Abemaciclib with fulvestrant for treating advanced hormone-receptor positive, HER2-negative breast cancer after endocrine therapy

ERRATUM

This report was commissioned by the NIHR HTA Programme as project number 17/141/03

BMJ Technology Assessment Group

This document contains errata in respect of the ERG report in response to the manufacturer's factual inaccuracy check.

The table below lists the page to be replaced in the original document and the nature of the change:

Page No.	Change
29	The 0.60 value in Table B was changed to 0.59
29	The text "However, when compared with the TTD curves by starting dose of abemaciclib (i.e. 150mg vs 200mg populations) shown in Figure I, the ABE-FUL curve for the ITT population is considerably underestimating the time on treatment for patients receiving 150mg of abemaciclib … In fact, using the ITT TTD data leads to considerable underestimation of the ABE-FUL costs in the economic analysis" was amended to "However, when compared with the TTD curves by starting dose of abemaciclib (i.e. 150mg vs 200mg populations) shown in Figure I, the ABE-FUL curve for the ITT population is underestimating the time on treatment for patients receiving 150mg of abemaciclib … It is the ERG's opinion that using the ITT TTD data leads to considerable underestimation of the ABE-FUL costs in the economic analysis"
90	The tick for Zhang 2016 in the limited network for OS has been changed to a cross.
100	The caption changed to refer to Table 16 of the CS
143	The text "In fact, using the ITT TTD curve leads to a considerable underestimation of the ABE-FUL costs in the economic analysis" was amended to "It is the ERG's opinion that using the ITT TTD curve leads to a considerable underestimation of the ABE-FUL costs in the economic analysis".
144	The 0.60 value in Table 26 was changed to 0.59
148	The word "PFS" was changed to "OS" in the caption of Table 27 and in the name of the second column of the table.
150	The text "As a scenario analysis, the ERG included the first-order FP OS curve with a power of -0.5. As explained in Section 4, the FP curve for p = -0.5 has a higher DIC statistic, indicating a worse fit when compared to the ERG's base case of p = -1.5. Nonetheless, given the uncertainty around the relative treatment effect for ABE-FUL compared with the other treatments in the OS NMA, the ERG considered the p = -1.5 curves to be relevant for a scenario analysis as these show a smaller, and thus more conservative, relative treatment effect for ABE-FUL compared with other treatments (Figure 42)." was amended to "The ERG's base case results included the first-order FP OS curve with a power of -0.5, while the power of -1.5 was used as a scenario analysis. As explained in Section 4, the FP curve for p = -0.5 has a higher DIC statistic, indicating a worse fit when compared to the curve for p = -1.5. Nonetheless, given the uncertainty around the relative treatment effect for ABE-FUL compared with the other treatments in the OS NMA, the ERG chose the p = -0.5 curves to as the base case OS curves as these show a smaller, and thus more conservative, relative treatment effect for ABE-FUL compared with other treatments (Figure 42)."

Table A. Median TTD and PFS across comparator treatments

	PFS	TTD	Source
ABE-FUL	16.4	[redacted]	MONARCH 2
FUL	9.3	[redacted]	MONARCH 2
EXE	3.2	3.2	BOLERO 2
EXE-EVE	7.8	EXE:6.8 EVE:5.5	BOLERO 2
EXE-EVE	8.4	Overall: 6.3	BOLERO 6
TMX	9.2	9.2	Milla-Santos, 2001[1]
Chemotherapy	9.6	9.6	BOLERO 6

Most importantly, the estimates shown in Table A indicate that the only treatments where there might be a difference (as far as medians are concerned) between PFS and TTD curves is ABE-FUL, FUL, and EXE-EVE. Therefore, the PFS curves for EXE, TMX and chemotherapy can be used as proxies to estimate TTD in the economic analysis. To estimate TTD for ABE-FUL, FUL, and EXE-EVE, the ERG used the company's proposed methodological approach, but used the MONARCH 2 and BOLERO 2 PFS curves instead of the HR NMA-derived ones. Table B reports the calculations undertaken by the ERG and the resulting HRs used to estimate TTD curves in the economic analysis.

Table B. ERG's HRs to estimate TTD curves

	PFS	TTD	PFS % at median TTD	HR
ABE-FUL	16.4	[redacted]	[redacted]	[log(0.5) / log([redacted])] = [redacted]
FUL	9.3	[redacted]	[redacted]	[log(0.5) / log([redacted])] = [redacted]
EXE-EVE (BOLERO 2)	7.8	EXE:6.8 EVE:5.5	PFS (6.8) = 0.55	[log(0.5) / log(0.55)] = 1.16
EXE-EVE (BOLERO 6)	8.4	6.3	PFS (6.3) = 0.59	[log(0.5) / log(0.59)] = 1.31

Figure F shows the TTD curves when the ERG's HRs were applied to the ERG's FP NMA PFS curves. The TTD curve for ABE-FUL has a considerable separation from the ABE-FUL PFS curve. This is a direct translation of the separation in TTD and PFS KM curves in the ITT analysis of MONARCH 2 data (Figure G and Figure H). However, when compared with the TTD curves by starting dose of abemaciclib (i.e. 150mg vs 200mg populations) shown in Figure I, the ABE-FUL curve for the ITT population is underestimating the time on treatment for patients receiving 150mg of abemaciclib. Given that abemaciclib will be given in 150mg regimens in clinical practice, and that the 150mg population sample size in MONARCH 2 was considerably bigger than the 200mg population, the ERG considers that the 150mg TTD data would have been a more appropriate choice to model TTD for ABE-FUL. It is the ERG's opinion that using the ITT TTD data leads to considerable underestimation of the ABE-FUL costs in the economic

Table 15. Summary of trials used to inform the network meta-analysis (adapted from CS Appendix D.1.3. Table 19)

Trial	Treatment A (ITT n)	Treatment B (ITT n)	Treatment C (ITT n)	Dichotomous		HR NMA		FP NMA			
								Full network*		Limited network**	
				ORR	CBR	OS	PFS	OS	PFS	OS	PFS
BOLERO-2[53]	EXE-EVE (485)	EXE (239)	NA	✓	✓	✓	✓	✓	✓	✓	✓
BOLERO-6[29]	EXE-EVE (104)	EVE (103)	CAP (102)	×	×	×	×	✓	✓	✓	✓
Buzdar 1997[43]	ANAS 1 mg (128)	ANAS 10 mg (130)	MGA 160 mg (128)	✓	✓	✓	✓	×	×	×	×
Buzdar 2001[42]	LTZ 0.5 mg (202)	LTZ 2.5 mg (199)	MGA 160 mg (201)	✓	×	✓	✓	×	×	×	×
Campos 2009[52]	EXE (65)	ANAS 1 mg (65)	NA	×	✓	✓	✓	×	×	×	×
CONFIRM[54, 55]	FUL 500 mg (362)	FUL 250 mg (374)	NA	✓	✓	✓	✓	✓	✓	✓	✓
Dombernowsky 1998[48]	LTZ 0.5 mg (188)	LTZ 2.5 mg (174)	MGA 160 mg (189)	✓	✓	✓	×	×	×	×	×
EFECT[56]	FUL 250 mg (351)	EXE (342)	NA	✓	✓	×	×	×	×	×	×
Hi-FAIR fx[57]	FUL 500 mg (52)	TOR 120 mg (53)	NA	✓	✓	✓	✓	✓	✓	×	×
Howell 2002[50]	FUL 250 mg (222)	ANAS 1 mg (229)	NA	✓	✓	✓	✓	×	×	×	×
Jonat 1996[44]	ANAS 1 mg (135)	ANAS 10 mg (118)	MGA 160 mg (125)	✓	✓	✓	✓	×	×	×	×
Kaufmann 2000[47]	EXE (366)	MGA 160 mg (403)	NA	✓	✓	✓	×	×	×	×	×
Milla-Santos 2001[1]	TMX 40 mg (111)	TOR 60 mg (106)	NA	×	×	×	×	✓	×	×	×
MONARCH 2[27]	ABE-FUL (446)	FUL 500 mg (223)	NA	✓	✓	✓	✓	✓	✓	✓	✓
Muss 1990[46]	MGA 160 mg (86)	MGA 800 mg (84)	NA	✓	×	✓	×	×	×	×	×
PALOMA 3[45]	PAL-FUL (347)	FUL 500 mg (174)	NA	✓	✓	✓	✓	×	×	×	×
Rose 2003[49]	LTZ 2.5 mg (356)	ANAS 1 mg (357)	NA	✓	✓	✓	×	×	×	×	×
SoFEA[58]	FUL 250 mg (231)	EXE 25 mg (249)	NA	✓	✓	✓	✓	✓	✓	✓	✓
Trial 0021[51]	FUL 250 mg (206)	ANAS 1 mg (194)	NA	✓	✓	✓	✓	×	×	×	×
Yamamoto 2013[59]	TOR 120 mg (46)	EXE 25 mg (45)	NA	✓	✓	✓	✓	✓	✓	×	×
Zhang 2016[60]	FUL 500 mg (111)	FUL 250 mg (110)	NA	✓	✓	×	✓	×	✓	×	✓

Figure 11. Forest plot of treatment effects relative to FUL 500 for OS using a fixed-effects HR NMA

Footnotes: The results presented give the median of the posterior distributions as these are less skewed by outlying observations compared to the mean. FUL 250 is not a licensed dose in the UK but was included in the NMA to help connect the full network of comparators.
Abbreviations: ABE: abemaciclib; CrI: credible interval; EVE: everolimus; EXE: exemestane; FUL: fulvestrant; OS: overall survival; PFS: progression-free survival.

Table 16. Adjusted indirect comparison results for TMX vs FUL 500 mg based on Milla-Santos 2001 and the HR NMA (adapted from CS Table 16)

	OS, HR (95% CrI or CI)*	**PFS/TTP, HR (95% CrI or CI)***	**Source**
TOR vs TMX	[redacted]	[redacted]	Milla-Santos 2001[1]
TOR vs FUL 500 mg	[redacted]	[redacted]	NMA
Adjusted indirect comparison TMX vs FUL 500 mg	[redacted]	[redacted]	

*For TOR vs TMX the uncertainty is presented as 95% CI, for TOR vs FUL 500 mg the uncertainty is presented as 95% CrI, but for TMX vs FUL 500 mg the ERG is unsure of the unit of the interval quantifying the uncertainty as it is calculated based on a combination of CrI and CI
Abbreviations: CrI: credible interval; FUL: fulvestrant; HR: hazard ratio; NMA: network meta-analysis; OS: overall survival; PFS: progression-free survival; TOR: toremifene; TMX: tamoxifen.

BOLERO 6 shows a much higher separation in median TTD and median PFS estimates than BOLERO 2, however the company did not include BOLERO 6 in the discussion and therefore did not discuss the differences in median survival estimates. Nonetheless, BOLERO 2 trial's design is superior than that of BOLERO 6, thus the former is likely to be a more robust source of data.

Most importantly, the estimates shown in Table 25 indicate that the only treatments where there might be a difference (as far as medians are concerned) between PFS and TTD curves is ABE-FUL, FUL, and EXE-EVE. Therefore, the PFS curves for EXE, TMX and chemotherapy can be used as proxies to estimate TTD in the economic analysis. To estimate TTD for ABE-FUL, FUL, and EXE-EVE, the ERG used the company's second proposed methodological approach, but used the MONARCH 2 and BOLERO 2 PFS curves instead of the HR NMA-derived ones. Table 26 reports the calculations undertaken by the ERG and the resulting HRs used to estimate TTD curves in the economic analysis.

Figure 34 shows the TTD curves when the ERG's HRs were applied to the ERG's FP NMA PFS curves. The TTD curve for ABE-FUL has a considerable separation from the ABE-FUL PFS curve. This is a direct translation of the separation in TTD and PFS KM curves in the ITT analysis of MONARCH 2 data (Figure 35 and Figure 36). However, when compared with the TTD curves by starting dose of abemaciclib (i.e. 150mg vs 200mg populations) shown in Figure 37, the ABE-FUL curve for the ITT population is considerably lower than the 150mg ABE-FUL TTD curve. Given that abemaciclib will be given in 150mg regimens in clinical practice, and that the 150mg population size in MONARCH 2 was considerably bigger than the 200mg population, it is the ERG's opinion that the 150mg TTD curve would have been a more appropriate choice to model TTD for ABE-FUL. It is the ERG's opinion that using the ITT TTD curve leads to a considerable underestimation of the ABE-FUL costs in the economic analysis. During the clarification period, the ERG asked the company to provide the TTD data for the 150mg and the 200mg populations, however the company has not provided these.

Furthermore, the HRs for the TTD and PFS curves for ABE-FUL and EXE-EVE in BOLERO 2 ([redacted] vs 1.16) suggest that patients in ABE-FUL discontinue treatment before progression at higher rates that EXE-EVE patients.

Given that the HR used to estimate TTD curves in the economic analysis is one of the key model drivers, the ERG advises that the Committee considers the clinical plausibility of the assumptions underlying these clinical data. The ERG also recommends that the 150mg TTD data are used by the company to generate a more robust estimation of the costs of ABE-FUL in the economic analysis.

Finally, the ERG notes the caveat in the approach undertaken to estimate HRs in order to derive TTD curves. The starting point in this approach is to compare median TTD with median PFS values. However, comparison of medians is a reasonably weak approach, as equivalence (or difference) in

median survival estimates does not inform the difference in the curves' shape and doesn't necessarily translate an accurate picture of differences in mean survivals. Nonetheless, given that TTD data were not available for the comparator treatments, the use of median TTD estimates is necessary.

Table 26. ERG's HRs to estimate TTD curves (in months)

	PFS	**TTD**	**PFS % at median TTD**	**HR**
ABE-FUL	16.4	[redacted]	[redacted]	[log(0.5) / log([redacted)] = [redacted]
FUL	9.3	[redacted]	[redacted]	[log(0.5) / log([redacted)] = [redacted]
EXE-EVE (BOLERO 2)	7.8	EXE:6.8 EVE:5.5	PFS (6.8) = 0.55	[log(0.5) / log(0.55)] = 1.16
EXE-EVE (BOLERO 6)	8.4	6.3	PFS (6.3) = 0.59	[log(0.5) / log(0.59)] = 1.31

Figure 34. PFS and TTD in ERG's anaysis

Table 27. Hazard ratios (95% credible interval) for OS

Comparator	OS HR (CrI)
EXE (25 mg) (NMA)	[redacted]
EXE (25 mg)-EVE (10 mg) (NMA)	[redacted]
FUL (500 mg)	Reference
TMX (adjusted indirect comparison)	[redacted]
Abbreviations: ABE: abemaciclib; CrI: credible interval; EVE: everolimus; EXE: exemestane; FUL: fulvestrant.	

5.4.6.1 ERG critique

The CS reports that the CONFIRM population was more pre-treated and thus expected to be at a more advanced stage of the disease compared with the MONARCH 2 population. Clinical expert opinion sought by the ERG agreed that the CONFIRM population was more pre-treated and thus clinical outcomes could be expected to be worse relatively to outcomes in MONARCH 2. Nonetheless, the company used the CONFIRM data to adjust the extrapolated tails of the FUL and ABE-FUL curves in their base case analysis.

The CONFIRM data are considerably rich and complete, with a follow-up period close to seven years, whereas the MONARCH 2 OS data are very immature (with median OS not reached for either treatment arms at the end of the follow-up period of two years and four months). Interestingly, OS for the FUL arm of MONARCH 2 reached 54% at 28 months, while CONFIRM median survival was approximately 27 months (Figure 39). An earlier data cut-off analysis of the CONFIRM data showed a median survival of 25 months.[54, 55] Although the numbers at risk at 28 months in the FUL arm of MONARCH 2 (one patient) require caution when interpreting the OS curve, the 54% survival estimate is not dissimilar to the median OS for the shorter and longer follow-up analysis of the CONFIRM OS data.

Given the immaturity of OS data in MONARCH 2, the ERG advises caution when interpreting all analysis undertaken involving these data. Furthermore, the ABE-FUL and FUL OS curves in the trial show a very small – if any – benefit for ABE-FUL (with the OS HR not being statistically significant), potentially due to data immaturity. Therefore, the ERG sees the additional value in using CONFIRM data in the economic analysis. Furthermore, CONFIRM was included in the HR (and FP) NMA, therefore it should, to a reasonable degree, provide a comparable source of effectiveness for FUL.

Similar to the company's PFS analysis, the ERG disagrees with the company's decision to jointly fit the OS curves to the ABE-FUL and FUL arms of MONARCH 2 instead of using the HR obtained in their base case NMA to estimate the ABE-FUL OS curve. Moreover, given the immaturity of OS data in MONARCH 2, the company could have also considered using the CONFIRM FUL 500mg curve as the baseline FUL curve in the model (rather than the MONARCH 2 FUL curve) to then apply the NMA

forever. The plateau of the OS curves is clearly implausible, and given that it occurs at ~15%, compared to the plateau in PFS curves at ~35%, it also means that PFS and OS curves cross, which is equally implausible. Furthermore, the ABE-FUL and FUL curves cross, indicating that FUL patients might die at slower rates than ABE-FUL patients. This could be a result or the immature shape (and close tracking) of the ABE-FUL and FUL KM curves in MONARCH 2. The company did not provide a discussion of the clinical plausibility of their FP-NMA curves. Instead, it clarified that PFS curves were capped by OS curves, so the former would not cross the latter.

Figure 40. Company's FP-derived OS curves

Figure 41 reports the ERG's FP NMA-derived survival curves. The ERG used a first-order FP NMA, which produced more clinically plausible long-term extrapolations of OS, with virtually all patients being dead at approximately 13 years (160 months). As explained in Section 4, the ERG used the simplified FP NMA which excluded TMX from the network. The CAP curve crosses the ABE-FUL curve at approximately 30 months; however, CAP results should be interpreted with caution, as mentioned in Section 5.4.5.2.

The ERG's base case results included the first-order FP OS curve with a power of -0.5, while the power of -1.5 was used as a scenario analysis. As explained in Section 4, the FP curve for p = -0.5 has a higher DIC statistic, indicating a worse fit when compared to the curve for p = -1.5. Nonetheless, given the uncertainty around the relative treatment effect for ABE-FUL compared with the other treatments in the OS NMA, the ERG chose the p = -0.5 curves to as the base case OS curves as these show a smaller, and thus more conservative, relative treatment effect for ABE-FUL compared with other treatments (Figure 42). The curves also portray a more conservative scenario overall, as OS curves plateau close to zero much earlier than the