SEARCH GRID FOR TYPE 1 ERROR CALIBRATION

In Step 1 of the algorithm for calculating statistical power, a grid search is employed for every considered Stage 1 sample size $n\_{1}$ to find the corresponding threshold estimate $τ$ that calibrates the type 1 error rate to the desired value. This is achieved in several steps described below and outlined in Figure S1.

First, the probability of rejecting $H\_{0} $is estimated at a starting $τ$ using simulations where all arms exhibit a mean equal to zero (the null hypothesis). Depending on whether the estimated value is too large or too small, the algorithm then adds or subtracts a step size from the current $τ$ and estimates the rejection probability again. This is repeated until the estimated probability overshoots the target type I error rate $α$. At this point the direction of the grid walk is reversed and the step size is shrunk by a factor of 0.7. This process of walking, overshooting and shrinking is repeated until the target error rate is crossed 15 times. In order to protect this algorithm from parameter configurations that lead to always or never rejecting $H\_{0}$, grid walks include boundary conditions. If a step lands on a value larger than $1.0$, an anchor point with a threshold of $1.000001$ and a rejection rate of $0.0$ is added to the grid. Analogously, if a step lands on a value less than $0.0$ then an anchor point is added at $-0.000001$ with a rejection rate of $1.0$.

Next, a snapshot of the grid walk that accumulates points around the target error rate is selected. This is achieved by discarding those initial $τ$ values that were generated before the grid walk had overshot the target type I error rate twice. The ultimate goal of the algorithm is to fit a smoothed curve through points in this snapshot. In order to reduce leverage of outer points on the fit, six filling iterations are performed. In each of these iterations, distances are calculated between all neighboring thresholds. A point is then added to the grid at the center between those neighbors with the largest distance and the corresponding rejection rate is estimated.

After the filling iterations, the most suitable polynomial regression model of up to order three is selected, modeling $τ$ as a function of rejection rate. This fit is then used to predict which $τ$ achieves the target error rate for the given Stage 1 sample size. This predicted value is the raw threshold estimate for the target $n\_{1}$ the grid search provides for further processing in Step 1 of the algorithm.



*Figure S1. A flowchart of the grid search employed to calibrate* $τ$ *for a particular Stage 1 sample size. A: At the starting value of* $τ$ *the simulated type I error* $α$ *is too low so the algorithm walks left. B: After 2 steps the target α is overshot and the algorithm switches directions and lowers the step size. C: After 3 more steps the target* $α$ *is overshot again; the direction is switched and the step size is reduced. D: Search grid after crossing the target* $α$ *several times. Points before* $α$ *was crossed twice are grayed out. E: Search grid after selecting the relevant snapshot; grayed out values from D were removed. F: Search grid after performing filling iterations to reduce leverage of outer points. A smoothed fit is used to predict the target* $τ$*.*