MFIT Trial Design

Trial Sample Size and Operating Characteristics

James Totterdell

2023-08-11

Table of contents

1 Introduction

M-FIT is a 4-arm parallel trial evaluating the effect of structured exercise interventions on fatigue as measured by FACIT-Fatigue.

2 Primary Outcome

FACIT-Fatigue is a 13-item 5-point (0 to 4) Likert scale with an aggregate score ranging from 0 to 52. Higher scores imply less fatigue.

The distribution of FACIT-Fatigue scores is skewed with a long tail towards lower scores.

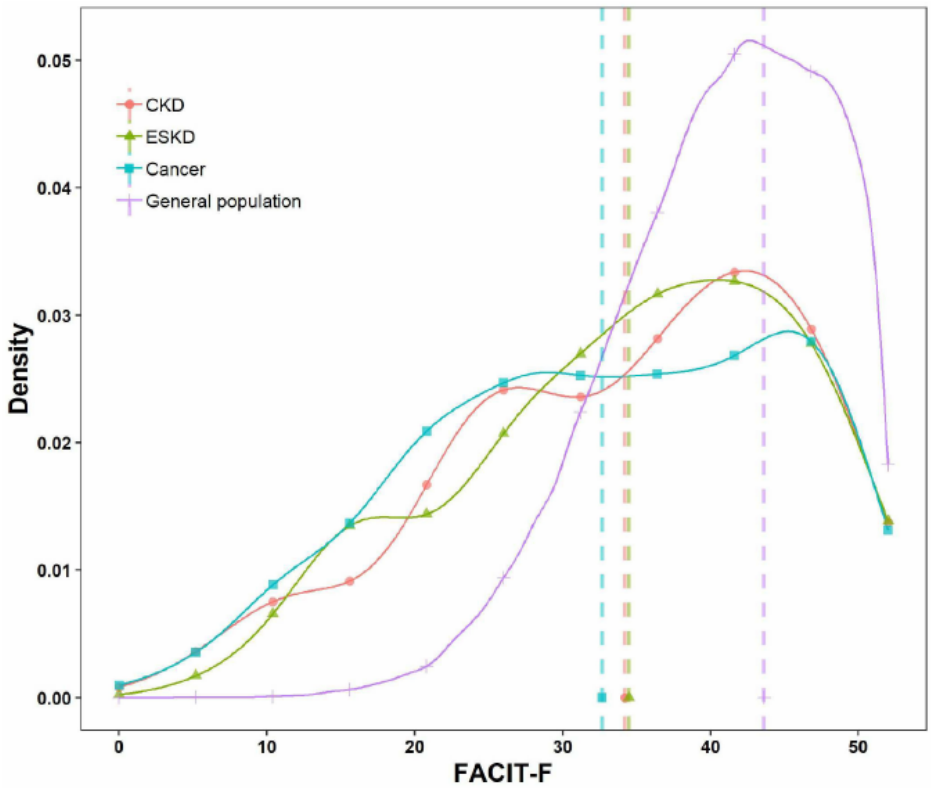

Fig 1.
Distribution of fatigue scores among patients with advanced CKD, ESKD and cancer compared to the US general population. Higher FACIT-F score indicate less fatigue. Density on y-axis shows the probability of distribution of patients. Dotted line depicts the mean FACIT-F score in each patient cohort.
* General population curve extrapolated from Cella (2002)

Figure 1: Example distribution of FACIT-Fatigue scores, (taken from Jhamb M, Abdel-Kader K, Yabes J, Wang Y, Weisbord SD, Unruh M, Steel JL. Comparison of Fatigue, Pain, and Depression in Patients With Advanced Kidney Disease and Cancer-Symptom Burden and Clusters. J Pain Symptom Manage. 2019 Mar;57(3):566-575.e3. doi: 10.1016/j.jpainsymman.2018.12.006. Epub 2018 Dec 13. PMID: 30552961; PMCID: PMC6382584).

The outcome is ordinal, so a ordinal longitudinal model would be ideal. However, given the interest in mean FACIT-Fatigue amongst the groups, for simplicity a normal linear model will be specified for the primary outcome.

3 Fixed-Sample Size Power

Ignoring repeated measures, a sample size of 100 participants per arm, assuming drop-out of 20%, would provide power 0.8 for an effect size of $\Delta = \mu_1 - \mu_0 = 0.45\sigma$. For example, assuming that FACIT-Fatigue scores are distributed with standard deviation of 5, then the sample size would have power 0.8 for an increase of 2.25 in group means from control to active treatment.

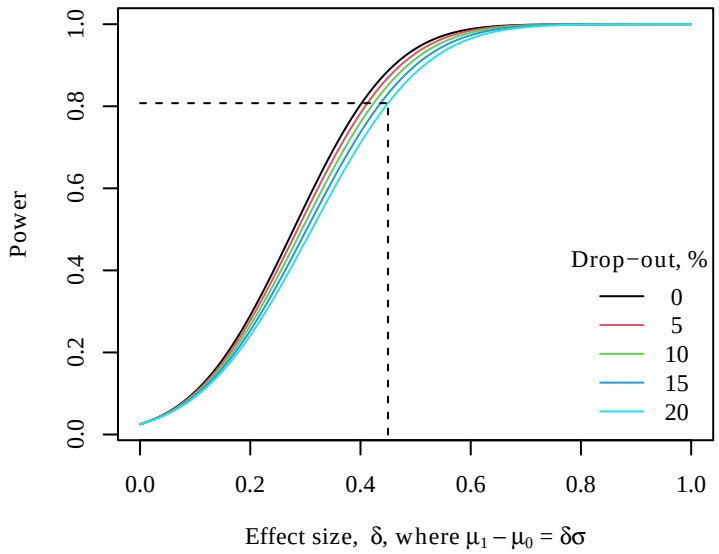

Figure 2: Fixed sample size power for two sample t-test of size $\alpha = 0.025$ with $n = 100$ per arm, assuming varying drop-out and effect size, δ, relative to the standard deviation σ.

4 Group-Sequential Design

4.1 Setup

The study consists of four arms: one control group and three active interventions. Accrual is assumed to be on average 3 participants per week up to a maximum of 400 participants.

Denote the treatment group means at each time-point by $\mu_{jt},\ j = 0, 1, 2, 3$ for $t = 0, 1, 2, 3$ corresponding to baseline, 4, 8, and 12 weeks after randomisation. The primary endpoint is at 12-weeks after randomisation.

The primary quantity of interest is the difference in FACIT-Fatigue at 12-weeks for each active treatment group relative to the control group, that is,

$$\Delta_j = \mu_{j,3} - \mu_{0,3}, \quad j = 1, 2, 3.$$

Another quantity of interest is the relative effectiveness of each active treatment compared to all the other active treatments.

$$\delta_j = \mu_{j,3} - \max_{j' \neq j} \mu_{j',t}, \quad j = 1, 2, 3.$$

and the final quantities of interest are the pairwise comparisons between the active interventions

$$\zeta_{jkl} = \mu_{jk} - \mu_{jl}, \quad j = 1, 2, 3, \quad k, l \in \{1, 2, 3\},$$

in particular, the comparison of the active interventions against the intervention with the highest mean 12-week FACIT-Fatigue scores.

We define an active treatment to be superior if mean FACIT-Fatigue is highest (amongst all active treatments) under that treatment, that is, $\delta_j > 0$, we define an active treatment to be effective, if it has higher mean FACIT-Fatigue compared to control, $\Delta_j > 0$, and ineffective if it has lower mean FACT-Fatigue compared to control, $\Delta_j < 0$.

The probability of these statements are quantified under the assumed model according to

$$\omega_j = \Pr(\Delta_j > 0 | \text{data})$$
$$\pi_j = \Pr(\delta_j > 0 | \text{data})$$

such that ω_j is the posterior probability that treatment j is effective, and π_j is the posterior probability that treatment j is superior amongst the active treatments.

4.2 Model

The FACIT-Fatigue outcomes will be analysed assuming a multivariate normal model for the ordinal outcome. The simulation model focuses only on the 12-week scores.

The simulation model assumes

$$
\begin{aligned}
y_{i,3} | \alpha, \beta, \sigma &\sim \text{Normal}(\alpha + x_{\texttt{trt}[i]}^{\mathsf{T}} \beta, \sigma^2) \\
\mu_{j,3} &= \alpha + x_j^{\mathsf{T}} \beta \\
\alpha &\sim \text{Normal}(40, 5) \\
\beta &\sim \text{Normal}(0, 5) \\
\sigma &\sim \text{Half-}t(3, 0, 5)
\end{aligned}
$$

where $x_{\texttt{trt}[i]}$ for $\texttt{trt}[i] \in \{0, 1, 2, 3\}$ denotes the treatment design vector corresponding to participants i's assigned treatment.

4.3 Interim Analyses

- drop if ineffective
- drop if effective and superior?
- update RAR with bounds (min of 0.1 or 0.15)

5 Example Trials

In what follows, some example trials are presented to show what the results may look like. In these examples, the average FACIT-Fatigue score under each treatment varies while the standard deviation of FACIT-Fatigue scores is assumed constant across all treatments at $\sigma = 10$. Drop-out is assumed to be 80% independent of treatment; the number enrolled excludes drop-outs.

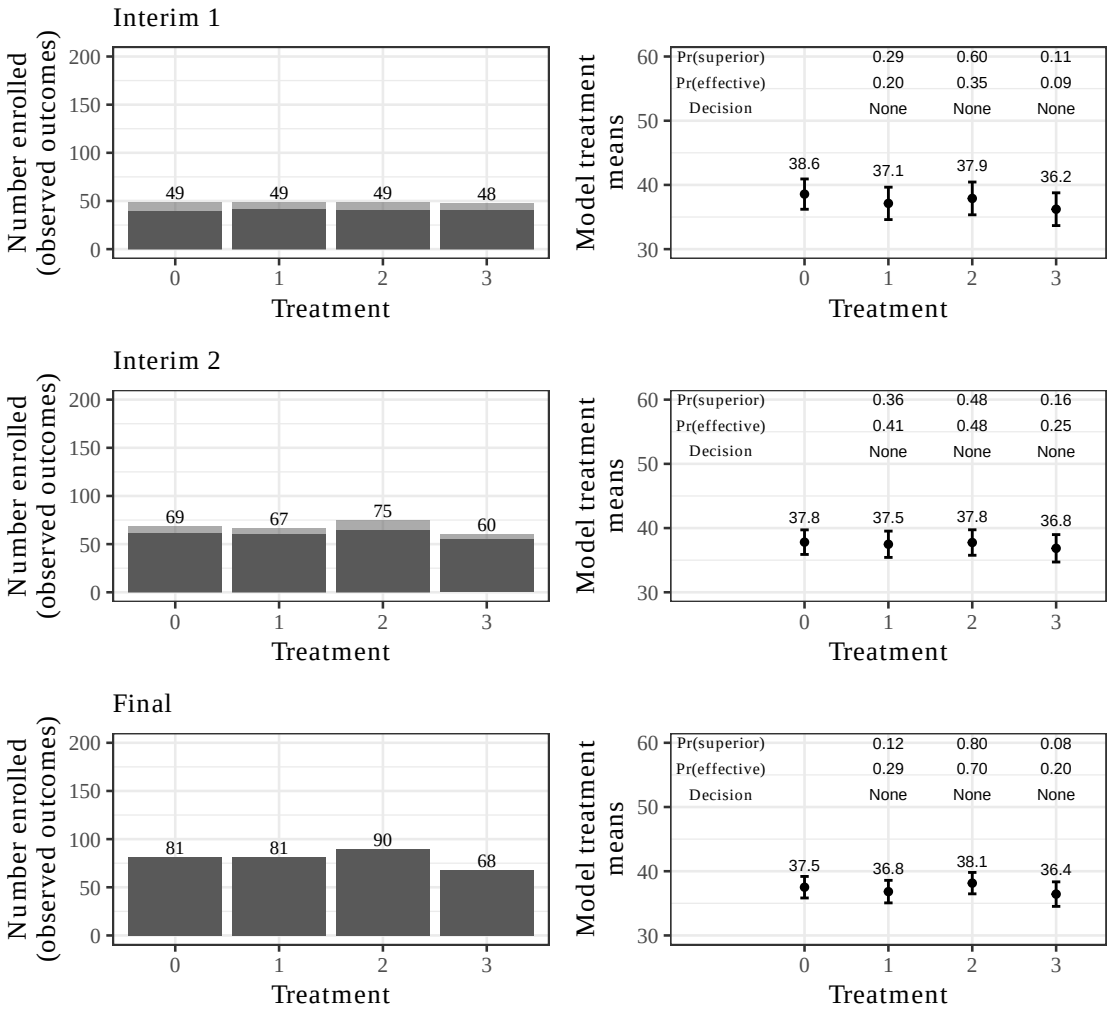

Figure 3: Example 0 - No effect

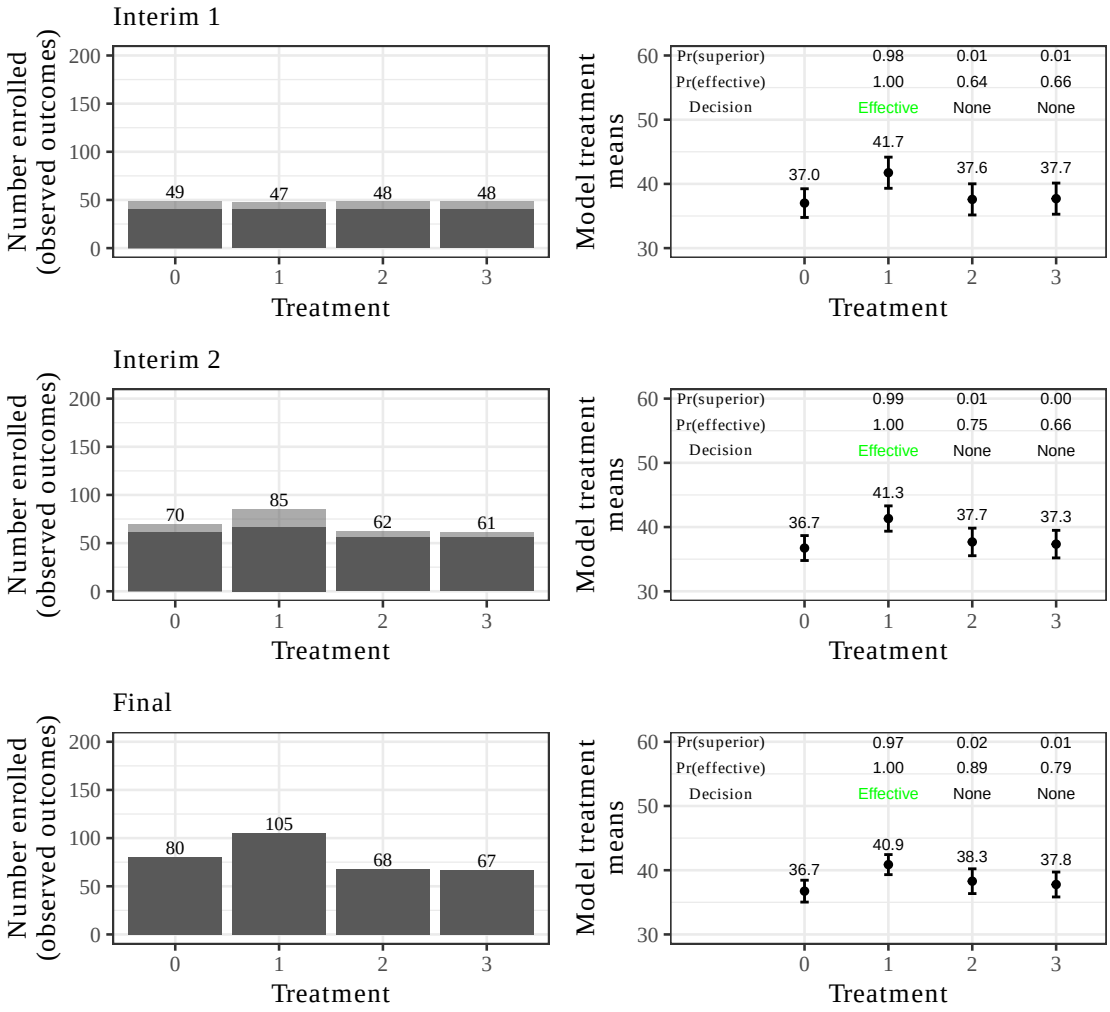

Figure 4: Example 1 - Effective

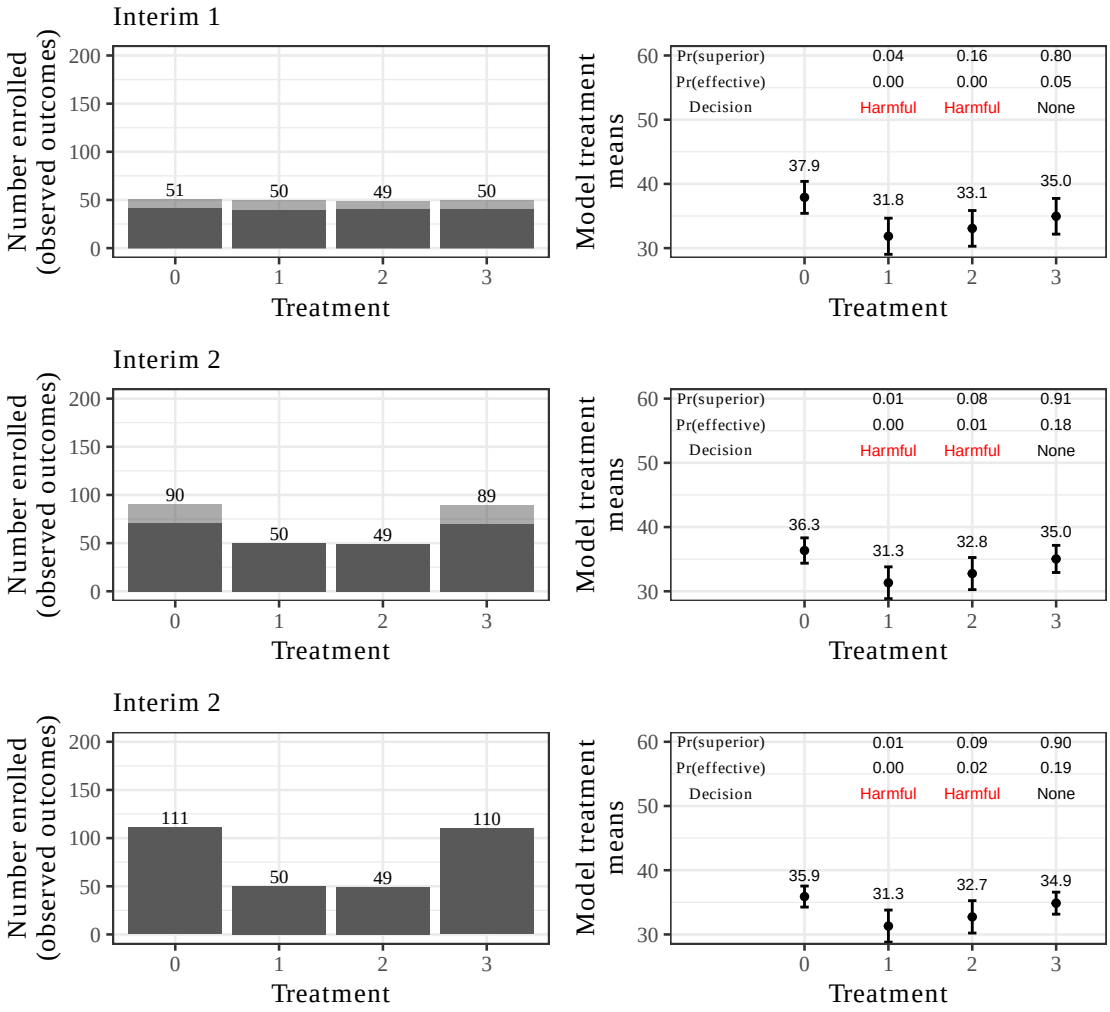

Figure 5: Example 2 - Harmful

6 Group-sequential Operating Characteristics

6.1 Keep Effective Treatment if Superior

Table 1: Summary of trial operating characteristics across all arms, n = 10,000 simulations, $\epsilon = 0.98$.

True means	Any superior and effective	Any superior	Any effective	All effective	Any harmful	All harmful
35.0, 35.0, 35.0, 35.0	0.00	0.00	0.04	0.00	0.07	0.00
35.0, 45.0, 35.0, 35.0	1.00	1.00	1.00	0.00	0.09	0.00
35.0, 40.0, 35.0, 35.0	0.78	0.78	0.92	0.00	0.07	0.00
35.0, 40.0, 40.0, 35.0	0.07	0.04	0.96	0.01	0.05	0.00
35.0, 40.0, 40.0, 40.0	0.01	0.00	0.97	0.54	0.00	0.00
35.0, 45.0, 40.0, 35.0	0.91	0.89	1.00	0.00	0.07	0.00
35.0, 45.0, 40.0, 40.0	0.81	0.78	1.00	0.58	0.00	0.00
35.0, 30.0, 30.0, 30.0	0.00	0.00	0.00	0.00	0.97	0.76

Table 2: Summary of trial operating characteristics, n = 10,000 simulations, BRAR bounded 0.15.

Treatment	Truth	Allocated	Superior	Effective	Ineffective	Active
0 - control	35.0	80				
1	35.0	80	0.00	0.01	0.03	0.97
2	35.0	80	0.00	0.01	0.03	0.97
3	35.0	80	0.00	0.01	0.02	0.98
0 - control	35.0	80				
1	45.0	105	1.00	1.00	0.00	0.99
2	35.0	67	0.00	0.01	0.05	0.95
3	35.0	67	0.00	0.01	0.05	0.95
0 - control	35.0	80				
1	40.0	103	0.78	0.92	0.00	1.00
2	35.0	68	0.00	0.01	0.04	0.96
3	35.0	68	0.00	0.01	0.03	0.97
0 - control	35.0	80				
1	40.0	86	0.02	0.84	0.00	1.00
2	40.0	86	0.02	0.84	0.00	1.00
3	35.0	67	0.00	0.01	0.05	0.95
0 - control	35.0	80				
1	40.0	80	0.00	0.79	0.00	1.00
2	40.0	80	0.00	0.79	0.00	1.00
3	40.0	80	0.00	0.79	0.00	1.00
0 - control	35.0	80				
1	45.0	104	0.89	1.00	0.00	1.00
2	40.0	69	0.00	0.77	0.00	1.00
3	35.0	67	0.00	0.00	0.07	0.93
0 - control	35.0	80				
1	45.0	103	0.78	1.00	0.00	1.00
2	40.0	69	0.00	0.73	0.00	1.00
3	40.0	69	0.00	0.73	0.00	1.00
0 - control	35.0	83				
1	30.0	68	0.00	0.00	0.88	0.12
2	30.0	68	0.00	0.00	0.88	0.12
3	30.0	68	0.00	0.00	0.88	0.12