AUTHORS' REPLY

We appreciate the insightful comments on our 2015 paper [1] and the linking of our methods to those used by Pilz et al. [2]. We agree that our approach to optimizing a trial design can be applied more generally. Indeed, the Dynamic Programming methodology has been applied to find optimal group sequential tests with fixed group sizes [3] and with adaptively chosen group sizes [4, 5], to optimize the group sizes in non-adaptive designs [5], and to construct optimal group sequential tests for a delayed response [6]. The above papers consider general group sequential tests, thereby "optimizing over the combination function", and the different optimization criteria considered include weighted combinations of expected sample size at a number of θ values and the integral of expected sample size with respect to a normal density for θ.

In some of these applications, particularly for designs with more than two analyses, it may not be helpful to express the optimisation problem as solving the appropriate version of equation (4) of the letter. For designs with more than two analyses, we do not have an analytic formula for conditional power to differentiate. In an adaptive design, the optimal group size may be at one end of the range of allowable values for $n_2(z_1)$ and the derivative of the conditional expected gain is not zero there — see, for example, the optimized sample size rule in Figure 10 of [1].

Given the possibility of optimising more generally, one may question why, in [1], we chose to optimize designs within the more restrictive framework used by Mehta and Pocock [7]. The answer is that there are substantial practical advantages in using simple, robust trial designs. One may, for example, choose to apply a design using a combination test if this is only marginally less efficient than a more general design which requires rules for the second stage sample size to be followed precisely in order to protect the type I error rate. In earlier work [4, 5, 6], we argued that choosing group sizes in a group sequential test adaptively has minimal benefit, and a well-chosen error spending design has additional practical advantages while being almost optimal in terms of efficiency. In [1], we acknowledged the appeal of Mehta and Pocock's "promising zone" approach and aimed to provide a decision theoretic basis for choosing the second stage group size, leading to more efficient trial designs within this framework. We are pleased to see this work has stimulated interest and motivated further research.

CHRISTOPHER JENNISON[1] and BRUCE W. TURNBULL[2]

[1]*Department of Mathematical Sciences, University of Bath, UK*

[2]*School of Operations Research and Information Engineering, Cornell University, USA*

References

[1] Jennison C, Turnbull BW. Adaptive sample size modification in clinical trials: start small then ask for more? *Statistics in Medicine* 2015; **34**:3793–3810.

[2] Pilz M, Kunzmann K, Herrmann C, Rauch G, Kieser M. A variational approach to optimal two-stage designs. *Statistics in Medicine* 2019; **38**:4159–4171.

[3] Barber S, Jennison C. Optimal asymmetric one-sided group sequential tests. *Biometrika* 2002; **89**:49–60.

[4] Jennison C, Turnbull BW. Adaptive and nonadaptive group sequential tests. *Biometrika* 2006; **93**:1–21.

[5] Jennison C, Turnbull BW. Efficient group sequential designs when there are several effect sizes under consideration. *Statistics in Medicine* 2006; **25**:917–932.

[6] Hampson LV, Jennison C. Group sequential tests for delayed responses (with discussion). *Journal of the Royal Statistical Society: Series B* 2013; **75**:3–54.

[7] Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with examples. *Statistics in Medicine* 2011; **30**:3267–3284.