Statistical Aspects of Futility Analyses

Kevin J Carroll

Spring 2013

Contents

- Introduction
- The Problem in Statistical Terms
- Defining Futility
- Three Common Futility Rules
- The Maths
- An Example
- Comparing the Rules
- Another way of looking at futility and Conditional Power
- More than one futility analysis
- Generic Use of Predictive Power
- Summary

KJC STATISTICS

Introduction

- Where efficacy of a drug in Phase III can be reliably judged from preceding Phase II work, Phase III can be designed under an assumed alternative hypothesis with reasonable confidence the final outcome will be positive.

- Where the true efficacy of drug is less certain, sometimes Phase III trials are commenced at somewhat increased risk. In such situations an early interim for futility may be of value.

- The goal is then to ensure developments are sensibly risk managed, such that the number of Phase III trials that complete and yet fail is minimized.

- As with any interim analysis, best scientific practice is to employ an *independent* DMC to oversee the analysis and reduce the potential for bias

The Problem in Statistical Terms

- Assume we are comparing two treatments, experimental (E) and control (C).
- The hypothesis to be tested is $H_0 : \theta_{TRUE}=0$ vs $H_1 : \theta_{TRUE} \neq 0$.
- For the purposes of sizing assume $\theta_{TRUE} = \theta$ (>0) under the alternative.
- Let x be a sufficient statistic for θ with distribution $f(x|\theta) \sim N(\theta, V^{-1})$.
- Trial size is then governed by Type I and Type II errors, α and β, and the need to deliver the required information content, $V = \left(z_\alpha + z_\beta\right)^2 / \theta^2$.
- The null hypothesis is rejected when $x > c = \frac{z_\alpha}{\sqrt{V}}$
- An interim analysis is planned with V_i information where $V_i = V \times i$, $0 < i \leq 1$
- Futility will be declared if x_i, the observed treatment effect, is 'small', say less than some value c_i.

Defining Futility

- What is ‘futile’?
- p=NS at the interim? No since we cannot prove the null of ‘no difference’
- The 95% CI at the interim excludes the hypothesised difference θ? Possible but how conservative (or not) might this be?
- Low power to test for θ given the data we have seen at the interim?
- None of the above – we rather simply look at the data and make an experienced judgement regarding the apparent strength of (all the) evidence?
- And do we need to make an alpha adjustment and, if so, how much?

KJC STATISTICS

Three Common Futility Rules

Three rules are generally used to define 'futility':

I. Stop if Pr(p<0.025 at end of trial | V_i; θ) is 'small'.
II. Stop if Pr(p<0.025 at end of trial | V_i; x_i) is 'small'.
III. Stop if Pr(p<0.025 at end of trial | V_i) is 'small'.

I: Conditional Power = given the interim data and assuming $\theta_{TRUE} = \theta$ as hypothesised, what is the chance of showing p<0.025 at the end of the trial?

II: UnConditional Power = given the interim data and assuming $\theta_{TRUE} = x_i$ as observed at the interim, what is the chance of showing p<0.025 at the end of the trial?

III. Predictive Power = given the interim data and assuming $f(\theta) \sim N\left(x_i, \frac{1}{Vi}\right)$ as our best estimate of θ_{TRUE}, what is the chance of showing p<0.025 at the end of the trial?

The Maths

- In general, assume there are k interims at information times

  $0 < i_1 < i_2 < \cdots < i_k = 1$, with cumulative information content at interim i_s of $V_s = V \times i_s,\ s = 1, \ldots, k.$

- Then observed outcomes $x_1, \ldots, x_k$ are MVN with mean θ,

  $var(x_s) = V_s^{-1}$ and $corr(x_s, x_r) = \sqrt{\frac{s}{r}}$ for $s \leq r.$

- Therefore, with just one interim and a final $f(x, x_i; \theta)$ is bivariate normal:

  $$\begin{pmatrix} x \\ x_i \end{pmatrix} \sim N\left[\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} V^{-1} & V^{-1} \\ V^{-1} & V_i^{-1} \end{pmatrix}\right]$$

- Overall Power = $\Pr(x_i \geq c_i \cap x \geq c) = Pr\left[\begin{pmatrix} x_i \\ x \end{pmatrix} \geq \begin{pmatrix} (c_i - \theta)\sqrt{Vi} \\ (c - \theta)\sqrt{V} \end{pmatrix}\right]$

- Since $f(x|x_i;\theta) \sim N\left(ix_i + \theta(1-i), \frac{1-i}{V}\right)$ then:

  Conditional Power = $\Phi\left(\frac{c - ix_i - \theta(1-i)}{\sqrt{\frac{1-i}{V}}}\right)$

  UnConditional Power = $\Phi\left(\frac{c - x_i}{\sqrt{\frac{1-i}{V}}}\right)$

- Since $f(x|x_i) = \int_\theta f(x|x_i;\theta) f(\theta) d\theta$ where $f(\theta) \sim N\left(x_i, \frac{1}{Vi}\right)$ then $f(x|x_i) \sim N\left(x_i, \frac{1}{Vi} - \frac{1}{V}\right)$ so that:

  Predictive Power = $\Phi\left(\frac{c - x_i}{\sqrt{\frac{1}{Vi} - \frac{1}{V}}}\right)$

KJC STATISTICS

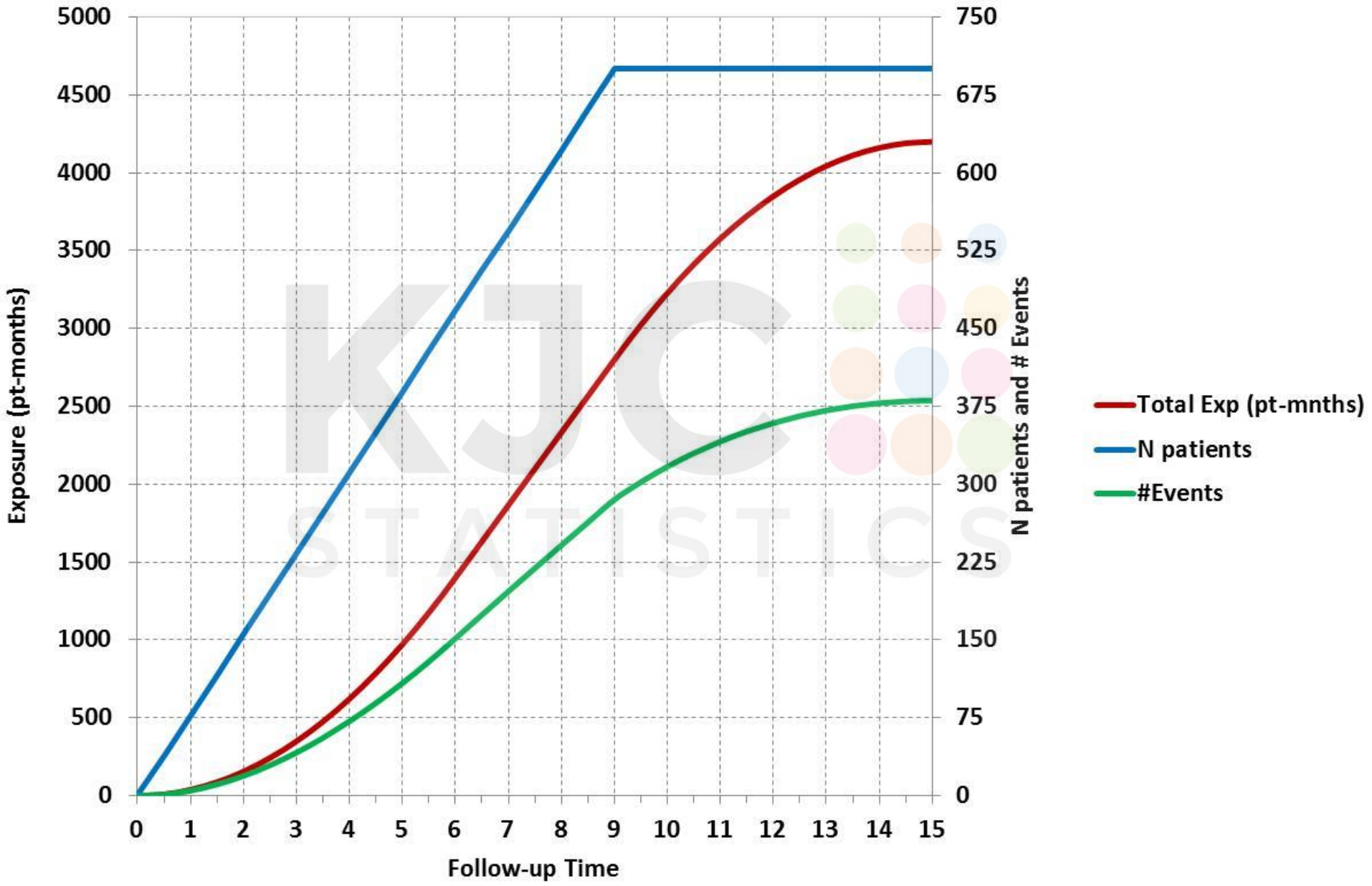
An Example

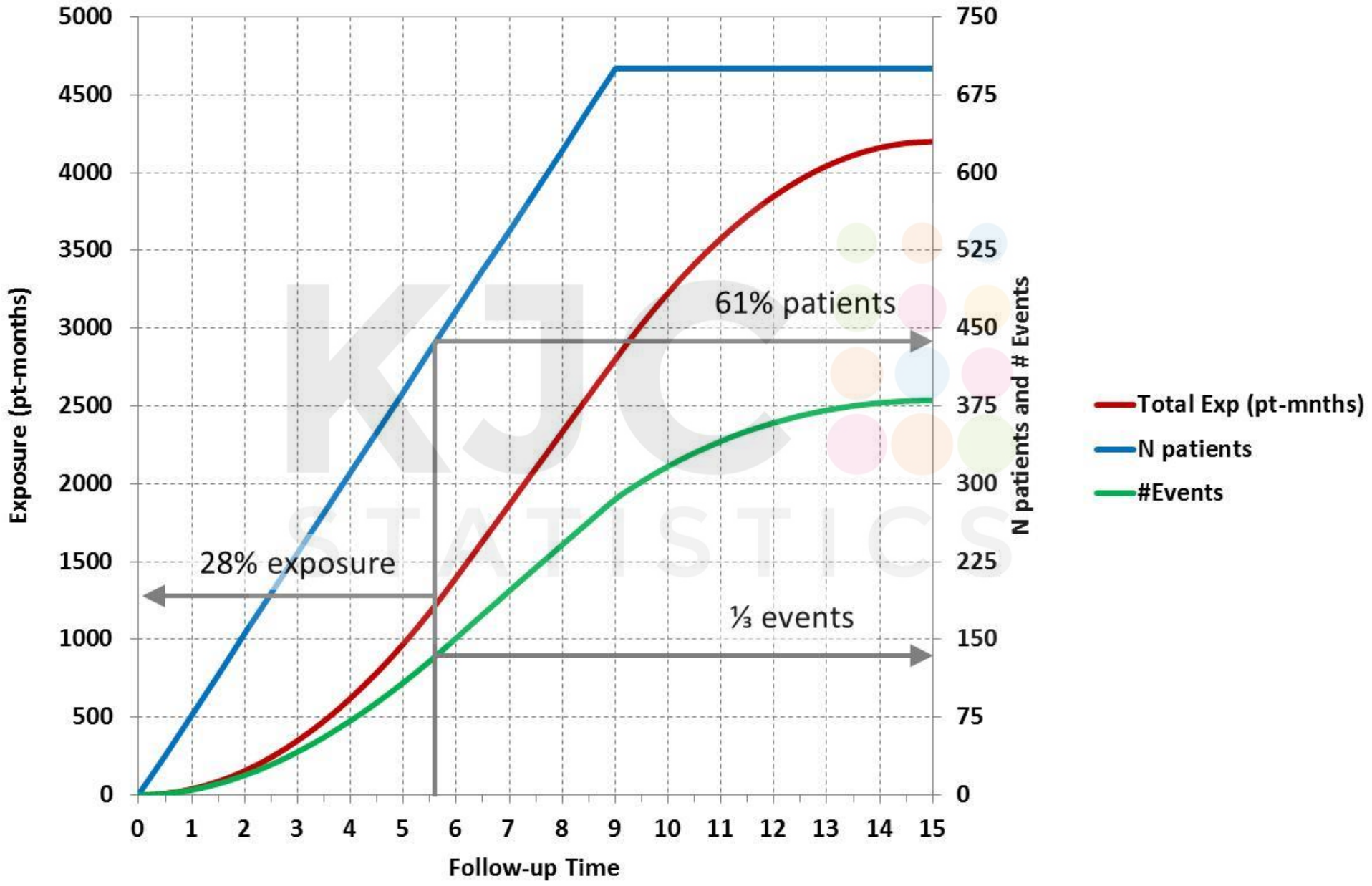
- H_0 : HR$=1$ vs H_1 : HR = 0.75. Hence, θ=ln(0.75)= –0.288. With α=2.5% and β=10%, V=94.75 $\Rightarrow$ 379 events. Therefore $c = -0.201$ so that $HR_c = e^{-0.201}$ = 0.818.

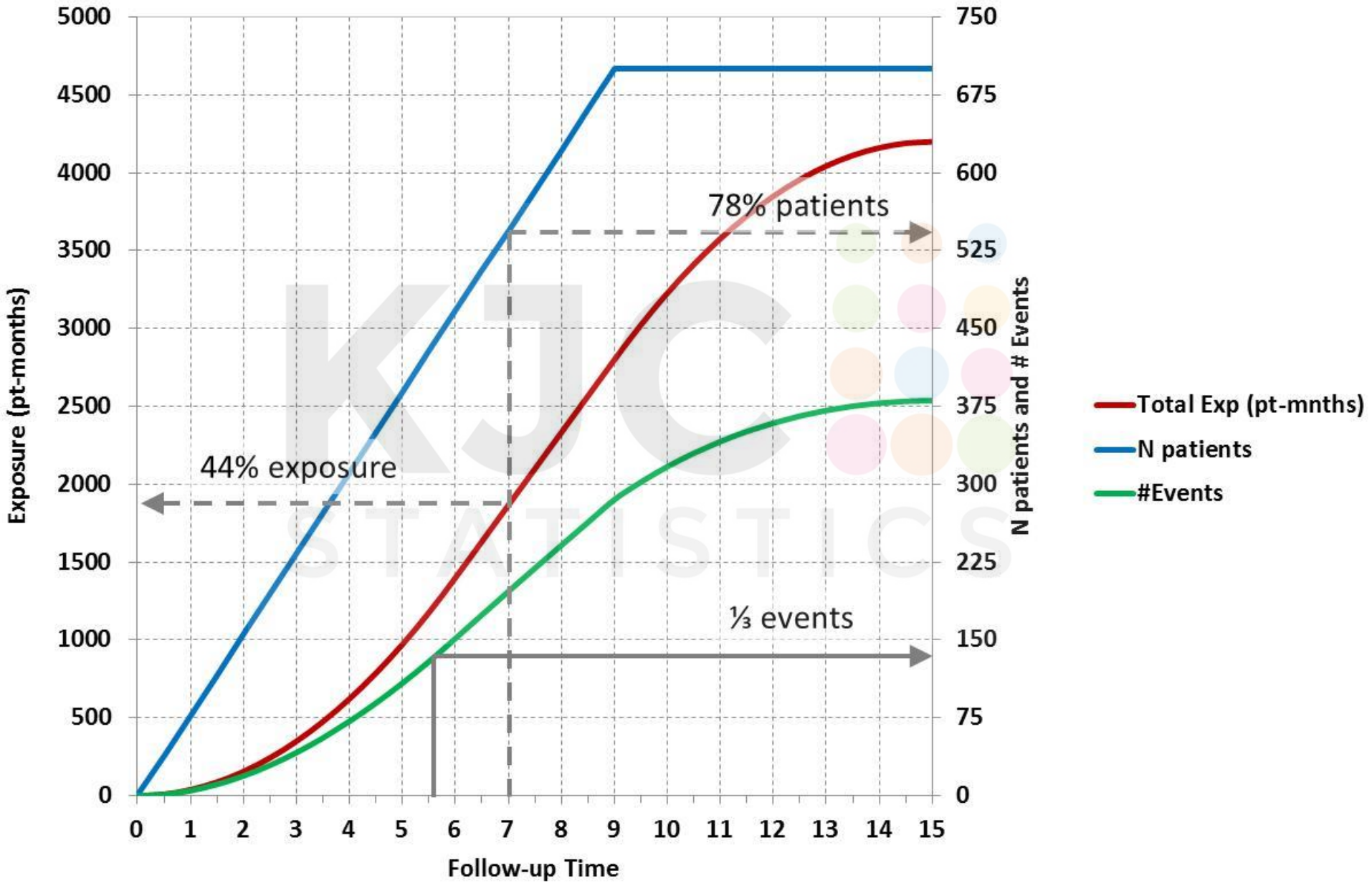
- With a 60% event rate at 6 months on control and with 9 months accrual and 6 months maximum follow-up, N=700 patients need to be randomised.

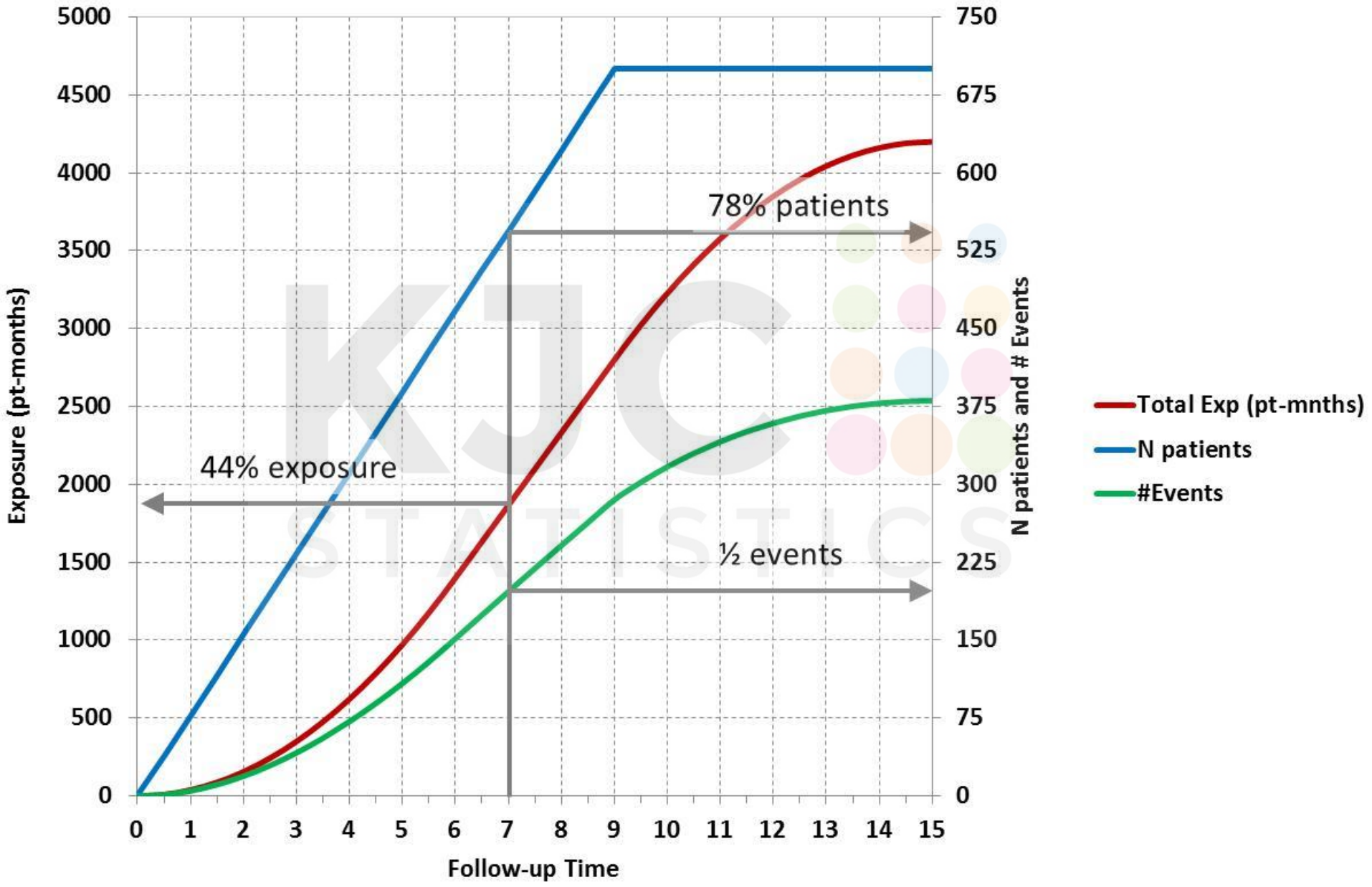
- A futility analysis is desired – at what faction of the events might such an analysis made sense (or not)?

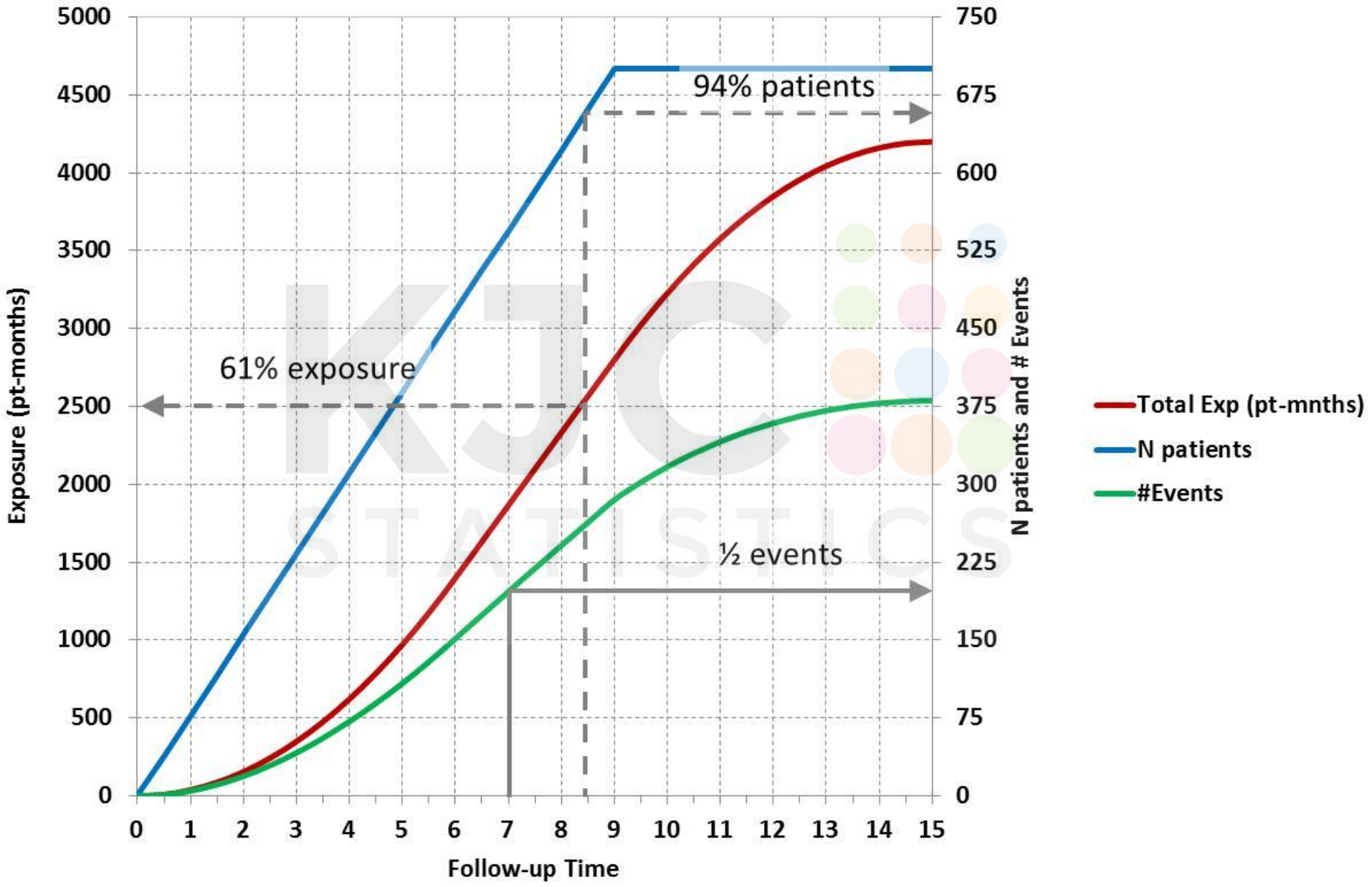
KJC STATISTICS


5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Exposure (pt-months)
750
675
600
525
450
375
300
225
150
75
0
N patients and # Events
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up Time
Total Exp (pt-mnths)
N patients
#Events


5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Exposure (pt-months)
750
675
600
525
450
375
300
225
150
75
0
N patients and # Events
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up Time
61% patients
28% exposure
⅓ events
Total Exp (pt-mnths)
N patients
#Events


5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Exposure (pt-months)
750
675
600
525
450
375
300
225
150
75
0
N patients and # Events
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up Time
78% patients
44% exposure
⅓ events
Total Exp (pt-mnths)
N patients
#Events

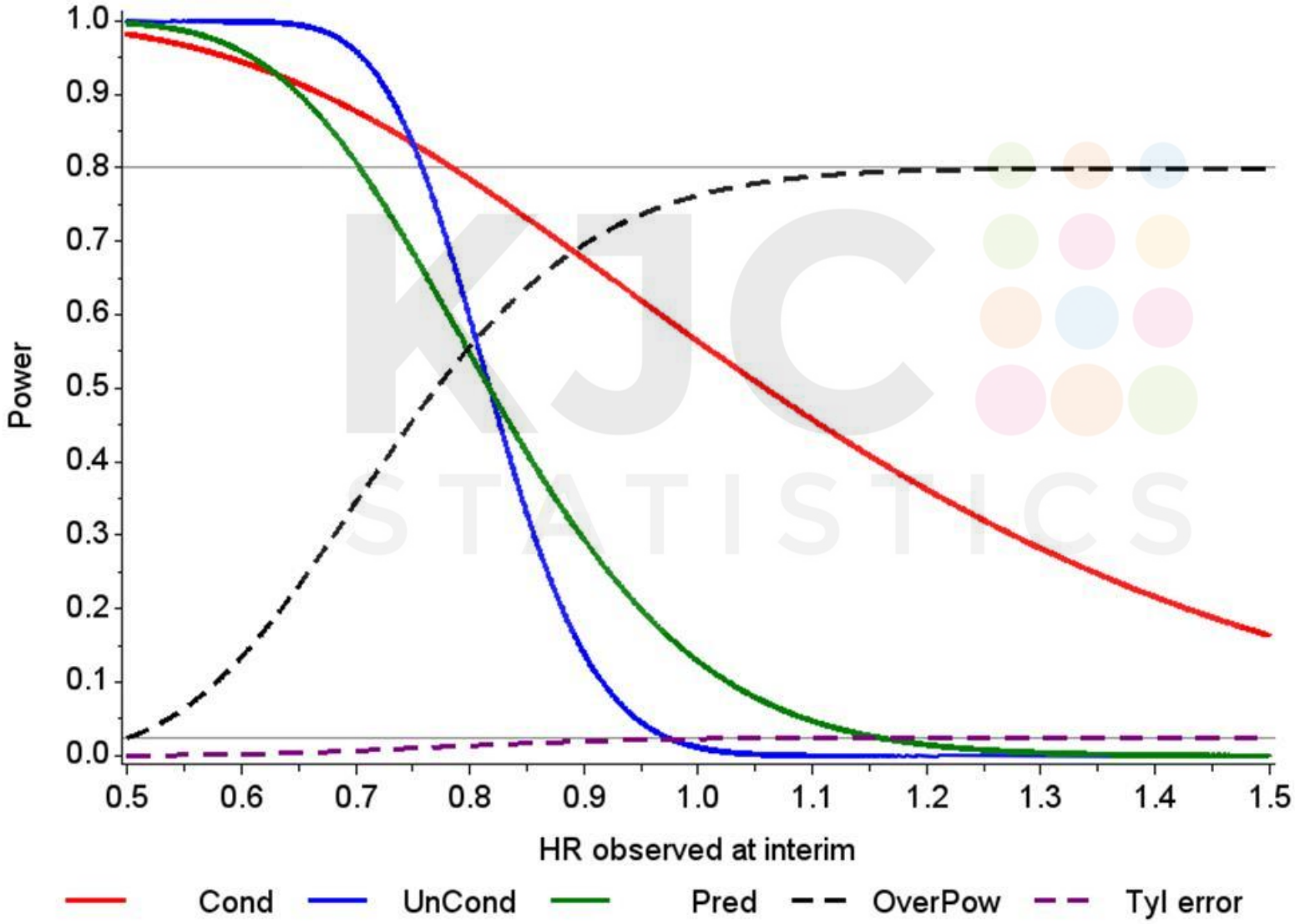

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Exposure (pt-months)
750
675
600
525
450
375
300
225
150
75
0
N patients and # Events
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up Time
78% patients
44% exposure
½ events
Total Exp (pt-mnths)
N patients
#Events


5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Exposure (pt-months)
750
675
600
525
450
375
300
225
150
75
0
N patients and # Events
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up Time
94% patients
61% exposure
½ events
Total Exp (pt-mnths)
N patients
#Events

- With ⅓ events:
  - Interim analysis results expected to be available only when 78% patients entered and 44% exposure accrued
- With ½ events:
  - Interim analysis results expected to be available only when 94% patients entered and 61% exposure accrued
- At what time point would you recommend a futility interim?

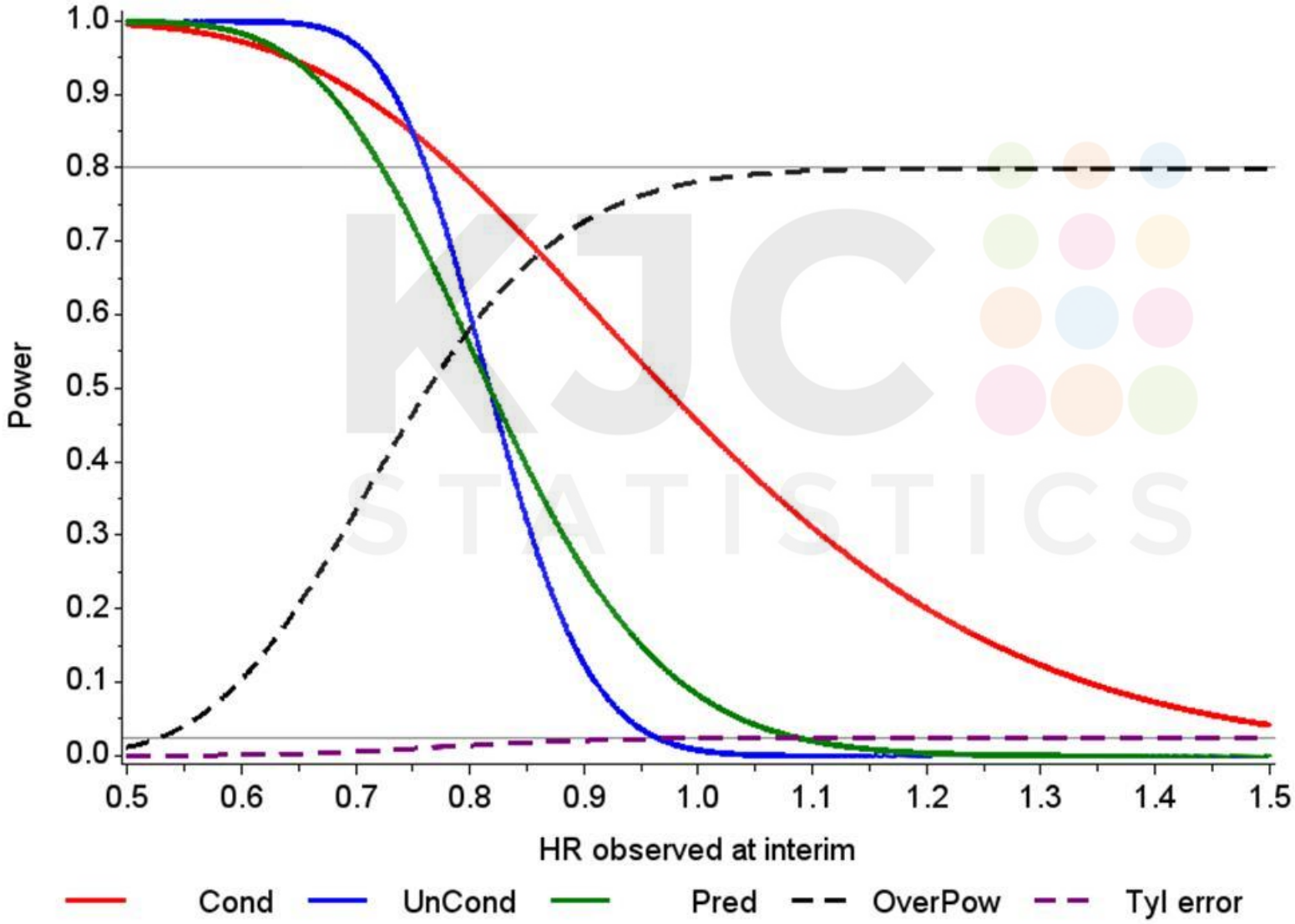
KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=25%</u>



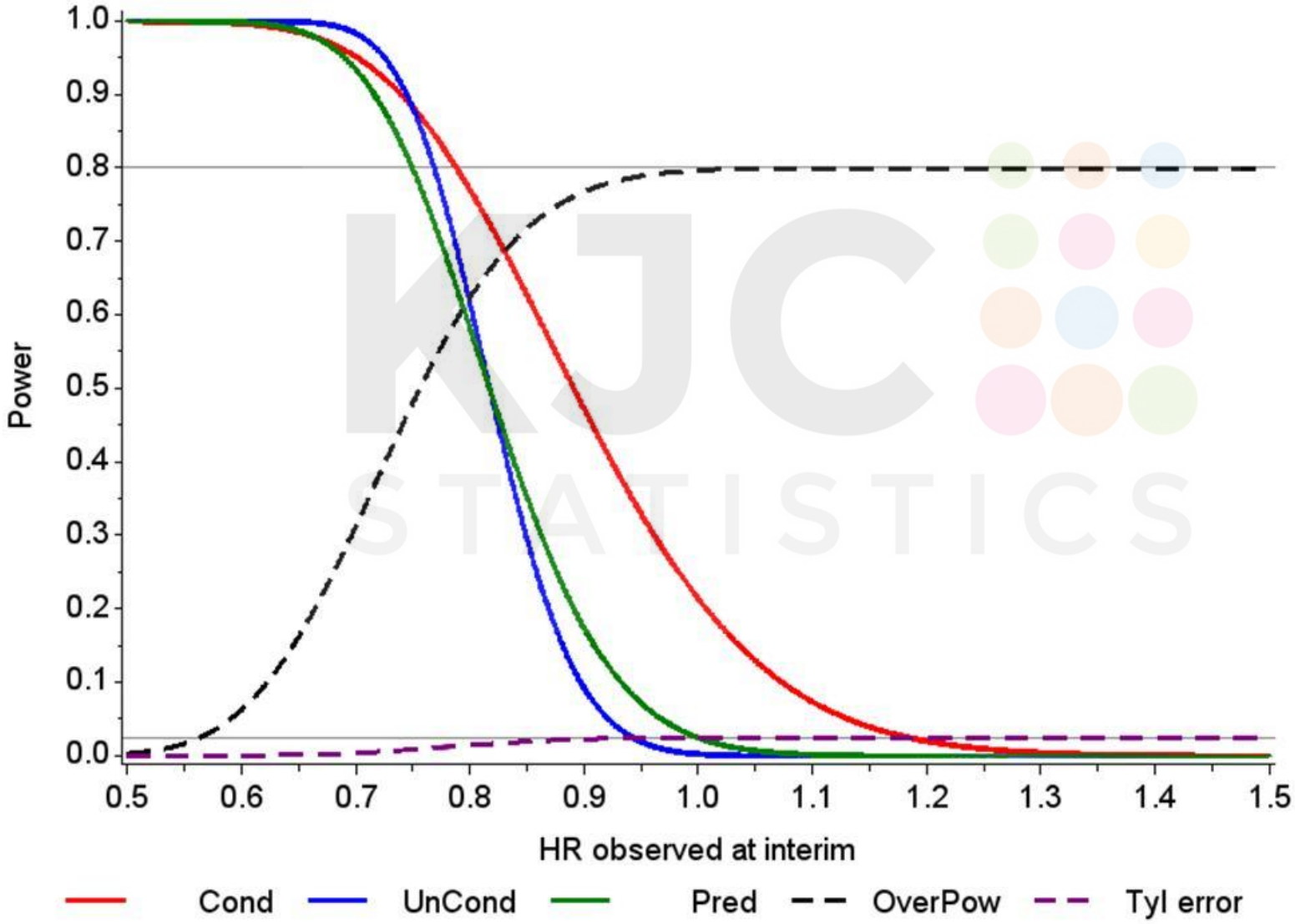
KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=33%</u>



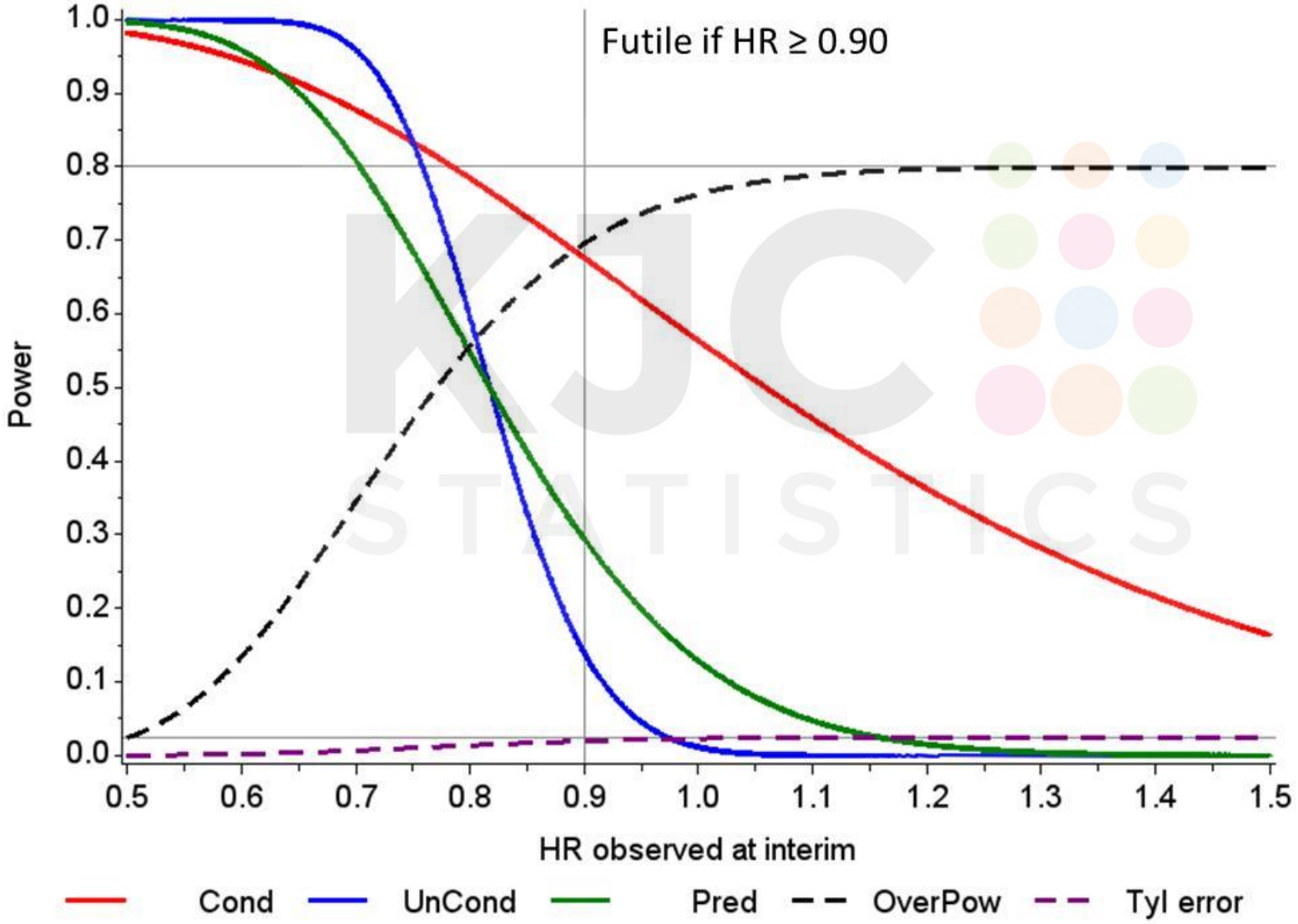
KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=50%</u>



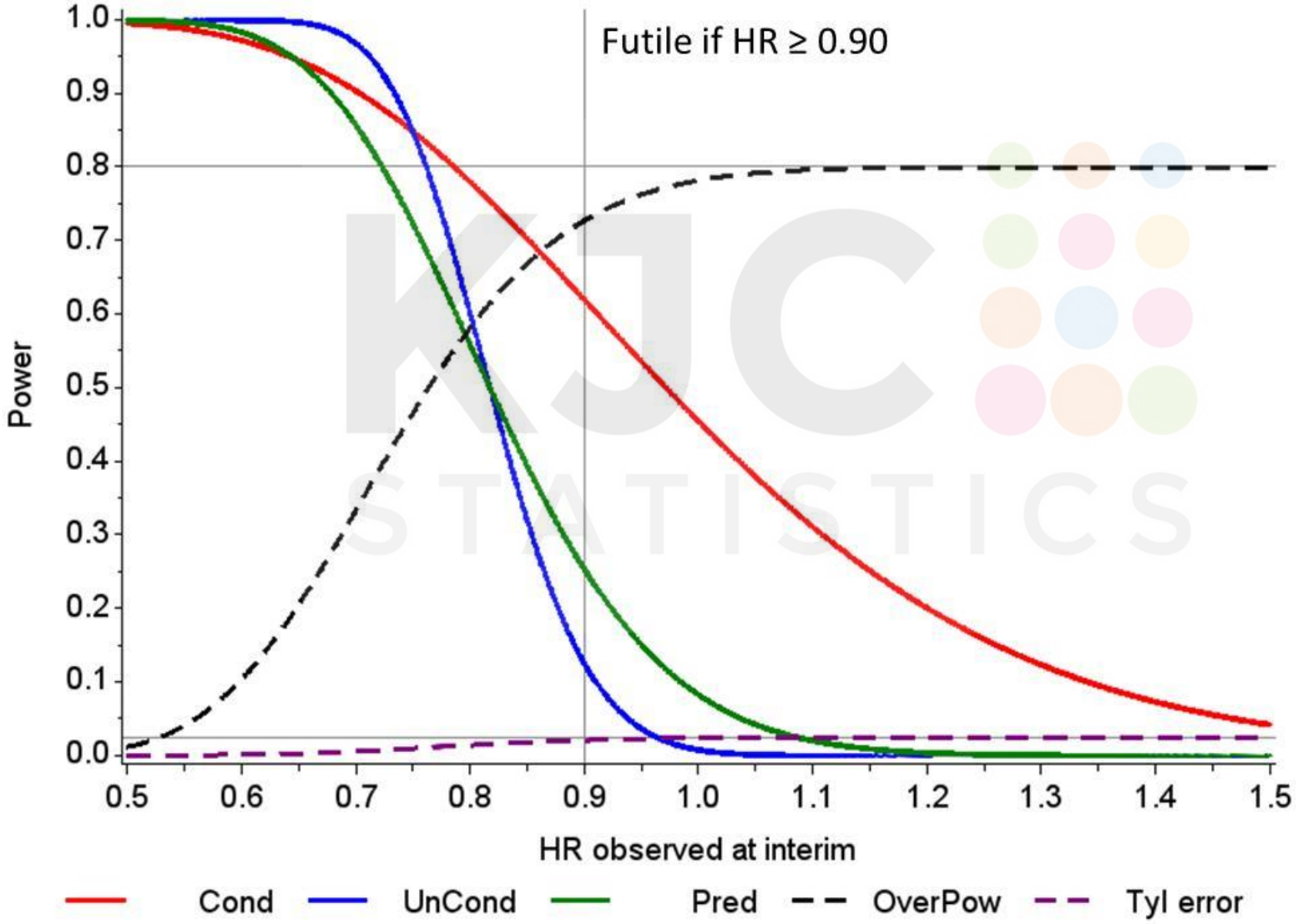
KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=25%</u>



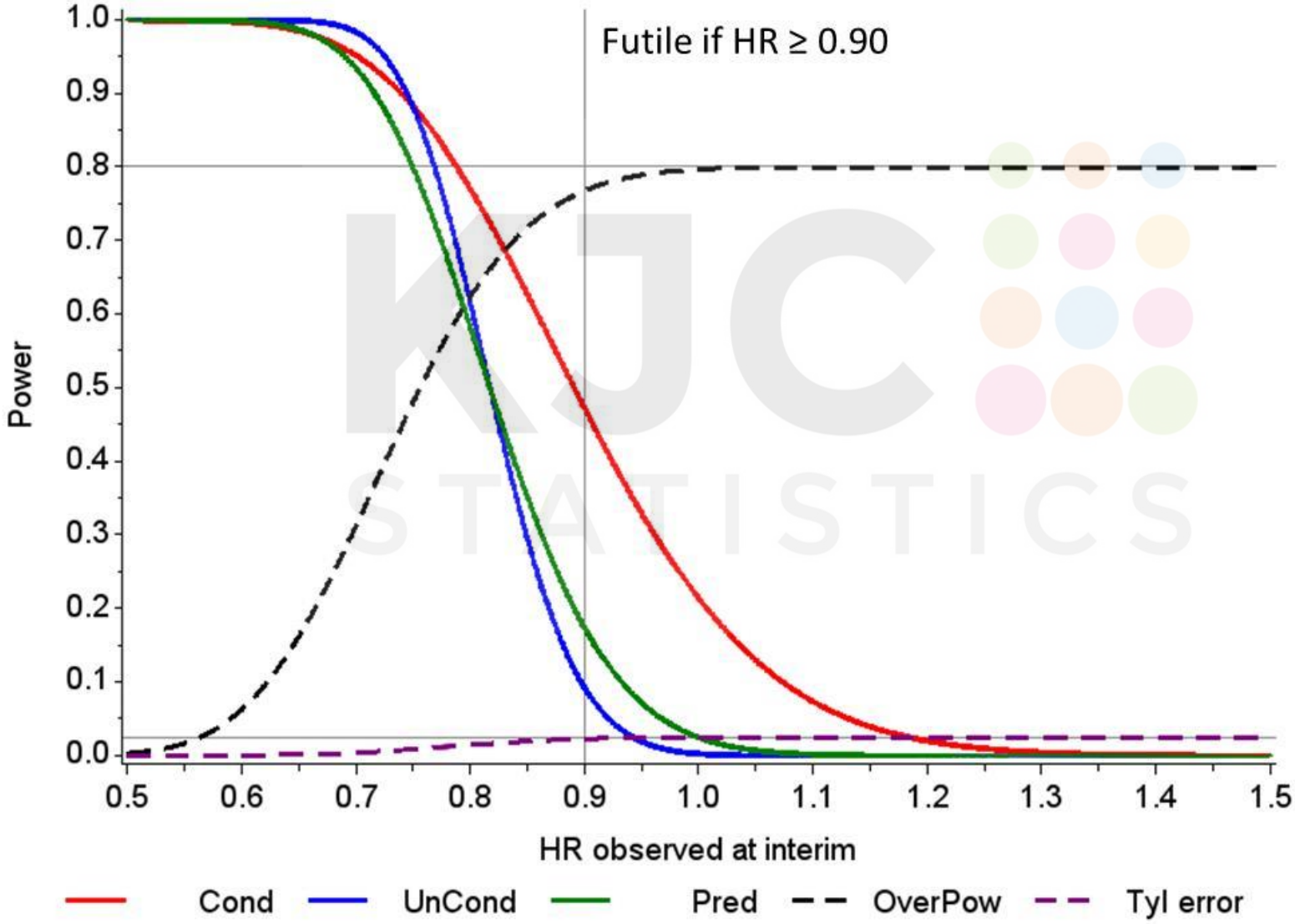
KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=33%</u>



KJC STATISTICS

θ_{TRUE} = 0.75, α=0.025, 1-β=0.80, Events=379, <u>futility=50%</u>



KJC STATISTICS

Comparing rules

α	1-β	θ	Events	C	% events at interim	C_i	Cond Power	UnCond Power	Pred Power	Overall Power	Overall Type I error
2.50%	80%	0.75	379	0.818	25%	0.900	0.677	0.140	0.295	0.695	0.0193
2.50%	80%	0.75	379	0.818	33%	0.900	0.620	0.126	0.254	0.727	0.0204
2.50%	80%	0.75	379	0.818	50%	0.900	0.473	0.093	0.175	0.767	0.0224
3.29%	80%	0.75	379	0.821	25%	0.900	0.685	0.161	0.310	0.689	0.0250
3.10%	80%	0.75	379	0.820	33%	0.900	0.630	0.142	0.268	0.722	0.0250
2.81%	80%	0.75	379	0.819	50%	0.900	0.483	0.101	0.184	0.766	0.0250
2.50%	80%	0.75	379	0.818	25%	1.027	0.535	0.005	0.100	0.772	0.0232
2.50%	80%	0.75	379	0.818	33%	0.985	0.479	0.013	0.100	0.777	0.0234
2.50%	80%	0.75	379	0.818	50%	0.933	0.378	0.035	0.100	0.783	0.0236

- Note that overall power is reduced when an interim for futility is performed
- Type I error is not inflated – the reverse is true
- Some Type I error may therefore be 'reclaimed' by increasing the α level applied

Comparing rules

α	1-β	θ	Events	C	% events at interim	C_i	Cond Power	UnCond Power	Pred Power	Overall Power	Overall Type I error
2.50%	80%	0.75	379	0.818	25%	0.900	0.677	0.140	0.295	0.695	0.0193
2.50%	80%	0.75	379	0.818	33%	0.900	0.620	0.126	0.254	0.727	0.0204
2.50%	80%	0.75	379	0.818	50%	0.900	0.473	0.093	0.175	0.767	0.0224
3.29%	80%	0.75	379	0.821	25%	0.900	0.685	0.161	0.310	0.689	0.0250
3.10%	80%	0.75	379	0.820	33%	0.900	0.630	0.142	0.268	0.722	0.0250
2.81%	80%	0.75	379	0.819	50%	0.900	0.483	0.101	0.184	0.766	0.0250
2.50%	80%	0.75	379	0.818	25%	1.027	0.535	0.005	0.100	0.772	0.0232
2.50%	80%	0.75	379	0.818	33%	0.985	0.479	0.013	0.100	0.777	0.0234
2.50%	80%	0.75	379	0.818	50%	0.933	0.378	0.035	0.100	0.783	0.0236

- Note that overall power is reduced when an interim for futility is performed
- Type I error is not inflated – the reverse is true
- Some Type I error may therefore be 'reclaimed' by increasing the α level applied

Comparing rules

α	1-β	θ	Events	C	% events at interim	C_i	Cond Power	UnCond Power	Pred Power	Overall Power	Overall Type I error
2.50%	80%	0.75	379	0.818	25%	0.900	0.677	0.140	0.295	0.695	0.0193
2.50%	80%	0.75	379	0.818	33%	0.900	0.620	0.126	0.254	0.727	0.0204
2.50%	80%	0.75	379	0.818	50%	0.900	0.473	0.093	0.175	0.767	0.0224
3.29%	80%	0.75	379	0.821	25%	0.900	0.685	0.161	0.310	0.689	0.0250
3.10%	80%	0.75	379	0.820	33%	0.900	0.630	0.142	0.268	0.722	0.0250
2.81%	80%	0.75	379	0.819	50%	0.900	0.483	0.101	0.184	0.766	0.0250
2.50%	80%	0.75	379	0.818	25%	1.027	0.535	0.005	0.100	0.772	0.0232
2.50%	80%	0.75	379	0.818	33%	0.985	0.479	0.013	0.100	0.777	0.0234
2.50%	80%	0.75	379	0.818	50%	0.933	0.378	0.035	0.100	0.783	0.0236

- Note that overall power is reduced when an interim for futility is performed
- Type I error is not inflated – the reverse is true
- Some Type I error may therefore be 'reclaimed' by increasing the α level applied

Comparing rules

α	1-β	θ	Events	C	% events at interim	C_i	Cond Power	UnCond Power	Pred Power	Overall Power	Overall Type I error
2.50%	80%	0.75	379	0.818	25%	0.900	0.677	0.140	0.295	0.695	0.0193
2.50%	80%	0.75	379	0.818	33%	0.900	0.620	0.126	0.254	0.727	0.0204
2.50%	80%	0.75	379	0.818	50%	0.900	0.473	0.093	0.175	0.767	0.0224
3.29%	80%	0.75	379	0.821	25%	0.900	0.685	0.161	0.310	0.689	0.0250
3.10%	80%	0.75	379	0.820	33%	0.900	0.630	0.142	0.268	0.722	0.0250
2.81%	80%	0.75	379	0.819	50%	0.900	0.483	0.101	0.184	0.766	0.0250
2.50%	80%	0.75	379	0.818	25%	1.027	0.535	0.005	0.100	0.772	0.0232
2.50%	80%	0.75	379	0.818	33%	0.985	0.479	0.013	0.100	0.777	0.0234
2.50%	80%	0.75	379	0.818	50%	0.933	0.378	0.035	0.100	0.783	0.0236

- Note that overall power is reduced when an interim for futility is performed
- Type I error is not inflated – the reverse is true
- Some Type I error may therefore be 'reclaimed' by increasing the α level applied

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ with information $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(V - Vi)^{-1}}\right\} \approx$ Conditional Power
- For a HR:
  - approx # events on drug $= E_{total}/(HR + 1)$
  - approx # events on control $= HR \cdot E_{total}/(HR + 1)$

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ with information $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(V - Vi)^{-1}}\right\} \approx$ Conditional Power
- For a HR:
  - approx # events on drug $= E_{total}/(HR + 1)$
  - approx # events on control $= HR \cdot E_{total}/(HR + 1)$

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ with information $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(V - Vi)^{-1}}\right\} \approx$ Conditional Power
- For a HR:
  - approx # events on drug $= E_{total}/(HR + 1)$
  - approx # events on control $= HR \cdot E_{total}/(HR + 1)$

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ with information $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(V - Vi)^{-1}}\right\} \approx$ Conditional Power
- For a HR:
  - approx # events on drug $= E_{total}/(HR + 1)$
  - approx # events on control $= HR \cdot E_{total}/(HR + 1)$

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ with information $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(\text{V} - \text{Vi})^{-1}}\right\} \approx$ Conditional Power
- For a HR:
  - approx # events on drug $= \text{E}_{total}/(\text{HR} + 1)$
  - approx # events on control $= \text{HR} \cdot \text{E}_{total}/(\text{HR} + 1)$

Another way of looking at futility and Conditional Power

						Final Events			Interim Events			What we would need to see in 2nd block and probability of observing this					
α	β	θ	Total Events	% events at int	# events at int	C	Drug	Control	Ci	Drug	Control	Events remaining	Drug	Control	HR	P(HR | θ)	Cond Power
0.025	0.2	0.75	379	25%	95	0.818	170	209	1.027	48	47	284	123	162	0.758	0.534	0.535
0.025	0.2	0.75	379	33%	126	0.818	170	209	0.985	63	64	253	108	145	0.745	0.479	0.479
0.025	0.2	0.75	379	50%	190	0.818	170	209	0.933	91	98	190	79	110	0.717	0.378	0.377

- **Given the data observed at the interim, how would the remaining data have to fall out to still deliver an overall positive result with p<0.025?**
- $\widehat{\theta}_{\text{post int}} = \left(\widehat{\theta}_{\text{final}} - \mathbf{f} \cdot \widehat{\theta}_{\text{int}}\right)/(1 - \mathbf{f})$ **with information** $(V - Vi)$
- $\Pr\left(\widehat{\theta}_{\text{post int}} | \theta\right) = \Phi\left\{\left(\widehat{\theta}_{\text{post int}} - \theta\right)/\sqrt{(V - Vi)^{-1}}\right\} \approx$ **Conditional Power**
- **For a HR:**
  - **approx # events on drug** $= E_{total}/(HR + 1)$
  - **approx # events on control** $= HR \cdot E_{total}/(HR + 1)$

More than one futility analysis

- Straightforward to incorporate 2 or more futility analyses
- R (or S+) software very easy to use and provides excellent flexibility to design bespoke rules (superior to EaST, AddPlann, nTerim)

		Under the null			Under the alternative		
Futile at 1st if HR >	Futile at 2nd if HR >	Pr(stop futility at 1st)	Pr(stop futility at 1st or 2nd)	Overall Type I error	Pr(stop futility at 1st)	Pr(stop futility at 1st or 2nd)	Overall Power
1.027	0.9327	45%	72%	2.23%	6.3%	10.2%	76%

- 1.027 and 0.9327 chosen to provide 10% Predictive Power
- 45% chance to stop correctly at 1st and 72% chance at 1st or 2nd
- 6% chance to stop incorrectly at 1st and 10% chance at 1st or 2nd
- Overall power reduced from 80% to 76%
- Type I error less than 2.5% as expected

Generic Use of Predictive Power

- **It is possible to derive a generic framework for the use of Predictive power**

V_i as % of V	Predictive power	Z-value at interim	Overall power (assuming V sufficient for 90% power)	Relative increase in hypothesised effect to maintain 90% overall power
10%	≤ 10%	≤ -0.5960	86.1%	1.016
	≤ 20%	≤ -0.1786	81.1%	1.042
15%	≤ 10%	≤ -0.4224	86.8%	1.013
	≤ 20%	≤ -0.0168	82.5%	1.033
20%	≤ 10%	≤ -0.2697	87.3%	1.011
	≤ 20%	≤ 0.1238	83.5%	1.025
30%	≤ 10%	≤ 0.0013	87.9%	1.010
	≤ 20%	≤ 0.3694	79.7%	1.048

Generic Use of Predictive Power

- **Thus it is possible to conceive of pragmatic rules that could be applied generically.**

Rule A: '10/10' rule. With 10% information, the trial would be stopped for futility if the interim z-value ≤ -0.5960 (p=0.55 2-sided) since predictive power would be no more than 10%. This would provide 86.1% overall power, or alternatively, 90% overall power if the alternative was in increased in magnitude by a factor of 1.016.

Rule B: '20/20' rule. With 20% information, the trial would be stopped for futility if the interim z-value ≤ 0.1238 (p=0.90 2-sided) since predictive power would be no more than 20%. This would provide 83.5% overall power, or alternatively, 90% overall power if the alternative was in increased in magnitude by a factor of 1.025.

Summary

- From time to time, there may be development scenarios where Phase III trials are commenced at somewhat increased risk.
- In these situations the efficacy assumed under the alternative will likely be a best guess and may be overestimated.
- For such situations, an early interim analysis to assess futility (the chance of achieving a positive ($p<0.025$ 1-sided) outcome at the end of the trial) may be appropriate.
- Several futility definitions are in use at the present time: Conditional, UnConditional and Predictive
- Of these Predictive Power is most appealing

Summary

- The exact choice of when to do the analysis and definition of a ‘low’ chance of success will vary depending on the particular project and circumstances. A suggestion might be
  a) the analysis should take place with at least 10% but not more than 33% of the required statistical information and
  b) the threshold futility probability of success to trigger a consideration to stop should be 20% or less.
- As is usual and best scientific practice, the interim analysis should need to be overseen by an independent DMC.

Summary

- In situations where the primary endpoint is a time to event variable and there is good reason to believe that the treatment effect will not emerge smoothly over time, then futility analyses are generally not be advisable.

- Rather, it would be expected that the trial be designed and sized to take account of a gradual emergence of a treatment effect and that this be taken into account when deciding to proceed to Phase III or not.

KJC STATISTICS