Supplement A: Post-Hoc Power

Post-hoc power was calculated for the non-inferiority test e using nQuery Version 8.4.0.0 (Stat. Solutions Ltd. & South Bank, Crosses Green, Cork, Ireland). Results are shown in table 2, input values are listed in table 3. Descriptive statistics (see table 3) were calculated using EquivTest (StatCon GmbH, Witzenhausen, Germany).
The nQuery output was:

> When the sample sizes in the groups are 9 and 8, a two group one-sided 0.05 significance level t-test will have 22.26 % power to reject the null hypothesis that the test [(performance of supported non-professionals)] is inferior to the standard [(performance of unsupported professionals)] in favor of the alternative hypothesis that the treatment [(support of non-professionals)] is non-inferior [(to no support of professionals)], assuming that the expected difference in means is 0, a non-inferiority margin of 1 and the common standard deviation is 2.23.

	realistic	optimistic
Test Significance Level, α (one-sided)	0.050	0.050
Non-Inferiority Limit Difference Δ_0	1.000	1.000
Expected Difference, Δ_1	0.000	0.000
Difference of Deltas, $\Delta_0 - \Delta_1$	1.000	1.000
Common Standard Deviation, σ	2.230	1.770
Effect Size, $\delta = \lvert\Delta_0 - \Delta_1\rvert / \sigma$	0.448	0.565
Power (%)	22.26 [80]	29.65 [80]
Group 1 Sample Size, n_1	9 [63]	9 [40]
Group 2 Sample Size, n_2	8 [63]	8 [40]
Sample Size Ratio, n_2/n_1	0.889 [1]	0.889 [1]
Total Sample Size, $N = n_1 + n_2$	17 [126]	17 [80]

Table 2: Post-hoc power calculation results for test e using model *MTE0U-1 / Two Group t-test of Non-Inferiority in Means Unequal n's*. The realistic power calculation assumes the common standard deviation σ found by pooling the two underlying standard deviations (see table 3). The optimistic power calculation assumes the smaller of the two underlying standard deviations (that of the group of supported non-professionals). To achieve a power of 80 % with $\sigma = 2.230$ and sample size ration $n_2/n_1 = 1$, a total sample size of 126 would have been required. With a $\sigma = 1.770$ and sample size ration $n_2/n_1 = 1$, a total sample size of 80 would have been required.

Group	supported non-professionals	unsupported professionals
Mean	11.8333	10.6875
Standard Error (Mean)	0.5892	0.9398
Geometric Mean	11.7135	10.3237
Median	11.5000	11.7500
Standard Deviation	1.7677	2.6583
Variance	3.1250	7.0669
Min	9.0000	5.5000
Max	14.0000	13.0000
Range	5.0000	7.5000
n	9	8
Common standard deviation σ	2.22998	

Table 3: Descriptive statistics of the groups of supported non-professionals and unsupported professionals also used in test e (see results in figure 3). Common standard deviation was estimated by pooling standard deviations of both groups.

Landgraf P, *et al. BMJ Open* 2019; 9:e027563. doi: 10.1136/bmjopen-2018-027563