Network Meta-Analysis Comparing Palbociclib Plus Fulvestrant With Everolimus Plus Exemestane For HR+/HER2- Advanced Breast Cancer That Has Become Resistant To Previous Endocrine Therapy

LeReun C1, Gentilini A2, Large S1, Chowdhury M3, Smith TA4, Rincio C2

1Independent Biostatistician, Sainte-Anne, Guadeloupe, France; 2Symmetron Limited, London, UK; 3Pfizer Limited, Walton Oaks, UK; 4Pfizer Inc, New York, USA

Introduction

- Fulvestrant (FUL) and everolimus plus exemestane (EVE+EXE) are licensed therapies for advanced breast cancer resistant to prior endocrine therapy.1,2 EVE+EXE was found to be superior in terms of overall survival (OS) compared to exemestane alone in the phase 3 BOLERO-2 trial (hazard ratio [HR]: 0.89; 95% confidence interval [CI]: 0.73, 1.10).3 Palbociclib is a first-in-class selective small-molecule inhibitor of the cyclin dependent kinases (CDKs) types 4 and 6, which play a role in proliferation of breast cancer cells. A phase 3 trial, PALOMA-3, found that palbociclib plus fulvestrant (PAL+FUL) was associated with clinically meaningful improvements in OS versus FUL (HR: 0.81, 95% CI: 0.64, 1.03).4 PAL+FUL is also licensed in breast cancer resistant to prior endocrine therapy,5-7 hence it is of interest to determine the comparative effectiveness of PAL+FUL versus EVE+EXE.

Methods

- An existing systematic literature review (SLR) was updated to identify all relevant published clinical data of pre/per/post-menopausal women with HR+, HER2- locally advanced or metastatic breast cancer receiving first- or second-line therapy and who have been exposed to prior endocrine therapy, either in the (neo)adjuvant or advanced/metastatic setting.8
- Searches were performed in EMBASE, MEDLINE, Medline In-Process, the Cochrane Library and relevant conference proceedings, published up to February 2019. The SLR was performed in accordance with standard guidelines.9
- The OS proportional hazards assumption was assessed using log cumulative hazard and Schoenfeld residual plots for all studies in the network.
- Two NMA methodologies were utilized to conduct analyses using a network of studies linking evidence from PAL+FUL to EVE+EXE.
- Proportional hazards (PH)-based NMA
  - First- and second-order FE models were fitted to the data representing all possible combinations of powers from the following set: −2, −1, −0.5, 0, 0.5, 1, 2, 3.0
  - The model with the lowest DIC was selected as the best-fitting model. Other models with a DIC with 5 points of the best model were also considered as possible candidates. The OS projections obtained from suitable candidates were also checked visually for clinical validity and were compared against Kaplan Meier OS observations.
- Both NMA methodologies were implemented in WinBUGS (version 1.4.3).

Results

- The SLR was conducted in February 2019, 142 publications reporting on 94 unique studies were ultimately found to be relevant. Four of these studies (BOLERO-2, CONFIRM,10 PALOMA-3, and SofEa11) were found eligible for inclusion in the OS NMA (Figure 1). A further study, EFFECT, was included after completion of the SLR using 2007 conference data. These data were not included in the SLR as only conference proceedings up to three years prior to the SLR were hand-searched. EFFECT was identified once the updated SLR was cross-checked against another SLR from NICE appraisal TA99 in a similar population.

Figure 1. Network plot

- Key patient characteristics were found to be similar across trials in the feasibility assessment.
- The PH assumption was met for three of the five studies, including PALOMA-3. The PH assumption was violated for two studies (BOLERO-2 and EFFECT), log cumulative hazard plots of the treatment arms were not parallel and produced significant p-values (p<0.05) when considering the Schoenfeld residuals (Figure 2).

Figure 2. Diagnostic plots showing proportional hazards assumption violations for two studies

- RE models were not run in the PH-based or FP-based NMA as the majority of the connections between interventions were informed by single studies in the OS network, and the ability to reliably estimate between-study variance was very limited. The RE NMA would therefore carry a lot of uncertainty and give results that are not clinically plausible due to extremely wide credible intervals.

PH-based model

- The results of the PH-based NMA showed an OS benefit for PAL+FUL over EVE+EXE with a median HR of 0.74 (95% credible interval [CrI]: 0.51, 1.09).

FP-based model

- Two models met the selection criteria. The best-fitting model was the second-order FE model with powers 0 and 0.5 (DIC: 2333.47), while the second-best-fitting model was the second-order FE model with powers 0.5 and −0.5 (DIC: 2337.39).
- In the best-fitting model, the median HR decreased from 0.87 at 6 months (95% CrI: 0.84, 1.74) to 0.43 at 60 months (95% CrI: 0.17, 1.59), as shown in Figure 3. The second-best fitting model produced similar results (Figure 3), the median HR decreased from 0.88 at 6 months (95% CrI: 0.80, 1.02) to 0.43 at 60 months (95% CrI: 0.13, 1.39).

Figure 3. Median HR (PAL+FUL versus EVE+EXE) and 95% credible interval over time for best FP-based model

- The results of the PH-based NMA are similar to the FP-based NMA at 24 months (difference in median HR: 0.04). However, the 6-month and 12-month median HR estimates from the FP-based NMA are higher with larger credible intervals.

Discussion and Conclusion

- Both the FP-based PH and PH-based NMA found PAL+FUL was associated with increased OS compared to EVE+EXE.
- It was possible to compare PAL+FUL to EVE+EXE indirectly via three additional treatments in a small network using results from studies identified in the SLR.
- The PH assumption did not hold for two of the five studies in the network, suggesting that the FP-based NMA results may be more reliable and valid than those produced by the PH-based NMA approach.
- The results of the NMA are consistent with the Kaplan-Meier curves for both PAL+FUL and EVE+EXE, and PAL+FUL is observed to have a greater survival probability over time compared to EVE+EXE.
- Lack of statistical significance may be expected due to studies not being powered to detect differences in OS, between-trial variance and the small number of studies in the NMA network.