Simulating to aid decision making

Jane Temple (GSK)

outline

Two case studies using simulation

1. Go/ No Go decision rules (programmed in R)
2. Impact of partial data on the operational characteristics of stopping rules for interim analyses (programmed in SAS)

All work presented here was sponsored by GSK

Go-No/Go decision rules

Background

- Amyotrophic lateral sclerosis is a motor neuron disease characterised by muscle atrophy, difficulty in speaking and respiratory problems
- Median survival time of approximately 2.5 years from onset[1]
- Expect 10% of subjects to die during a 1 year study
- Designed Phase II study with N=296 subjects

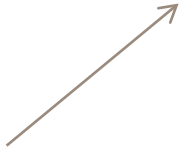
[1] Talbot K. Motor neuron disease: the bare essentials. Practical neurology. 2009; **9**(5): 303-9.

Combined assessment of function and survival (CAFS)

- Functional endpoint: ALSFRS-R
  - Patient reported
  - Scored from 0-48 (48 full health)
- Functional endpoint analysed using an MMRM are biased by missing data not missing at random
- CAFS: composite score of function and survival
  - compares each subject with every other subject
  - Survival better outcome than death
  - Small decline in function is a better outcome than a large decline in function
- Subjects compared at their last common visit
- Ranks the subjects based on their outcome

Calculating the CAFS score

Subject	ALSFRS-R score	Survival outcome
1	-12	Survived
2	-9	Survived
3	-5	Died (5 months)
4	-14	Died (7 months)
5	-11	Survived



Subject i	Subject j	score
1	2	-1
1	3	1
1	4	1
1	5	-1
CAFS score		0

Calculating the CAFS score

Subject	ALSFRS-R score	Survival outcome
1	-12	Survived
2	-9	Survived
3	-5	Died (5 months)
4	-14	Died (7 months)
5	-11	Survived

Subject i	Subject j	score
3	1	-1
3	2	-1
3	4	-1
3	5	-1
CAFS score		-4

Calculating the CAFS score

Subject	ALSFRS-R score	Survival outcome	CAFS score
1	-12	Survived	0
2	-9	Survived	4
3	-5	Died (5 months)	-4
4	-14	Died (7 months)	-2
5	-11	Survived	2

Score ranges from N-1 to 1-N

Impact of comparing subjects at their last common visit (LCV)

Subject	Change from BL at Visit					
	1	2	3	4	5	6
1	-1	-1	-3	x	x	x
2	0	-1	0	-2	-5	-7
3	0	-1	x	-4	-4	-6
4	x	x	x	x	x	x
5	-2	-4	-4	-5	-7	x

- Last common visit for Subject 1

Subject	LCV
2	3
3	2
4	BL
5	3

- Last common visit for Subject 2

Subject	LCV
1	3
3	6
4	BL
5	5

Purpose of the simulation

- To better understand the relationship between the two components of the CAFS
- To ensure that our go/ no go rules were appropriate

Simulation assumptions

- Need to simulation longitudinal functional scores at each of the 12 visits
- Need to simulate survival data: assumed exponential survival
- Need to simulate drop out subjects
- Assumed no correlation between function and survival data

- Each subject is compared with every other subject at their last common visit
  - 296 subjects = 31125 comparisons for each simulation

R coding

Loops within loops slow down the efficiency within R

If you get to 4 loops it will become very slow

Better to write a function, then apply that function to a matrix

```
cafs<-function(x, data, visits, names, lv=lv){

....

# returns the score for subject with details x
return(sum(score))
}
```

```
subj<-as.matrix(cbind(change, cen, os, strata, final.visit),
          colnames=c(names, "cen", "os", "strata", "final.visit"))

score<-apply(subj, 1, cafs, data=data, visits=visits,
           names=names, lv=lv)
```

Simulating to understand the components of the CAFS

- Made some assumptions about the placebo response for survival and functional decline based on meta analysis of placebo response from previous studies
  - Survival: 10% mortality
  - Function: 1.02 decline per month
- Set up a grid of values within a clinically relevant range
  - Survival: 5% increase to 10 % reduction, steps of 3%
  - Function: 10% decrease in slope to 40% decrease in slope, steps of 5%
- Simulated 250 studies at each grid point
- Assigned each simulated study a colour based on the go/no go decision grid

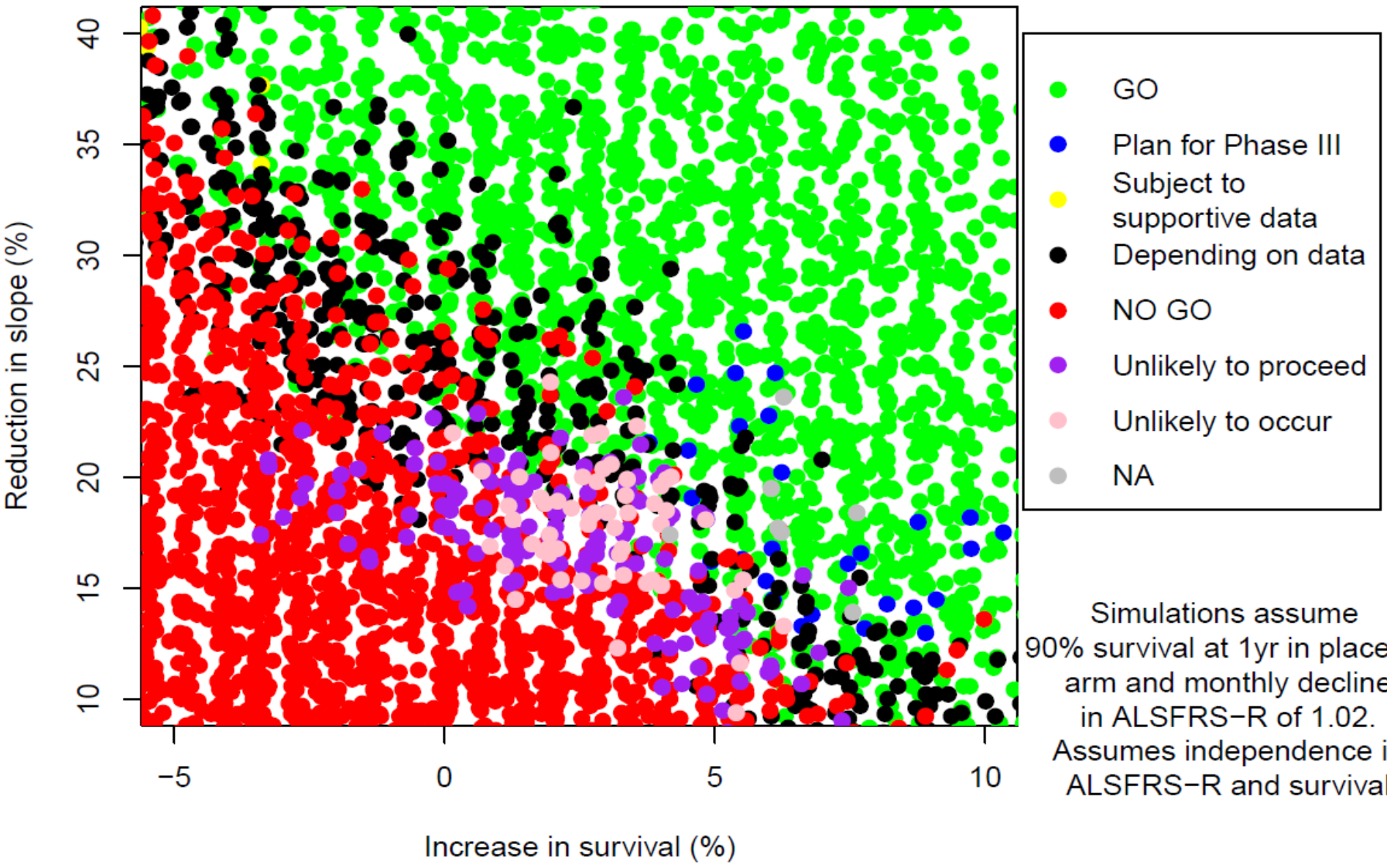
CAFS ranks significant

		Survival			
		+ve	Trend	No change	-ve
Function (ALSFRS-R)	+ve	GO	GO	GO	MAYBE? - subject to supportive data and secondary endpoints
	Trend	GO	GO	Unlikely to occur	NO GO / NA
	No change	GO	Unlikely to occur	NO GO / NA	NO GO / NA
	-ve	MAYBE? - subject to supportive data and secondary endpoints	NO GO / NA	NO GO / NA	NO GO / NA

CAFS ranks trend (p-value >0.05 &<=0.15)

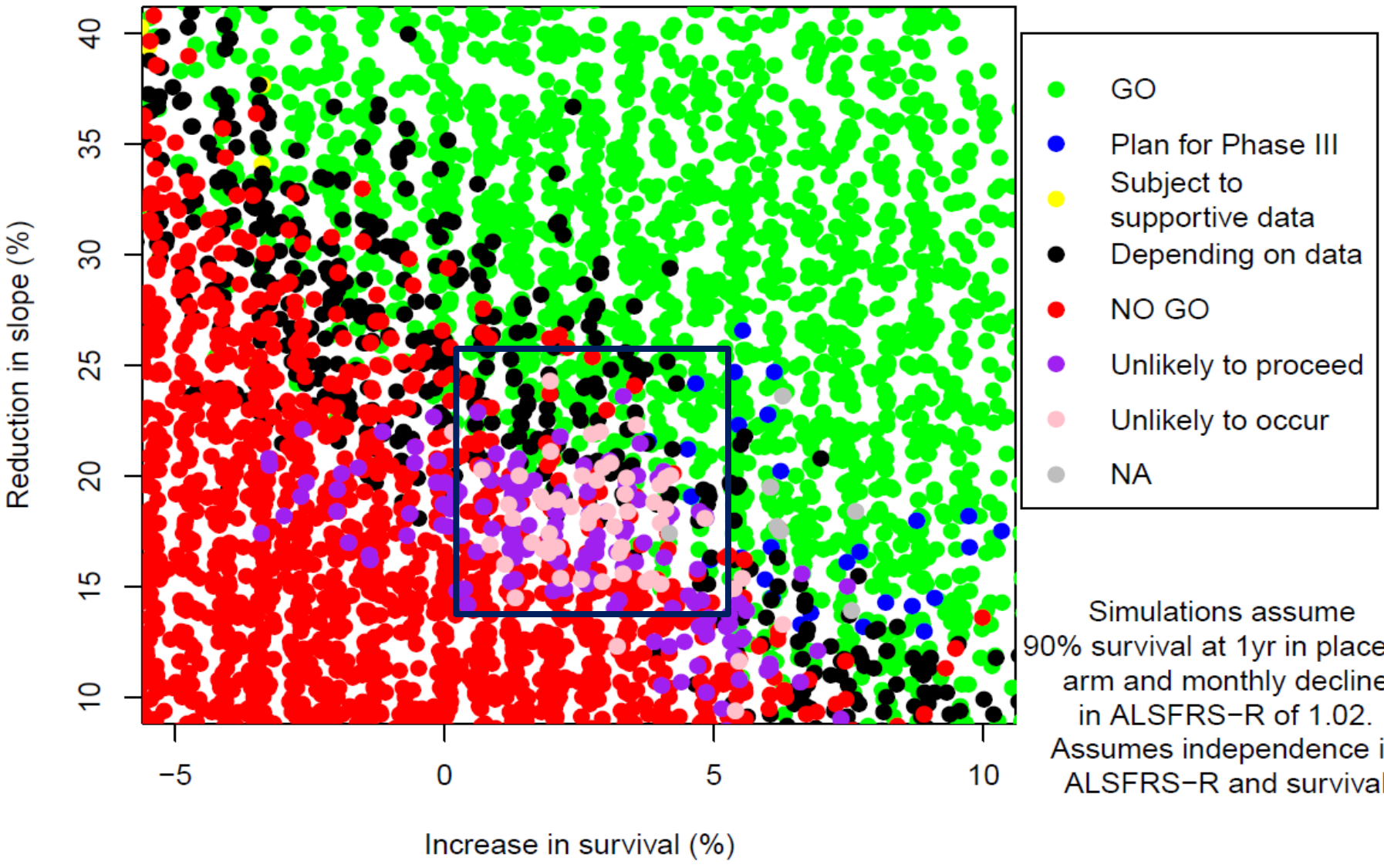
		Survival			
		+ve	Trend	No change	-ve
Function (ALSFRS-R)	+ve	N/A	Plan for phase 3	Maybe – depending on data	Maybe – depending on data
	Trend	Plan for phase 3	Maybe – depending on data	Unlikely to proceed	NO GO
	No change	Maybe – depending on data	Very unlikely to proceed	NO GO	NO GO
	-ve	Maybe – depending on data	NO GO	NO GO	NO GO

Application of Go / No Go Criteria Based on Simulated data



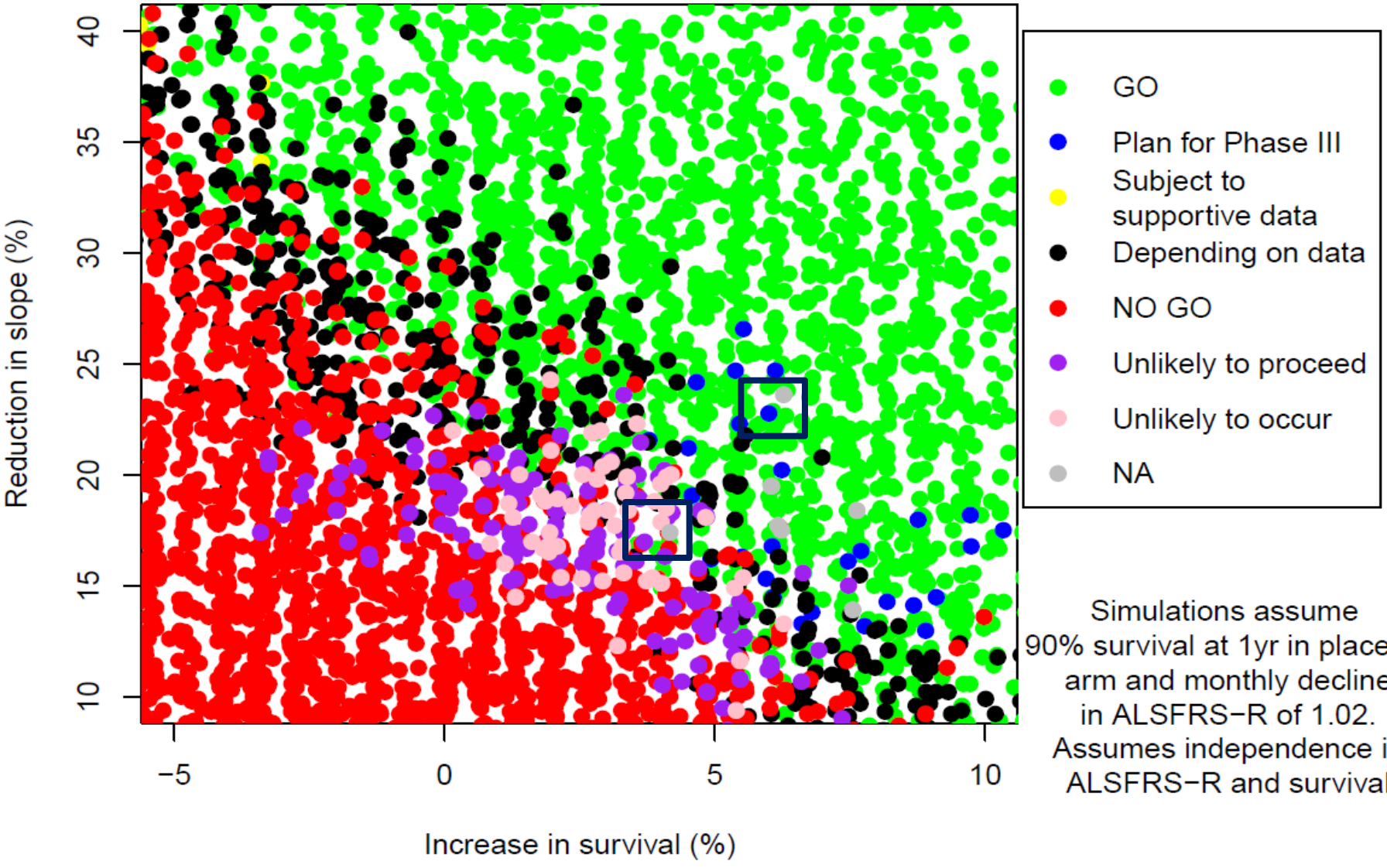
Simulations assume 90% survival at 1yr in placebo arm and monthly decline in ALSFRS−R of 1.02. Assumes independence in ALSFRS−R and survival.

Application of Go / No Go Criteria Based on Simulated data



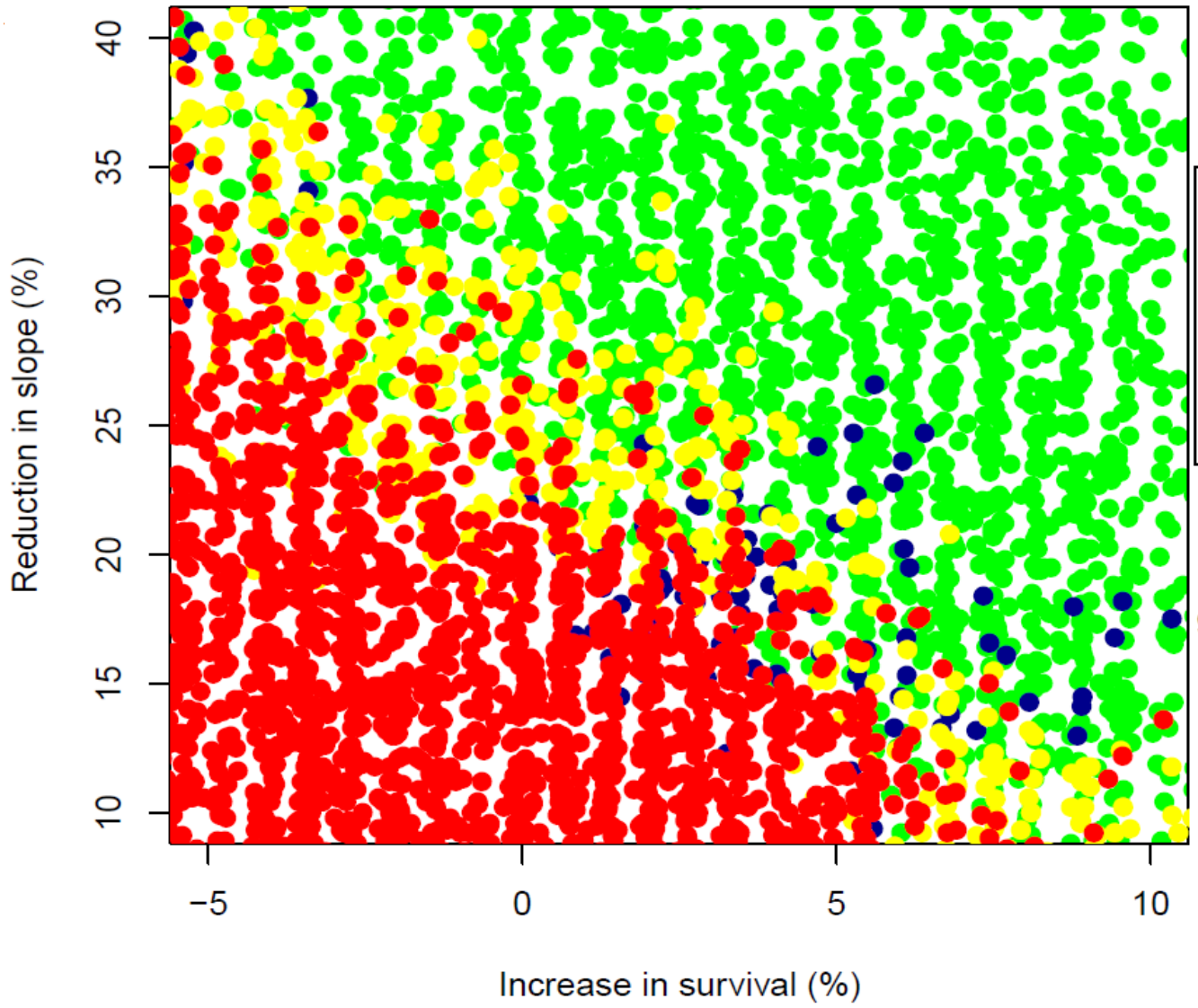
Simulations assume 90% survival at 1yr in placebo arm and monthly decline in ALSFRS−R of 1.02. Assumes independence in ALSFRS−R and survival.

Application of Go / No Go Criteria Based on Simulated data



Simulations assume 90% survival at 1yr in placebo arm and monthly decline in ALSFRS−R of 1.02. Assumes independence in ALSFRS−R and survival.

Application of Go / No Go Criteria Based on Simulated data



Simulations assume 90% survival at 1yr in placebo arm and monthly decline in ALSFRS−R of 1.02. Assumes independence in ALSFRS−R and survival.

Conclusions of go/ no go simulation

- The simulations helped the study team understand what magnitude of differences were needed in each of the components to see a significant CAFS score
- Allowed us to refine our go/ no go decision rules
- Allowed us understand where our competitors were in relation to these endpoints

Questions?

Impact of missing data on futility analyses

Background

- When designing a futility analysis, the study team want to understand the operational characteristics of the different rules
- When using predictive inference based rules, running simulations with patient level data are computationally expensive and time consuming
- Often people assume they have complete data at the interim as an approximation
- Questions to answer by simulating:
  1. When we have longitudinal data, what impact does the partial data have on the operational characteristics of the interim analysis?
  2. How does assuming you have complete data compare with the analysis incorporating partial data, that will actually be carried out at the interim analysis?

Spiegelhalter et al formula for calculating predictive power

Using a non-informative prior

$$p(S_{\varepsilon}^{C} \mid y_n, P(\theta)) = \Phi\left(\frac{\sqrt{n+m}}{\sqrt{m}} \frac{\sqrt{n}y}{\sqrt{2}\sigma} + \sqrt{\frac{n}{m}} z_{\varepsilon} \right)$$

Where:

- P(θ) The prior distribution. In this case uniform
- y is the observed mean difference
- σ is the observed standard deviation
- n is the number of subjects in each arm at the interim
- m is the number of subjects still to be observed
- Z_{ε} quantile from the standard normal distribution for a statistically significant result.
- S^{C}_{ε} is the event that a one sided p-value is less than ε under the null hypothesis

Spiegelhalter DJ, Abrams KR, Myles JP (2004) Bayesian approaches to clinical trials and health-care evaluation. John Wiley & Sons Ltd

3 approaches to estimate the predictive power

- Assume complete data only at the interim
  - Use formula to calculate the predictive power based on number of completers
  - Computationally cheap (can be done using summary statistics)
  - Doesn’t take into account partial data
- Use an MMRM including all partial data
  - Adjust the number of subjects observed in Spiegelhalter’s formula to account for partial data
  - Involves simulating patient level data, but computational cheap compared to Bayesian approach
  - Need some assumption of the correlation between visits
- Use a Bayesian approach (proc MCMC) including all partial data
  - Simulate missing data and calculated the predictive power
  - Computationally expensive
  - Reflective of what we actually do at the interim

MMRM approach with adjustment for partial data

Assuming equal variance across the two treatment arms. The standard error of the mean difference for the ANCOVA and MMRM respectively, are estimated as follows:

$$SE_{ANCOVA} = \sqrt{\frac{2\sigma^2}{n}}, \quad SE_{MMRM} = \sqrt{\frac{2\sigma^2}{\tilde{n}}},$$

n=subjects in each arm with complete data
N=total number of subjects in study

Rearranging we get

$$\sigma^2 = SE^2{}_{ANCOVA}\frac{n}{2} = SE^2{}_{MMRM}\frac{\tilde{n}}{2}$$

It follows that the adjusted sample size per arm including the partial data contributing to the MMRM is

$$\tilde{n} = \max\left(\frac{SE^2{}_{ANCOVA}}{SE^2{}_{MMRM}}n, N/2\right)$$

We can then use $\tilde{n},\ \tilde{m} = N/2 - \tilde{n}$ and the estimates from the MMRM in Spiegelhalter's formula to get the predictive power adjusted for partial data

MCMC approach

- Use proc MCMC to fit the longitudinal model with j=1,2 treatment groups and i=1,...,4 time points

$$Y_j \sim MVN(\underline{\theta}_j, \Sigma)$$

$$\theta_{ij} \sim N(0, 1e^6)$$

$$\Sigma \sim W^{-1}(I, 4)$$

- Use outpost to get samples of the predictive distribution from subjects with missing data
- Shared covariance matrix for the two treatment groups
- Start the MCMC at the estimates from the MMRM for both the θ and ∑

MCMC approach

- For each MCMC sample we can generate a predicted response for a subject at a missing time point

Sample	Subjid	Visit 1	Visit 2	Visit 3
1	1	1.25	2.56	3.78
2	1	1.25	2.56	3.78
3	1	1.25	2.56	3.78
1	2	0.859	1.59	2.37
2	2	0.859	1.81	2.15
3	2	0.859	1.62	2.29
1	3	1.89	2.67	3.56
2	3	1.89	2.12	3.56
3	3	1.89	2.78	3.56
1	4	1.64	2.98	3.82
2	4	1.64	2.98	4.01
3	4	1.64	2.98	3.66

Subject with complete data

Subjects where missing responses are predicted

Sample	Subjid	Visit 1	Visit 2	Visit 3
1	1	1.25	2.56	3.78
1	2	0.859	1.59	2.37
1	3	1.89	2.67	3.56
1	4	1.64	2.98	3.82

For each MCMC sample we then have a complete dataset and so can make end of study inferences for that sample.

MCMC approach

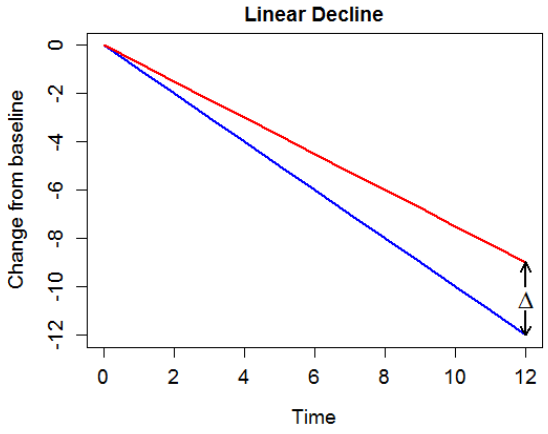
- To calculate the predicted power
  - Take the m=1,...,M samples for each missing data point and create M complete datasets
  - Run analysis on complete datasets
  - Predictive power=100*Number of datasets which are significant/M
- MCMC output is based on 10000 samples with a thin of 10

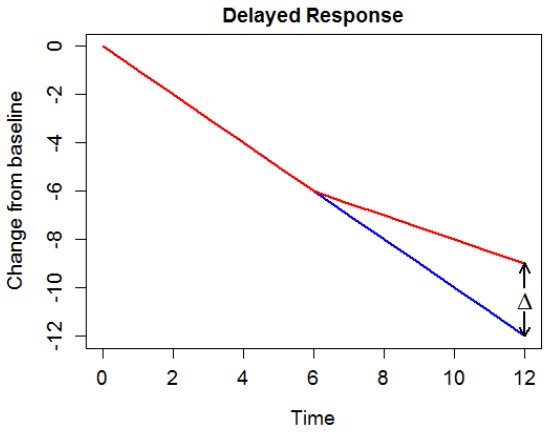
Simulation scenario

- N=140, interim after 50 subjects completed study
- Assume linear placebo decline of 1 point per 4 weeks
- SD of response=6
- Longitudinal data generated at weeks 12, 24, 36 and 48
- Stop for futility if predictive power is <=10%
- Currently simulations have included no missing data due to early withdrawals
- 500 simulations per scenario

Aim of simulations is to explore:

- Impact of length of recruitment
  - (40, 60, 80 weeks)
- Impact of correlation used to simulate repeated data
  - (Matrix 1: $\begin{pmatrix} 1 & 0.45 & 0.45 & 0.45 \\ 0.45 & 1 & 0.6 & 0.6 \\ 0.45 & 0.6 & 1 & 0.7 \\ 0.45 & 0.6 & 0.7 & 1 \end{pmatrix}$ Matrix 2: $\begin{pmatrix} 1 & 0.65 & 0.65 & 0.65 \\ 0.65 & 1 & 0.8 & 0.8 \\ 0.65 & 0.8 & 1 & 0.9 \\ 0.65 & 0.8 & 0.9 & 1 \end{pmatrix}$)
- Linear in decline
- Delay in response





Some notes on the actual simulation

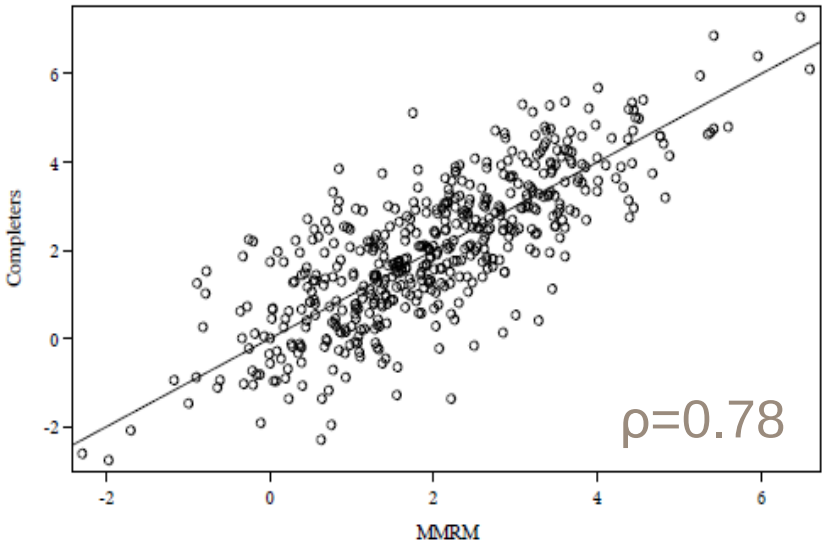
- 4 time points, 140 subjects, 500 simulations, 1000 samples from the posterior
- Added in loop around the mcmc analysis
- Need to change seed for each loop
- Not much more inefficient

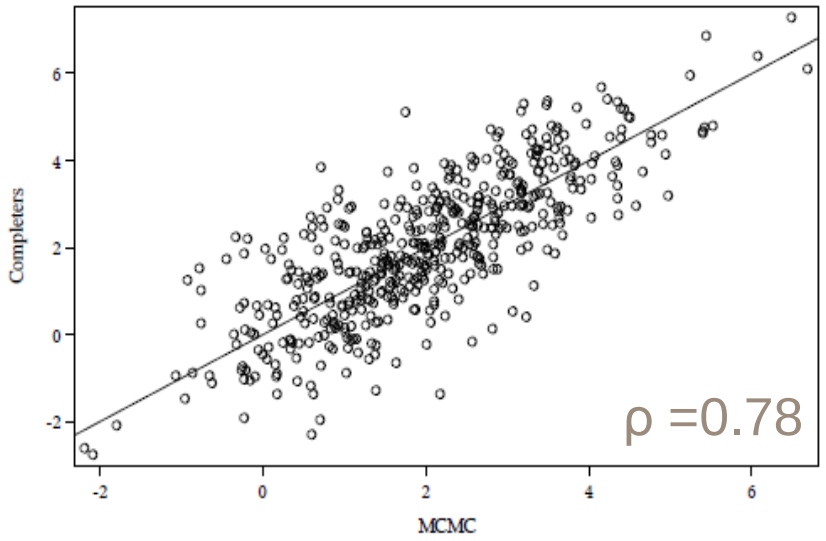
```
Data sim....

Proc mixed....
By sim;
Run;

data store;
` set _null_;
run;

%do k=1 %to &nsims.;

Proc mcmc ..... Seed=%eval(&seed.+3*&k.)

data out;
    set store mcmc (in=a);
     if a then sim=&k.;
run;

%end;
```

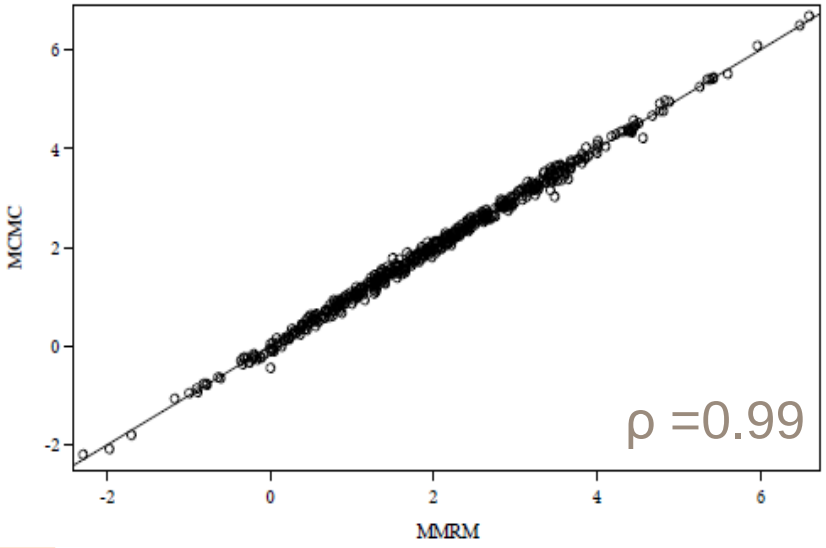
Number of subjects at each of the visits

Length of recruitment	Week 12	Week 24	Week 36	Week 48
40 weeks	140	136	94	50
60 weeks	136	106	78	50
80 weeks	108	88	68	50

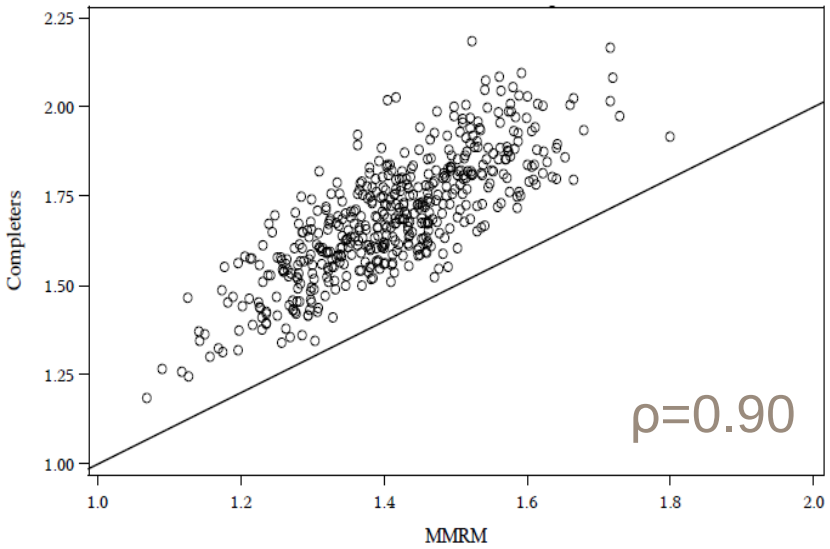
Output from one simulation – Estimated treatment difference

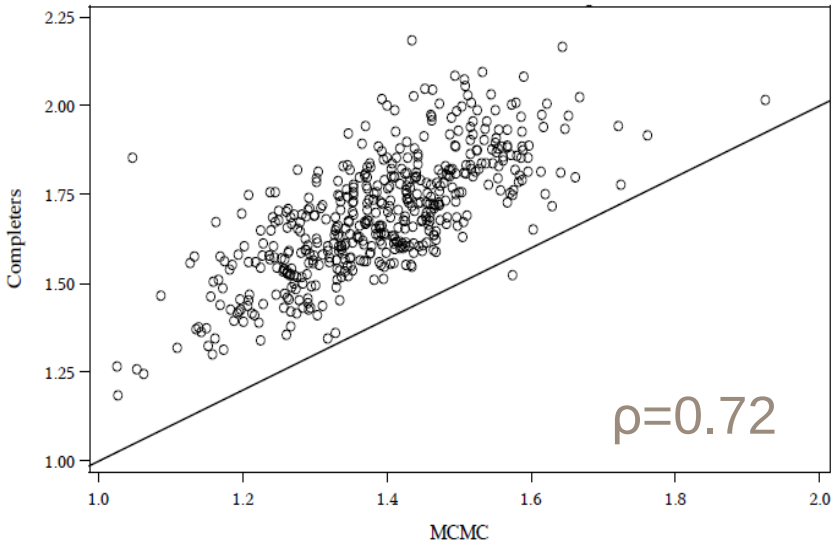


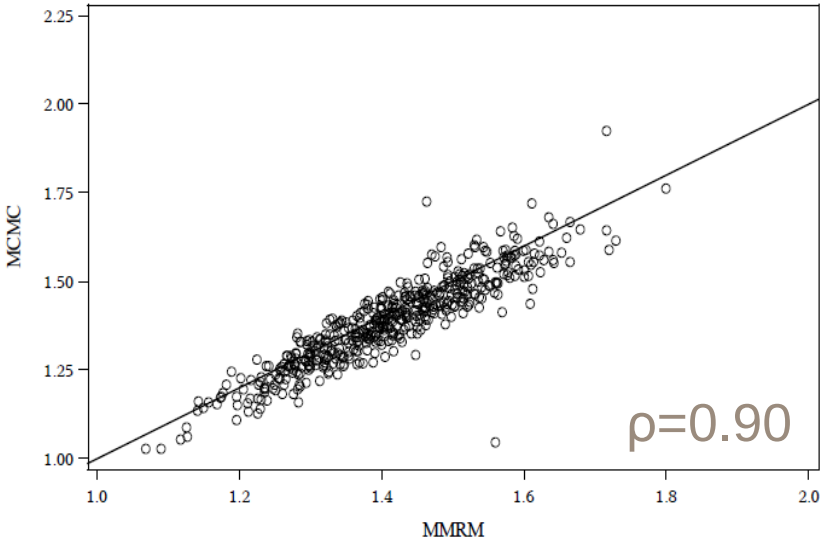




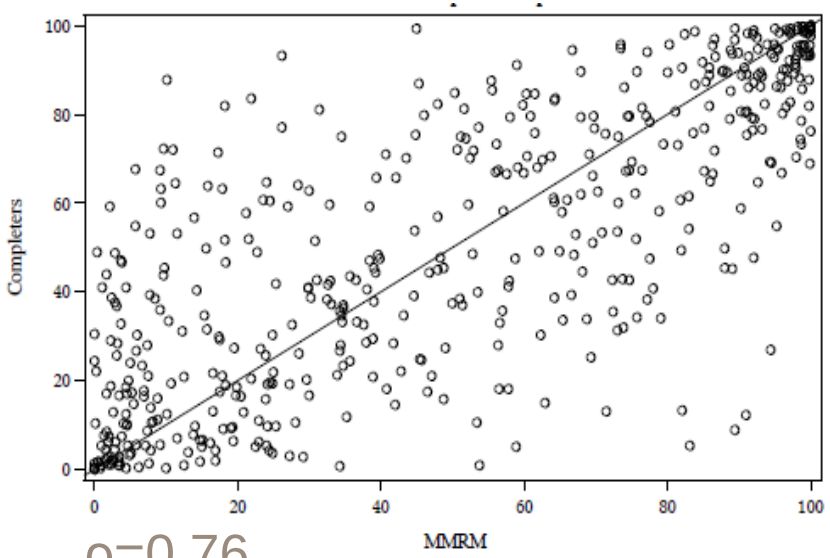
Output from one scenario– Estimated standard error



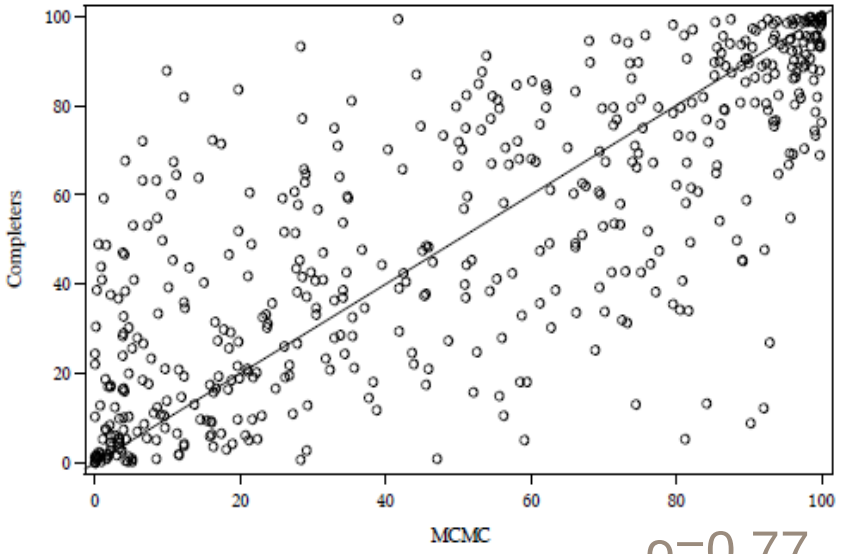




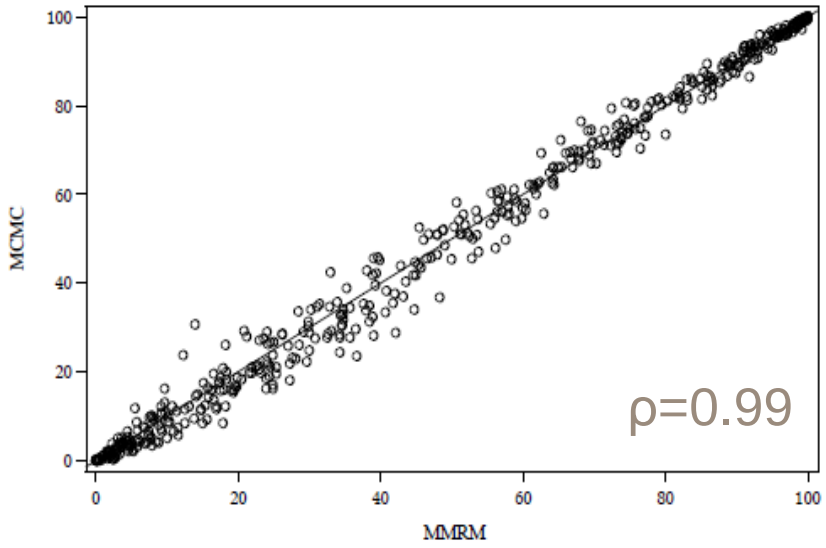
Output from one scenario– Predictive power



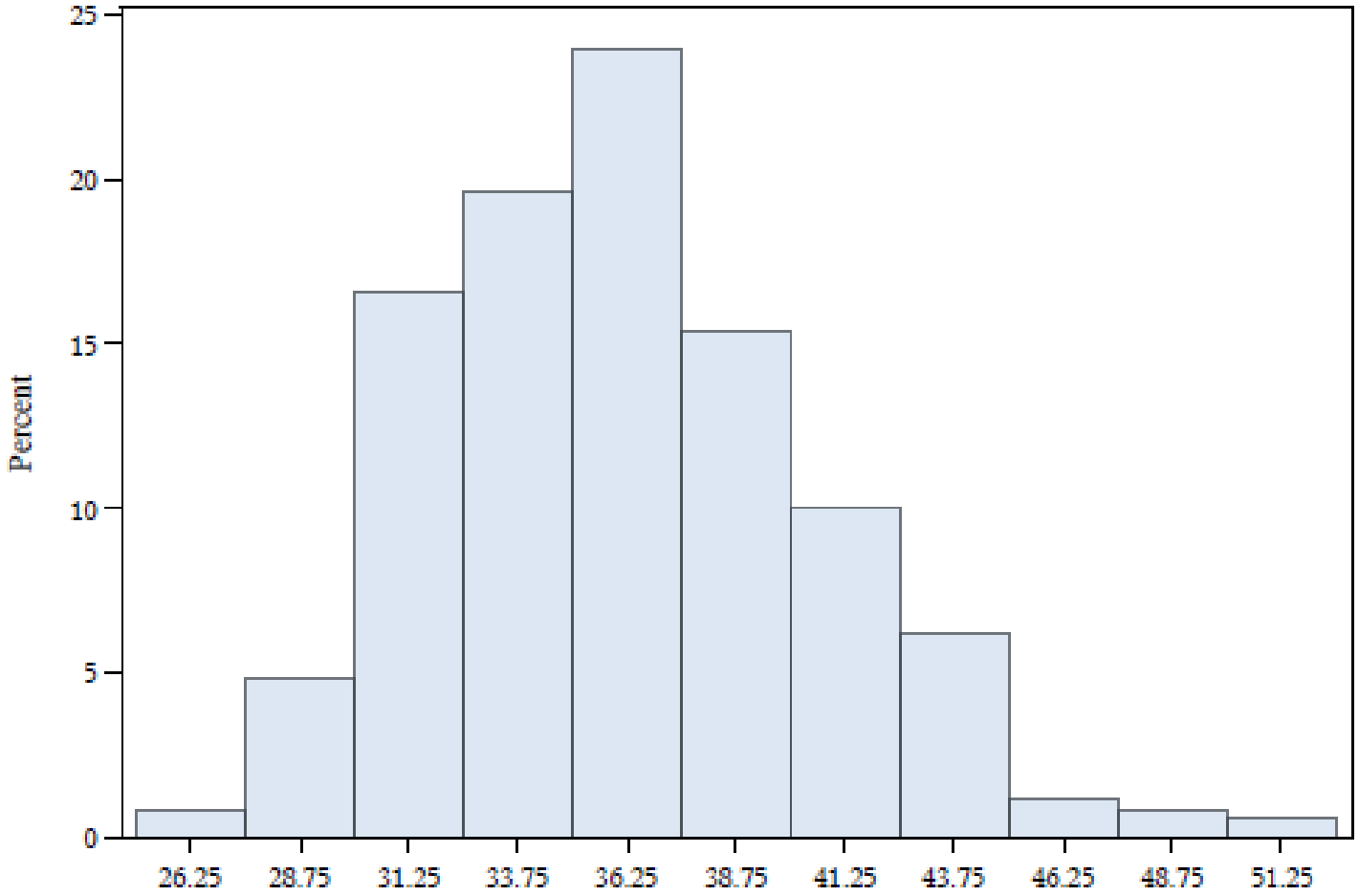
ρ=0.76



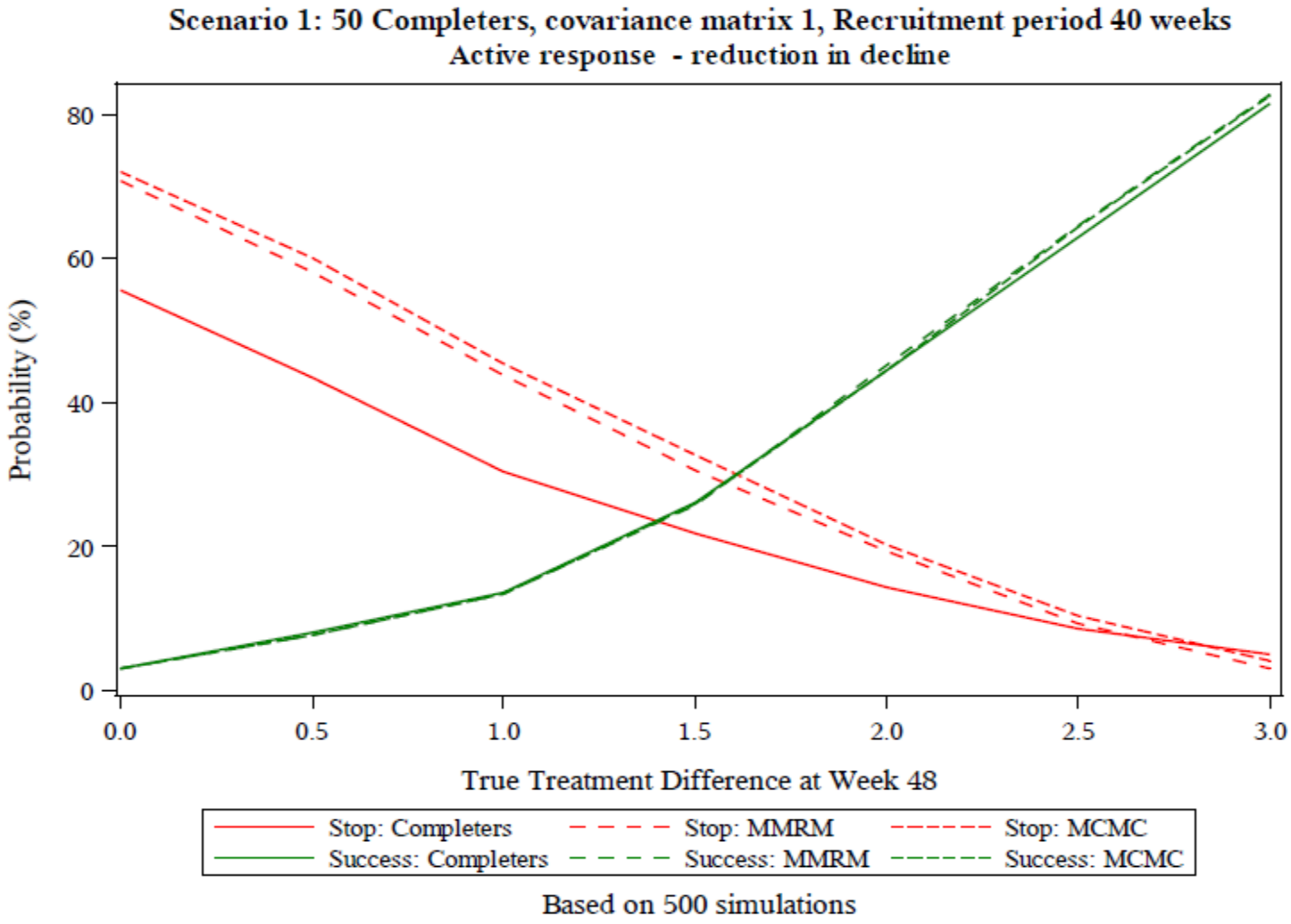
ρ=0.77



Effective sample size (recruitment=40 weeks)



Operating characteristics



Impact of length of recruitment: Change in decline

Covariance matrix	Length of recruitment	Probability of stopping when true treatment difference=0			Probability of stopping when true treatment difference=3		
		Method			Method		
		Comp.	MMRM	MCMC	Comp.	MMRM	MCMC
1	40 weeks	55.6	70.8	72.0	5.0	3.0	4.2
	60 weeks	56.2	67.2	69.0	5.2	2.8	3.6
	80 weeks	56.4	64.2	65.4	5.2	4.2	4.6
2	40 weeks	54.6	84.0	84.4	5.4	6.4	7.2
	60 weeks	55.8	79.6	81.4	5.4	4.2	3.6
	80 weeks	56.2	71.8	73.4	5.4	3.6	4.6

Conclusions of futility simulation

- Using all the data at our interim increases our probability of stopping when there is a small difference
- Estimating the effective sample size of the partial data is a reasonable approximation to using the MCMC method
- Although we stop more often, this does not impact the power
  - More likely to stop the studies that would have gone on to be non-significant

How simulations helped decision making

- In both cases simulations helped quantify some unknown
  - For the go/ no go decision rules it helped increase understanding of the endpoint
  - For the futility analysis it answered the question regarding which data to include and provided more realistic operating characteristics

- Simulating is a powerful tool to allow you to understand the impact of the assumptions you make

Questions?