

**S3 Figure: Predictive capabilities of sequential design (Pocock boundaries).**

	Optimistic effect size distribution (different approaches)	Pessimistic effect size distribution (different approaches)
Max. samples size per group (interim batch sizes)	Freq. seq. Pocock	Freq. seq. Pocock
<b>Probability of getting a significant study result <math>P(\text{significant})</math> (%)</b>		
18 (6 + 6 + 6)	19.3	9.4
36 (18 + 18)	28.0	12.2
<b>Positive Predictive value of a significant result <math>P(d \neq 0   \text{significant})</math> (%) (PPV)</b>		
18 (6 + 6 + 6)	86.5	55.9
36 (18 + 18)	90.9	66.7
<b>Positive Predictive value of a significant result <math>P(d \geq 0.5   \text{significant})</math> (%) (PPV)</b>		
18 (6 + 6 + 6)	78.7	47.2
36 (18 + 18)	83.3	57.1
<b>Probability of getting a significant study result that reflects a true effect of <math>d \neq 0</math> (<math>P_{\text{detect true effect } d \neq 0}</math>) (%)</b>		
18 (6 + 6 + 6)	16.7	5.3
36 (18 + 18)	25.4	8.2
<b>Probability of getting a significant study result that reflects a true effect of <math>d \geq 0.5</math> (<math>P_{\text{detect true effect } d \geq 0.5}</math>) (%)</b>		
18 (6 + 6 + 6)	15.2	4.5
36 (18 + 18)	23.3	7.0

Upper Left: ‘Optimistic’ scenario with more larger effect sizes. Upper right: ‘Pessimistic’ scenario with mostly effect sizes of 0. Bottom: Probability of getting a significant test result reflecting a true effect of  $d \neq 0$  or  $d \geq 0.5$ , respectively. First, the probabilities  $P(\text{significant})$  for getting any significant study results are given, then the corresponding positive predictive values and finally the product of both giving the corresponding overall probability of getting a significant study result that truly represents an effect of  $d \neq 0$  or  $d \geq 0.5$  ( $P_{\text{detect true effect}}$ ). Stopping rules that allowed early stopping for futility or success as given in Table S1.