**Supplement. Exact calculation of the probability of rejecting null hypothesis**

There are three routes defined in Section 3. Here, we present how we calculate the probability of rejecting hypotheses through each route.

For Route 1 and Route 2, the interim decision rule requires≥and ≥for both subgroups to proceed to the second stage. Recall that PAVA would pool the response rates if . We can express the interim decision rule by solving:

.

This yields the constraint on the observed  as:

.

Similar scenario might happen in the second stage for Route 1.The probability of rejecting the null hypothesis through Route 1 given the response rate of , , can be calculated using the two conditions below:

1. (Stage 1)  and  and,
2. (Stage 2)  and 

For Route 2, the Stage 1 decision will be the same. For Stage 2, there are two scenarios.

Scenario 1 is when  and . So we can solve for *X*+:

.

Thus the constraint on the observed as

.

In some rare cases, even when , we can still reject through Route 2 after applying PAVA. This would happen when

.

So to summarize, the probability we reject null hypothesis through Route 2 given the response rate of , , can be calculated using the conditions below:

1. (Stage 1)  and  and,
2. (Stage 2 option 1) and or,
3. (Stage 2 option 2) and .

Route 3 is to reject only  when only groups proceed to the second stage. Very similar to Route 2 but the second stage decision rules apply in the first stage. The second stage only involves testing about *Ne*+. the probability we reject null hypothesis through Route 3 given the response rate of , , can be calculated using the conditions below:

1. (Stage 1 option 1)  and  or,
2. (Stage 1 option 2)  and  and,
3. (Stage 2) .

Using the framework shown above, we can calculate the expected sample size and probability of early stopping. Using the conditions outlined above, define the probability of passing the interim analysis for both subgroup as and the probability of having only groups passing the interim analysis as , the probability of early stopping for futility (PET) can be calculated as:

.

The expected sample size  can be calculated as:

.