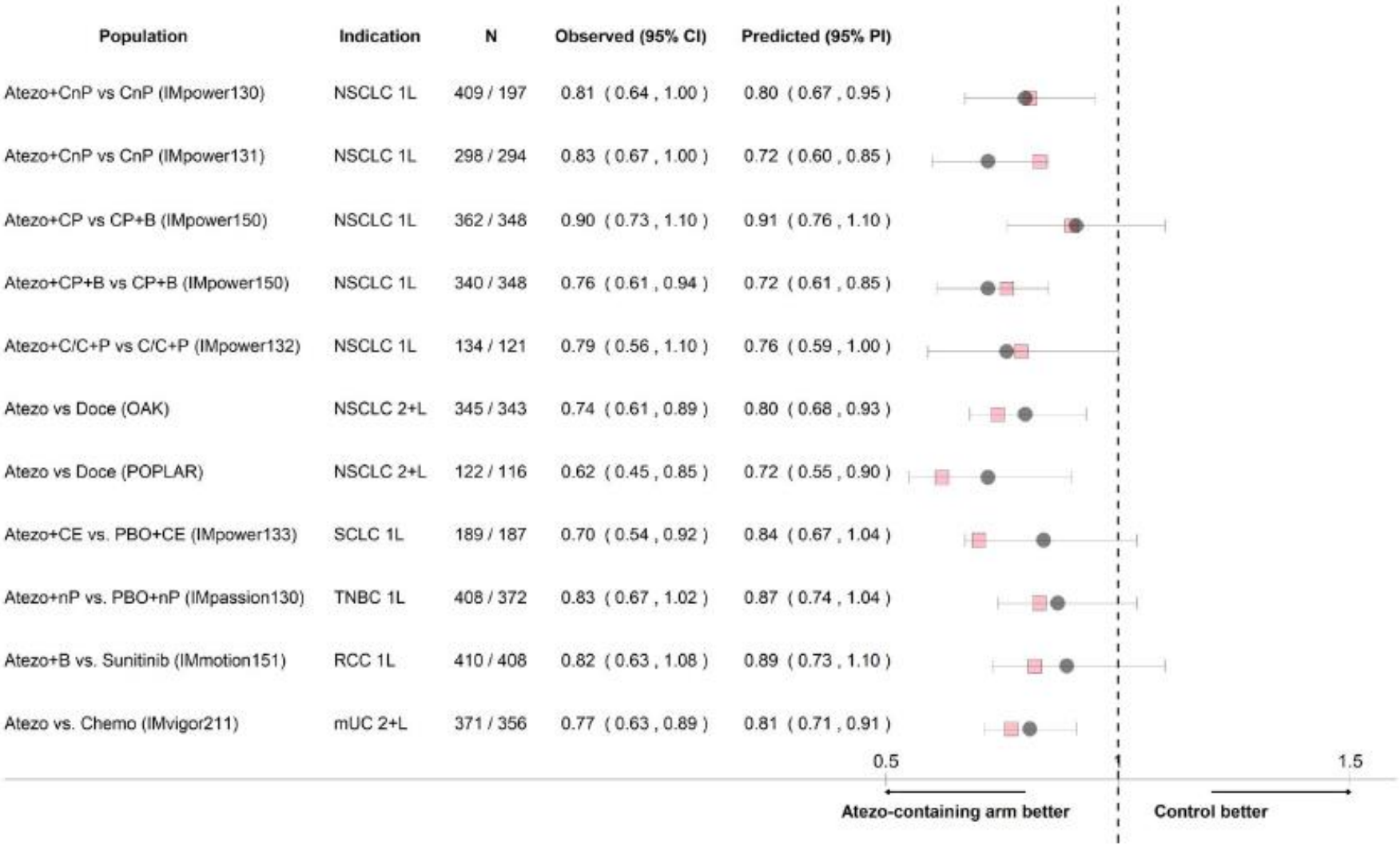


Model development (e.g. historical data)

Emerging tumor dynamic data → **OS simulation**

TGI-OS models predict atezolizumab vs. control HR in atezolizumab Phase II and III clinical studies

Population	Indication	N	Observed (95% CI)	Predicted (95% PI)
Atezo+CnP vs CnP (IMpower130)	NSCLC 1L	409 / 197	0.81 (0.64 , 1.00)	0.80 (0.67 , 0.95)
Atezo+CnP vs CnP (IMpower131)	NSCLC 1L	298 / 294	0.83 (0.67 , 1.00)	0.72 (0.60 , 0.85)
Atezo+CP vs CP+B (IMpower150)	NSCLC 1L	362 / 348	0.90 (0.73 , 1.10)	0.91 (0.76 , 1.10)
Atezo+CP+B vs CP+B (IMpower150)	NSCLC 1L	340 / 348	0.76 (0.61 , 0.94)	0.72 (0.61 , 0.85)
Atezo+C/C+P vs C/C+P (IMpower132)	NSCLC 1L	134 / 121	0.79 (0.56 , 1.10)	0.76 (0.59 , 1.00)
Atezo vs Doce (OAK)	NSCLC 2+L	345 / 343	0.74 (0.61 , 0.89)	0.80 (0.68 , 0.93)
Atezo vs Doce (POPLAR)	NSCLC 2+L	122 / 116	0.62 (0.45 , 0.85)	0.72 (0.55 , 0.90)
Atezo+CE vs. PBO+CE (IMpower133)	SCLC 1L	189 / 187	0.70 (0.54 , 0.92)	0.84 (0.67 , 1.04)
Atezo+nP vs. PBO+nP (IMpassion130)	TNBC 1L	408 / 372	0.83 (0.67 , 1.02)	0.87 (0.74 , 1.04)
Atezo+B vs. Sunitinib (IMmotion151)	RCC 1L	410 / 408	0.82 (0.63 , 1.08)	0.89 (0.73 , 1.10)
Atezo vs. Chemo (IMvigor211)	mUC 2+L	371 / 356	0.77 (0.63 , 0.89)	0.81 (0.71 , 0.91)

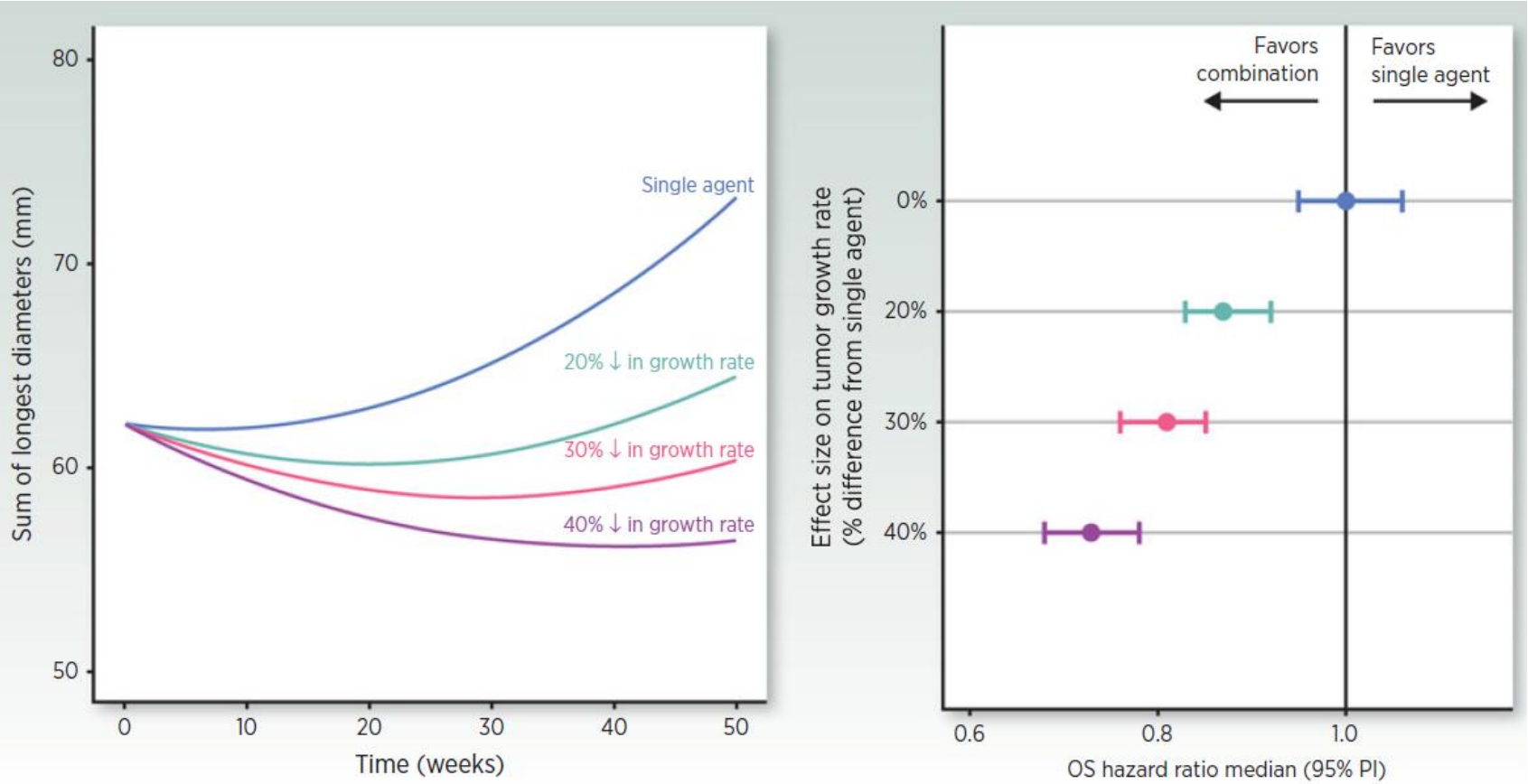


Model predictions (dots) and 95% prediction interval (1000 replicates, bars) with observed (squares)

Chan et al, CPT Pharmacometrics Syst Pharmacol. 2021;00:1–12

Background

- Typical tumor dynamic profiles following single agent or combination with a hypothetical agent that would reduce growth rate (KG) by 20%–40% and associated expected OS HR in randomized studies of the combination versus single agent
- We hypothesized that effect size in TGI metrics (relative to control) predicts for OS hazard ratio

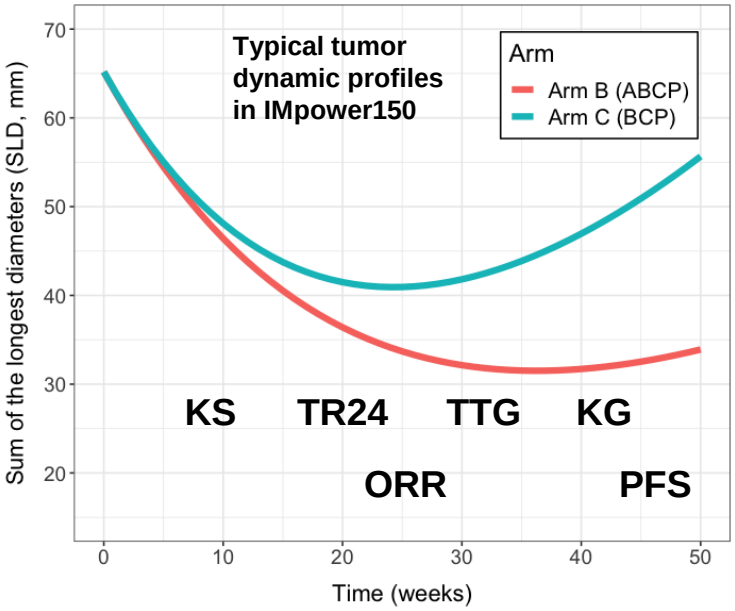


Bruno et al, Clin Cancer Res, 2020;26,1785-1795

Atezolizumab IMpower150 study in 1st-line NSCLC

- Atezolizumab in combination with bevacizumab plus carboplatin-paclitaxel (Arm B: ABCP) significantly improved PFS and OS compared to the control treatment (Arm C: BCP) (Socinski, N Engl J Med 2018;378, 2288-301)

- Typical tumor dynamic profiles showed a clear separation too and estimated KG predicted study outcome (OS distributions and hazard ratio) (Yoshida, ACoP 2019)



In black: tumor dynamic endpoints that may support early decisions

Methods

- Resampled tumor data from IMpower150
  - N=40 patients, 6 months recruitment, 24 weeks follow up with control
- Assessed probability to detect the effect **if true**
  - **Resampled in Arm B (ABCP) and Arm C (control, BCP)**
  - Estimated effect size based on based on **TGI metrics** (biexponential TGI model, Claret, CCR 2018)
    - Declared Go if significant difference (two-sided Wilcoxon test)
  - Predicted OS HR using integrated NSCLC TGI-OS model (Chan, CPT-PSP 2021)
    - Estimated effect size (HR)
    - Simulated a Phase III design (400 patients/arm resampled among the 40 patients)
    - Declared Go if significant difference (log-rank test)
  - **Correct go decision** (**power or sensitivity**) = % replicates with significant difference

X 500 replicates

- Assess probability to detect an effect **if absent (false positive)**
  - **Resampled in Arm C vs. Arm C**
  - **Incorrect go decision** (**Type I error or specificity**)
- Assessed **PFS** the same way

TGI metrics results

Metric	Median Effect size (%)	95%CI	Percent of success (%)[1]
KG (week^{-1})	-23.0	[-52.1 – 34.9]	73.8
KS (week^{-1})	6.7	[-37.0 – 75.0]	36.4
TR24	-19.3	[-38.1 – 3.49]	64.4
TTG (week)	18.8	[-24.9 – 87.2]	69.4

Power to detect a difference is fair except with KS, best with KG

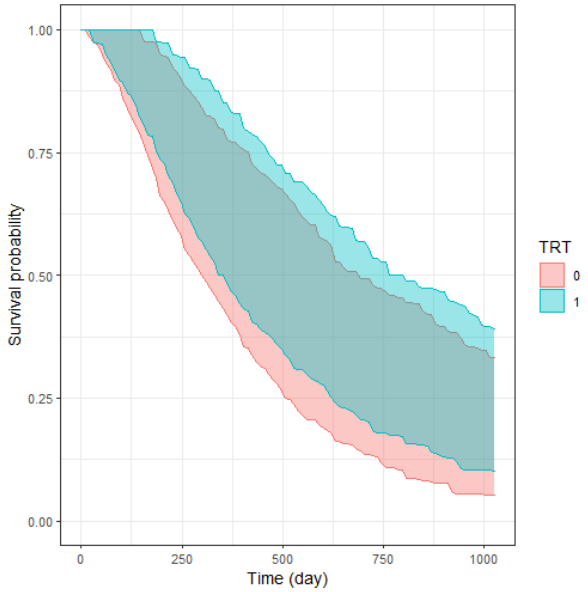
Metric	Median Effect size (%)	95%CI	Percent of success (%)[1]
KG (week^{-1})	0.3	[-13.1 - 17.8]	4.2
KS (week^{-1})	-0.4	[-22.7 - 27.8]	4.8
TR24	0.1	[-17.4 - 21.5]	4.4
TTG (week)	-0.4	[-15.5 - 14.9]	3.0

Type I error is acceptable

[1] % replicates with significant difference, two-sided Wilcoxon at $p<0.05$

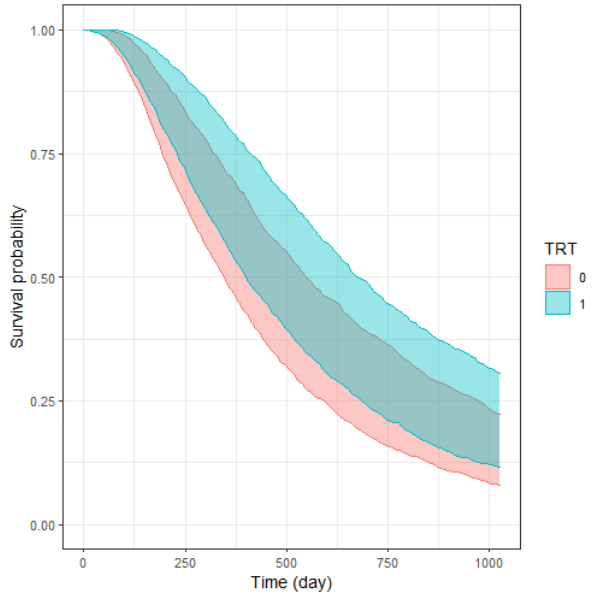
OS simulations

40 patients/Arm



HR predictions [95%PI]		
0.82 [0.47 – 1.57]		
Percent of success[1]		
p<0.05	**p<0.1**	**p<0.15**
18.4	25.8	32.2

400 patients/Arm



HR predictions [95%PI]		
0.82 [0.57 – 1.21]		
Percent of success[1]		
p<0.05	**p<0.1**	**p<0.15**
68.2	75.4	78.4

HR estimate quite uncertain

Power to detect a difference is fair in the Phase III simulations

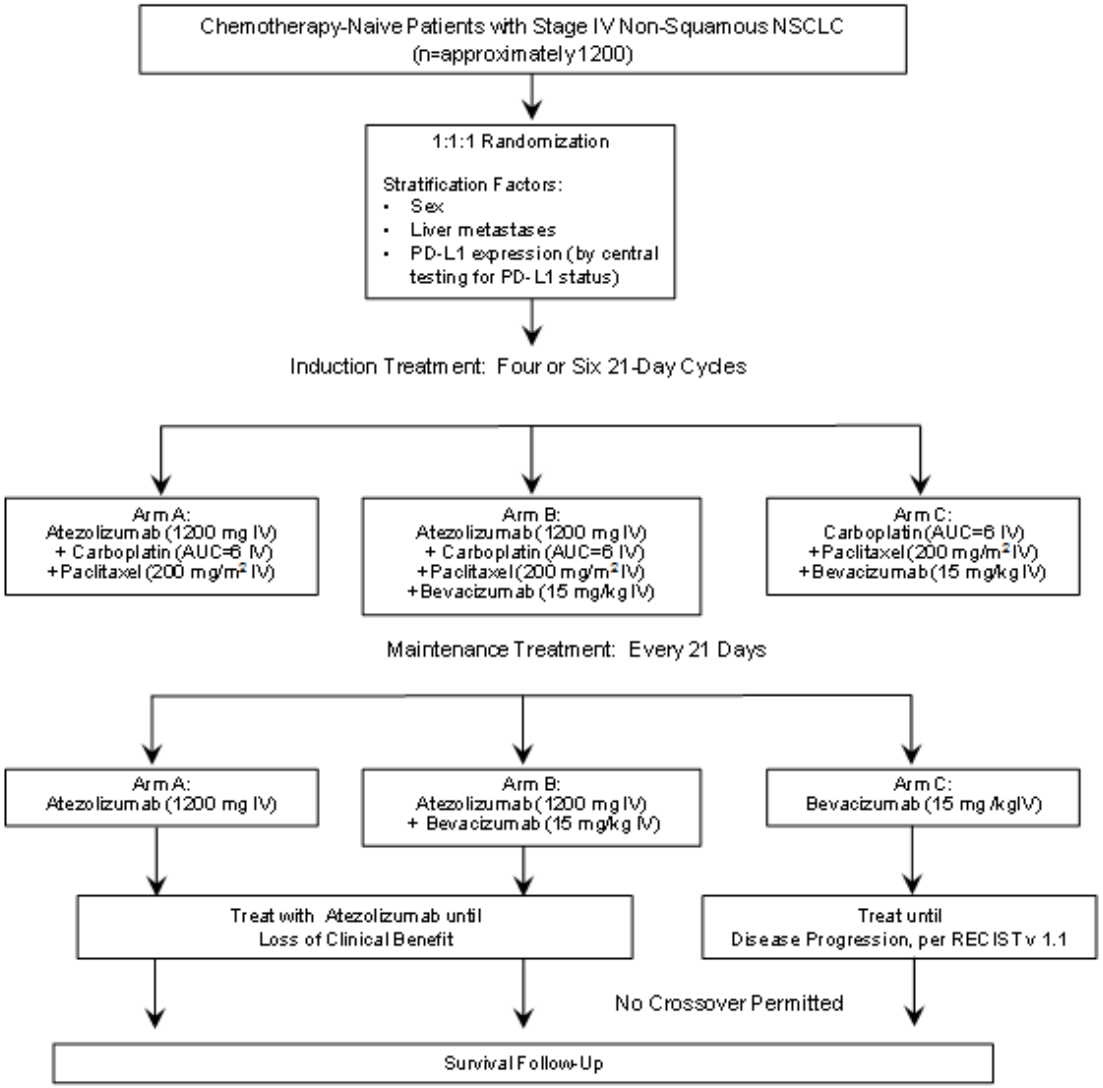
[1] % replicates with significant difference, two-sided log-rank

Conclusions so far…

- Data selection process based on IMpower150 mimics a Phase Ib design with control
  - N=40 patients, 6 months recruitment, 24 weeks follow up with control
- "Success" has been defined when TGI metrics were significantly different in experimental vs. control arm
  - As expected KG was the most sensitive metric to predict "success" but TTG and TR24 have good sensitivity too and could offer interesting, easier to estimate alternatives to KG
  - Type I error was close to 5% specified by the test
- OS simulations predicted a HR of 0.82, Phase III probability of success was fair for a 400 patient per arm trial
- The observed PFS analysis based on the same resampled and truncated data had a 28.0% power to show a difference across Arms (2-sided log rank test at p=0.05)
- **This evaluation suggests that model-based TGI metrics may be useful exploratory endpoints to inform early clinical decisions**
- Alternative designs are being investigated
  - Less patients (N=20, 30)
  - Shorter follow up (3 months)
  - Single arm study (options to generate virtual control are explored, Marchand, ACoP 2017)
- Expansion of this work is ongoing in other settings

Backups

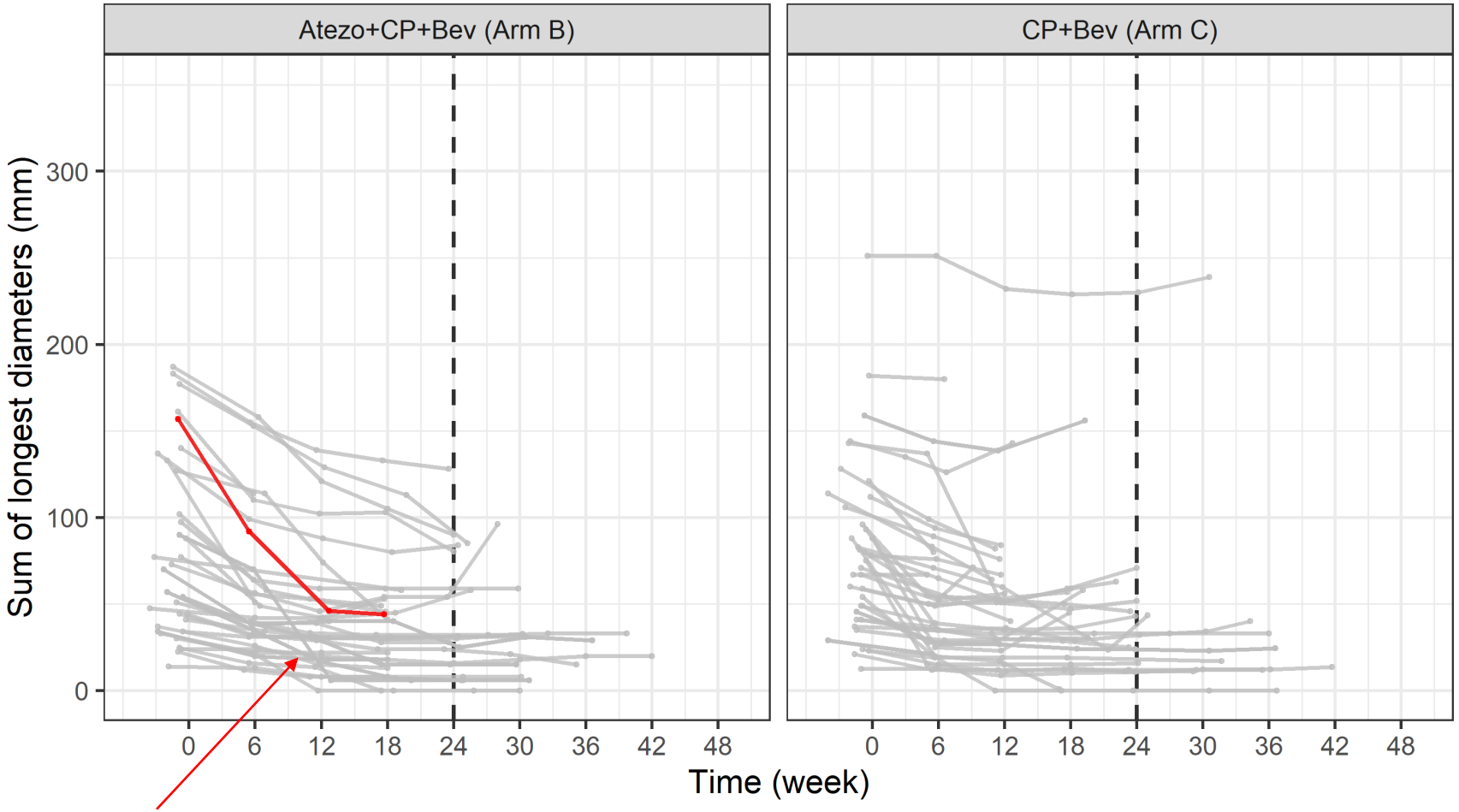
IMpower150 study design



Selection of data

- Select only Arm B (Atezo + CP+Bev) and Arm C (CP+Bev) data
  - Select TGI evaluable patients *i.e.* at least baseline and one post-treatment SLD assessment: 91 % of treated patients
- Arm B:
  - Rank the patients by date of first dose in each arm
  - Randomly sample one patient in Arm B
  - Define the 10-month period after first dose
  - Sample 39 patients with replacement
- Arm C:
  - Sample with replacement 40 patients in the same 6-month period
- Check the date of the SLD assessments for the last patient (selected in arm B) #40 and select the last assessment before 24w+1w weeks after first dose
  - Cut the dataset for SLD assessment visit after this date for the whole dataset (Arm B and Arm C)

Tumor profiles by Arm in one virtual analysis dataset



Tumor profile of the last patient selected
(SLD up to ~24w)

Follow-up and number of scans in the bootstraped datasets

- N 500 replicates

TRT	N replicates	Median FU (weeks)	Min FU (weeks)	Max FU (weeks)	Median Scan	Min Scan	Max Scan
Control	500	23.7	3.9	49.4	5	2	9
Atezo	500	24.8	-3.9	50.6	5	1	10

In the analysis datasets (N=500), the median follow-up is 24 weeks in both atezo and control arms, with a maximum of 50 weeks (~1 years).
The patients had a median of 5 scans for the tumor assessment.

Methods for TGI metrics

- The N datasets are used to estimate TGI metrics using Stein TGI model (biexponential model).
- The individual TGI metrics (KG, KS, TTG, and TR24) are summarized by treatment arms and replicates.
- An effect size for each metrics and each replicate is derived as follow:
  - $\frac{Metric\ Atezo - Metric\ Control}{Metric\ Control} * 100$
- A two-sided Wilcoxon test (α=5%) is performed on the TGI metrics of the two arms
- P-values of the test are used to derive the percent of success of each replicate if p-value < 0.05 (consider doing < 0.10)
- The percent of success is summarized for each metric
- To investigate the type I error, the same approach is used with randomly selected patients from Arm C compared to patients in Arm C.

TGI model estimates on one virtual analysis analysis dataset

Parameter	Estimate	SE	RSE	shrinkage	
KG[Atezo]	0.00740	0.0016	21.9	-	Atezo+CP+Bev
KS[Atezo]	0.0704	0.0068	9.72	-	
KG[control]	0.0137	0.0019	13.6	-	CP+Bev
KS[control]	0.0675	0.0099	14.7	-	
TS[0]	67	4.9048	7.32	-	
sigma^2	42.1	11.1456	26.5	-	
omega[KG, atezo]	0.611	0.2459	40.3	17.5	
omega[KS, atezo]	0.156	0.0469	30.1	23.4	
omega[KG, control]	0.261	0.1253	48	19.9	
omega[KS, control]	0.428	0.15	35	18.2	
omega[TS0]	0.41	0.0576	14.1	2.85	
omega[corrKGKS, atezo]	-0.0706	0.0831	118	-	
omega[corrKGKS, control]	-0.0443	0.1093	247	-	
Obj	2620	0	0	-	

Derivation of the % change ArmB versus ArmC on the individual TGI metrics values

N	parameter	ARMB	ARMC	%change
40	median.KG	0.00864	0.0142	-39.2
40	median.KS	0.0729	0.0710	2.7
40	median.TR24	0.44	0.651	-32.4
40	median.TTG	25.2	18.1	39.2

Derivation of the % change ArmB versus ArmC on the typical TGI metrics values

parameter	ARMB	ARMC	%change
KG	0.00740	0.0137	-46.0
KS	0.0704	0.0675	4.3

Methods for OS simulations

- Integrated NSCLC TGI-OS model (Chan, CPT-PSP 2021) has been re-estimated without IMpower150 data (no big change, see next slide)
- This model was used to simulate OS as follows:
  - 500 replicates of 40 patients by arm (Arm B vs. Arm C or Arm C vs. Arm C) with individual covariates and individual estimated KG
- For each of these scenario and across the 500 replicates, the followings analyses were preformed:
  - KM of OS distribution by arm
  - HR 95%PI
  - Percent of success using log-rang test at p-values of 0.05, 0.1 and 0.15

NSCLC TGI-OS integrated model

Pooled						without IMpower150				
	Value	**Std. Error**	**z**	**p**			**Value**	**Std. Error**	**z**	**p**
(Intercept)	3.47	0.173	20.1	8.41E-90		(Intercept)	3.41	0.201	17	6.44E-65
logKG	-0.616	0.0224	-27.4	1.35E-165		logKG	-0.621	0.0256	-24.3	4.46E-130
BCRP	-0.00385	0.000348	-11.1	1.75E-28		BCRP	-0.00374	0.000403	-9.29	1.60E-20
BECOG	-0.233	0.0298	-7.81	5.61E-15		BECOG	-0.179	0.0348	-5.14	2.74E-07
nsite5	-0.0764	0.0139	-5.51	3.53E-08		nsite5	-0.0723	0.0153	-4.73	2.21E-06
asian	0.244	0.0443	5.52	3.42E-08		asian	0.194	0.0508	3.82	0.000135
BALBUM	0.0135	0.00304	4.43	9.48E-06		BALBUM	0.0135	0.0036	3.74	0.000184
ICTC	0.119	0.0286	4.16	3.20E-05		ICTC	0.131	0.0331	3.96	7.64E-05
BLDH	-0.000141	4.00E-05	-3.53	0.00041		BLDH	-0.00012	4.17E-05	-2.8	0.00515
BNLR	-0.009	0.00262	-3.44	0.000582		BNLR	-0.0165	0.00379	-4.36	1.27E-05
line	-0.109	0.0341	-3.2	0.00138		line	-0.103	0.0357	-2.88	0.00403
mliver	-0.118	0.0401	-2.94	0.00332		mliver	-0.121	0.0452	-2.68	0.00734
Log(scale)	-0.264	0.0161	-16.3	4.95E-60		Log(scale)	-0.276	0.0185	-14.9	2.82E-50