SUPPLEMENTARY APPENDIX

Comparative effectiveness of CDK4/6 inhibitors in metastatic breast cancer: using the target trial emulation framework to investigate overall survival in routine care

Authors:

Full name	Affiliation
Adam M. Brufsky	University of Pittsburgh School of Medicine & University of Pittsburgh Medical Center, Division of Hematology/Oncology, Pittsburgh, PA, USA
Richard S. Finn	Division of Hematology/Oncology, David Geffen School of Medicine at University of California, Los Angeles, CA, USA
Otto Metzger	Breast Oncology Program, Dana-Farber/Brigham and Women's Cancer Center, Boston, MA, USA
Rodrigo Goncalves	Late Development Oncology R&D, AstraZeneca
Cynthia Huang-Bartlett	
Sameet Sreenivasan	Oncology Data Science R&D, AstraZeneca
Ula Nur	
Jessica Davies	
Alex Grigorenko	
Gráinne H. Long	

Correspondence to: Sameet Sreenivasan, Oncology Data Science R&D, AstraZeneca. sameet.sreenivasan@astrazeneca.com

CONTENTS

SUPPLEMENTAL METHODS

Statistical analyses – categorical variables

As the distribution of patient age was similar across the three CDK4/6i treatment groups, and for analytical purposes, we dichotomized age into < 65 years and ≥ 65 years buckets. This age threshold was chosen because it aligns with significant physiological and social milestones in the US healthcare system. American citizens become eligible for Medicare at 65 years old, so this age threshold marks a transition in healthcare access and utilization patterns. Moreover, this age threshold is commonly used to delineate an older adult population in studies ranging from those aiming to understand chronic disease burden in older adults to those identifying barriers to cancer clinical trial participation [1, 2].

To account for the varying durations of follow-up across treatment arms, the calendar year of first-line therapy initiation was included as a covariate in our analysis. Due to differences in the approval dates of the CDK4/6is, we stratified the year of first-line initiation into two approximately equal periods: 2018–2020 and 2021–2024. This binary categorization allowed us to mitigate potential temporal biases and ensure balanced populations across the treatment groups.

Sensitivity analyses

The leveraged dataset contained missing data in categorical variables of race, ECOG PS, socioeconomic status, and disease site. We mitigated the challenge of missing data by running a primary analysis, where missing values were assigned their own category (“Unknown”, missing category imputation), and two secondary sensitivity analyses. One of the secondary sensitivity analyses was a complete case analysis, where only patients with no missing covariate data were included. The other secondary sensitivity analysis was a MICE imputation, where missing values were imputed using the multiple imputation by chained equations approach [3]. We further investigated the type of missingness (MCAR, MAR, and MNAR) in each covariate. Consequently, we performed a sensitivity analysis to understand the implications of possible deviation from the MAR assumption for socioeconomic status and disease site variables. A stabilized inverse-probability weighting (sIPTW) was performed within each imputed dataset

to balance the characteristics of the three arms, and a multivariate Cox model was then fitted to each weighted dataset. The results were pooled using Rubin's rules to obtain hazard ratios and associated CIs.

A structural missing data investigation (using the smdi package [4]) was performed to characterize the missingness mechanisms in ECOG PS, race, disease site, and socioeconomic status. Three groups of diagnostics were evaluated: the average standardized mean difference (ASMD and p Hotelling), the predictability of the missing indicator based on observed covariates (AUC), and the univariate and multivariate association of the missing indicator with the outcome (beta univariate and beta).

A NARFCS sensitivity analysis was performed on socioeconomic status index to determine whether a systematic departure of $\boldsymbol{\delta}$ from the MAR mechanism would affect estimated hazard ratios. To perform the NARFCS sensitivity analysis, we systematically incremented the probability of a missing socioeconomic status being imputed to the lowest possible value (1) and investigated the effect of this probability increment $\boldsymbol{\delta}$ on the estimated hazard ratio for ribociclib.

In the MICE imputation, imputation of ECOG PS values was restricted to values (0,1) seen in the trial-eligible cohort. To determine whether the estimated hazard ratios were affected when this range restriction on imputation was lifted, we utilized an expanded cohort that adhered to all eligibility criteria except for ECOG PS. We then performed a MICE analysis, allowing missing values to be imputed in the full range (0–4) observed in the full patient cohort (trial eligible and non-eligible), filtered each MICE dataset to include only eligible patients (ECOG < 2), estimated the doubly robust hazard ratio for each dataset, and pooled the estimates using Rubin's rules.

Supplementary sensitivity analyses

Supplementary sensitivity analyses considering specific patient and disease characteristics were also conducted. As the primary analysis cohort excluded all patients with a TFI < 12 months associated with any adjuvant ET, we performed a sensitivity analysis (sIPTW) on a cohort that excluded patients with TFI < 12 months who received adjuvant AI only.

Another sensitivity analysis was performed to understand the effect of premenopausal status, as one of the key eligibility criteria that we were unable to emulate in this study was the exclusion of premenopausal

patients due to the unavailability of data on menopausal status in the leveraged dataset. Although ribociclib plus AI has been approved by the US FDA as first-line treatment for pre-/perimenopausal patients [5], there have been no studies explicitly accounting for the prognostic effect of menopausal status in a real-world patient population treated with CDK4/6i, to the best of our knowledge. We therefore resorted to a quantitative bias analysis to understand the effect of the unmeasured confounder of menopausal status. Specifically, we performed an E-value analysis, implemented via episensr package [6], where in our context, the E-value is defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to obscure a specific treatment–outcome association, conditional on the measured covariates [7]. In our scenario, given that our eligibility criteria were mostly based on trials that targeted postmenopausal women [8-10], we assumed that being premenopausal reduces the efficacy of the treatment, and investigated, under this assumption, the circumstances under which an actual difference between two arms (e.g. the palbociclib and ribociclib arms) would be masked in our analysis due to (a) an unobserved higher prevalence of premenopausal patients in the ribociclib arm, and (b) an unobserved association between being premenopausal and the risk of death. We also used the array approach [11]. The array approach is a sensitivity analysis method that assesses the effect of unmeasured confounding by examining a range of possible relationships between the unmeasured confounder and both the exposure and outcome, which can help identify possible scenarios that could alter the conclusions of our analysis. We used the array approach to generate a map of potential risk ratios that would be observed, assuming different degrees of imbalance in the prevalence of premenopausal women between the ribociclib and palbociclib arms, and different degrees of association between premenopausal status and death.

SUPPLEMENTAL RESULTS

Missingness mechanisms and additional sensitivity analyses

An investigation into the missingness mechanisms was performed using various diagnostics (**Supplementary Table 1**), the collective results of which were interpreted to determine the most likely missingness mechanism. For ECOG PS, the ASMD was small, association of the missingness with the outcome was negligible, and unaffected when adjusted for observed covariates. These suggest an MCAR mechanism for the missingness of ECOG PS data. For race and disease site, ASMD was > 0.1, the association with outcome was non-negligible and was affected by the presence of other covariates, indicating a MAR mechanism for missingness of race and disease site data. For socioeconomic status, the ASMD was small, however missingness had an association with the outcome, and was unaffected when adjusting for other covariates, suggesting an MNAR mechanism for missingness of socioeconomic status index data.

Supplementary sensitivity analyses

Excluding patients with a TFI < 12 months who received adjuvant AI only

When excluding patients with a TFI < 12 months who received adjuvant AI only, the resulting cohort consisted of 2828 patients, with 1826, 580, and 422 receiving palbociclib, ribociclib, and abemaciclib, respectively. After sIPTW weighting, the multivariate Cox model estimated the hazard ratios for ribociclib relative to palbociclib to be 1.04 (95% CI: 0.84–1.28) and for abemaciclib relative to palbociclib to be 0.95 (95% CI: 0.75–1.20), consistent with our other findings.

Sensitivity analysis to account for unmeasured menopausal status

In our scenario, given that our eligibility criteria were mostly based on trials that targeted postmenopausal women [8-10], we assumed that being premenopausal reduces the efficacy of the treatment, and investigated, under this assumption, the circumstances under which an actual difference between two arms (palbociclib and ribociclib arms) would be masked in our analysis due to (a) an unobserved higher prevalence of premenopausal patients in the ribociclib arm, and (b) an unobserved association between being premenopausal and the risk of death. Assuming a true hazard ratio between the ribociclib and

palbociclib arms of 0.7, we found that a shift of the hazard to the observed null result of 1.03 (95% CI: 0.82–1.29) (**Fig. 4**) would require that having premenopausal status should increase the likelihood of a patient receiving AI plus ribociclib (vs AI plus palbociclib) by a factor 1.94, and also result in a 1.94 times higher likelihood of death. This suggests that for the unmeasured menopausal status to mask even a moderate effect, it would require this covariate to have an implausibly strong association with death (on the risk ratio scale).

To further investigate the potential of a masked effect of premenopausal status on survival outcomes, we used the array approach [11]. Following this approach, we varied the imbalance in prevalence of premenopausal women between distinct CDK4/6i + AI arms and simultaneously varied the hypothetical association between premenopausal status and survival to estimate the "true" effect size (on the risk ratio scale) that is being masked for each resulting scenario of unmeasured confounding. Based on the age distribution in the palbociclib arm, we assumed the prevalence of premenopausal women in the palbociclib arm would be fixed at 40% and, in the ribociclib arm, which has a greater proportion of younger patients, to be as high as 90% (prevalence confounder: range 20–90%). Due to the absence of quantitative estimates on the prognostic effect of menopausal status, we assumed that being premenopausal has a negative effect on survival, with risk ratios ranging from 1.4 to 2.4 (RR_CD). With these ranges of parameters, an adjusted risk ratio surface was obtained (RR_adj). The example location indicated on the adjusted risk ratio surface shows that even for a moderate effect (0.76) to be obscured by the unmeasured menopausal status, it would require the premenopausal prevalence in the ribociclib arm to be 80% and the association between being premenopausal and death to be implausibly high (2.3 on the risk ratio scale) (**Supplementary Fig. 2**).

TFI and sensitivity analysis with TFI duration as a categorical covariate

In two pivotal CDK4/6i trials (MONALEESA-2 [9] and MONARCH 3 [10]), TFI was used as an exclusion criterion, ensuring that patients with rapid disease relapse following adjuvant ET, who have a poorer expected prognosis on first-line CDK4/6i + ET, were ineligible for enrollment. We emulated this criterion in our hypothetical trial by excluding patients with a TFI < 12 months from the end of any adjuvant endocrine therapy to metastatic recurrence, thereby ensuring that patients included in our cohort were not

endocrine-resistant. To further ensure that differences in the distribution of TFI among the retained patients did not affect the key results, we repeated the sIPTW-weighted analysis with an additional covariate that categorized patients by TFI status (< 24 months or ≥24 months). The results of this analysis were very similar to those of the original sIPTW-weighted analysis and did not change the study's conclusions.

After inclusion of the categorical TFI variable, median OS for abemaciclib, palbociclib, and ribociclib were 60.8 (95% CI: 52.4–not available), 56.0 (95% CI: 52.0–62.0), and 50.4 (95% CI: 42.4–69.0) months, respectively. Hazard ratios for the abemaciclib and ribociclib groups, with palbociclib as the reference group, were 0.96 (95% CI: 0.76–1.23) and 1.03 (95% CI: 0.81–1.3), respectively.

Sensitivity analysis with restricted cohort to align follow-up times

To further ensure that the finding of no observable difference in real-world OS between the CDK4/6is when combined with AI was maintained when treatment cohorts had similar median follow-up times, an additional sensitivity analysis was performed, where eligibility for the target trial was restricted to patients who had initiated first-line therapy between 2018 and 2020 (as opposed to the full study period of January 1, 2018 to August 30, 20242018). This restriction reduced the treatment group sizes (palbociclib: n = 873, ribociclib: n = 103; abemaciclib: n = 130), but median follow-up times were comparable across treatment groups: (palbociclib: 59.4 months, ribociclib: 62.6 months, abemaciclib: 56.8 months). This restriction did not alter the qualitative conclusions regarding overlapping CIs for median OS across treatment groups or hazard ratio CIs that cross 1.

In the unweighted analysis, median OS estimates for palbociclib, ribociclib, and abemaciclib were 55.7 months (95% CI: 51.4–62.2), 53.5 months (95% CI: 42.4–65.9), and 68.2 months (95% CI: 48.5–not available). Hazard ratios in the unweighted cohort were 1.09 (95% CI: 0.82–1.44) and 0.95 (95% CI: 0.72–1.27) for ribociclib and abemaciclib (with palbociclib as reference), respectively (**Supplementary Fig. 3**).

In the sIPTW-weighted cohort (**Supplementary Fig. 4**), median OS estimates for palbociclib, ribociclib and abemaciclib were 55.4 months (95% CI: 51.2–62), 52.6 months (95% CI: 42–not available), and 68.2

months (95% CI: 56.5– not available) respectively, with upper bounds for ribociclib and abemaciclib not estimable due to the reduced effective sample sizes (N_eff = 80 and 87, respectively). Hazard ratios from a Cox-model fitted to the sIPTW-weighted cohort (with palbociclib as the reference) were 1.09 (95% CI: 0.81–1.47) and 0.86 (95% CI: 0.61–1.22) for ribociclib and abemaciclib, respectively (**Supplementary Fig. 4**)

Patient numbers in the Kaplan-Meier risk table in the unweighted analysis for the restricted cohort (**Supplementary Fig. 3**), as well as those for the unweighted analysis presented in the main manuscript (**Fig. 3**), underscore a core limitation of the dataset: insufficient long-term follow-up, yielding identical small numbers at risk for ribociclib and abemaciclib beyond approximately 42 months, and widening CIs for median survival estimates at these later times. This limitation can only be addressed by a more mature dataset that has accrued more patients with extended follow-up and warrants a subsequent study.

SUPPLEMENTAL TABLES

Supplementary Table 1 Multiple diagnostic approaches to investigate missingness mechanisms in covariates

Covariate	ASMD (min/max)[a]	p Hotelling[a]	AUC[b]	Beta univariate (95% CI)[c]	Beta (95% CI)[c]
Metastatic sites	0.142 (0–0.233)	0.015	0.500	−0.32 (−0.65 to 0.01)	−0.36 (−0.69 to −0.02)
ECOG PS	0.057 (0.005–0.168)	< 0.001	0.499	0.02 (−0.14 to 0.17)	−0.01 (−0.17 to 0.15)
Race	0.165 (0.038–0.311)	< 0.001	0.500	0.40 (0.21–0.59)	0.35 (0.16–0.54)
Socioeconomic status	0.085 (0.042–0.269)	0.017	0.500	0.44 (0.22–0.66)	0.43 (0.21–0.65)

ASMD, average standardized mean difference; AUC, area under the curve; beta, beta coefficient; CI, confidence interval; ECOG PS, Eastern Cooperative Oncology Group performance status

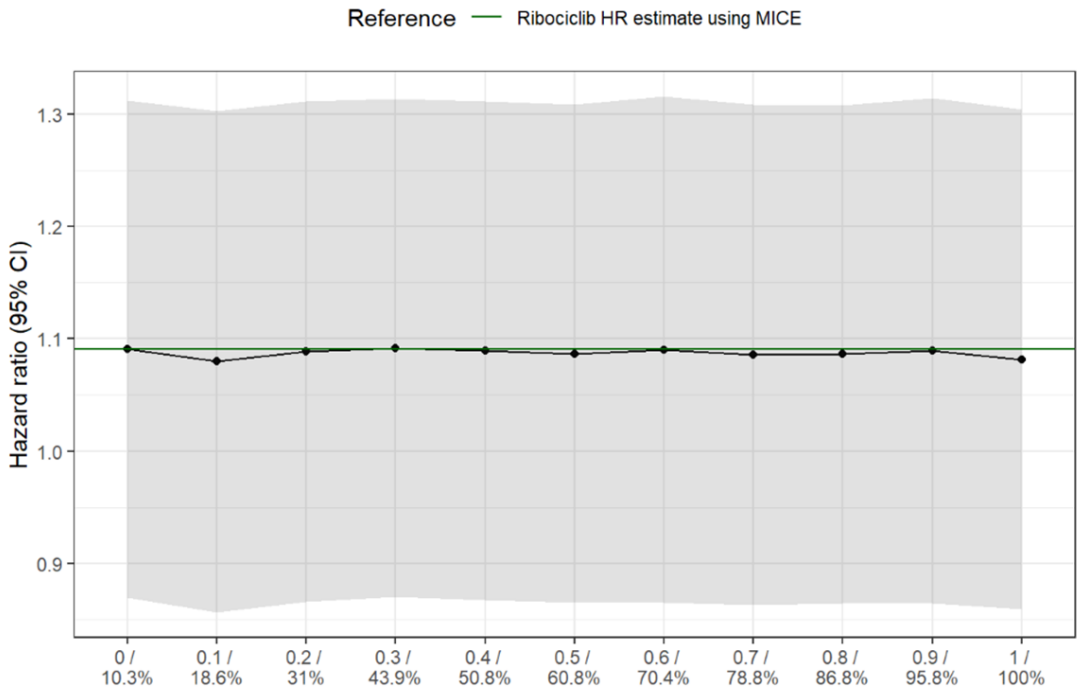
[a]Group 1 diagnostic: differences in patient characteristics between patients with and without covariate

[b]Group 2 diagnostic: ability to predict missingness

[c]Group 3 diagnostic: assessment if missingness is associated with the outcome (univariate, adjusted)

SUPPLEMENTAL FIGURES

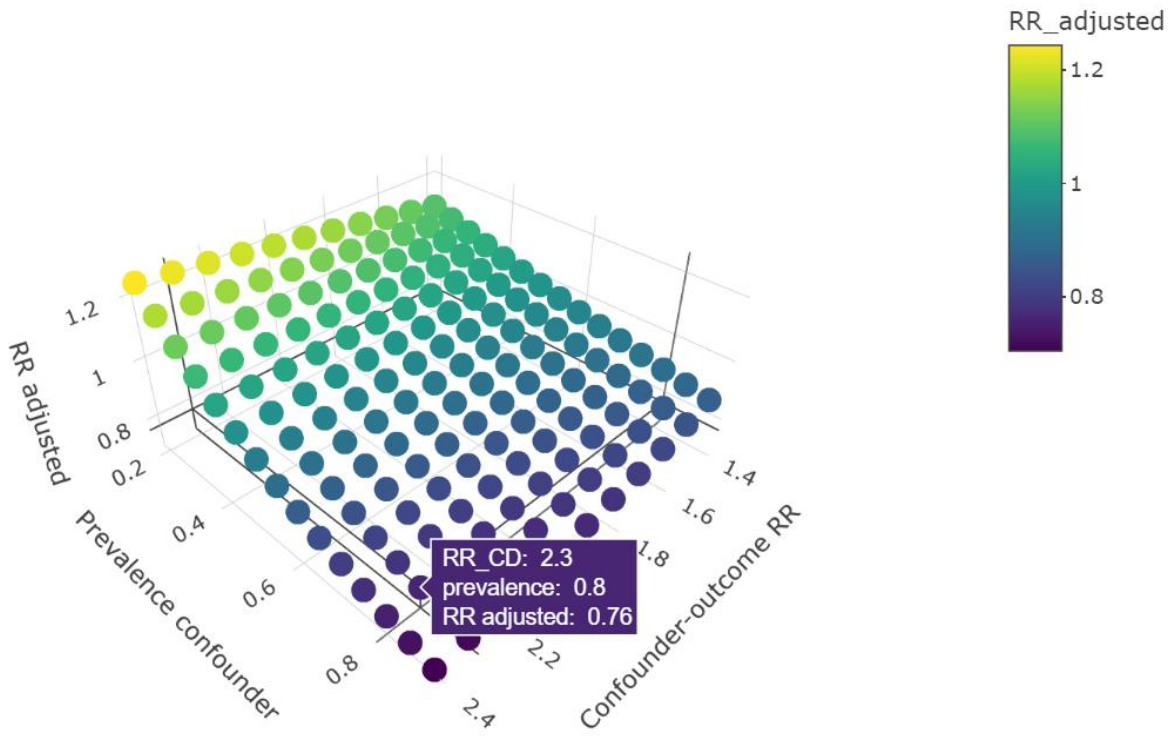
Supplementary Fig. 1 Sensitivity analysis on impact of an MNAR (delta-shifted) socioeconomic status imputation using the NARFCS method



CI, confidence interval; HR, hazard ratio; MICE, multiple imputation by chained equations; MNAR, missing not at random; NARFCS, not at random fully conditional specification

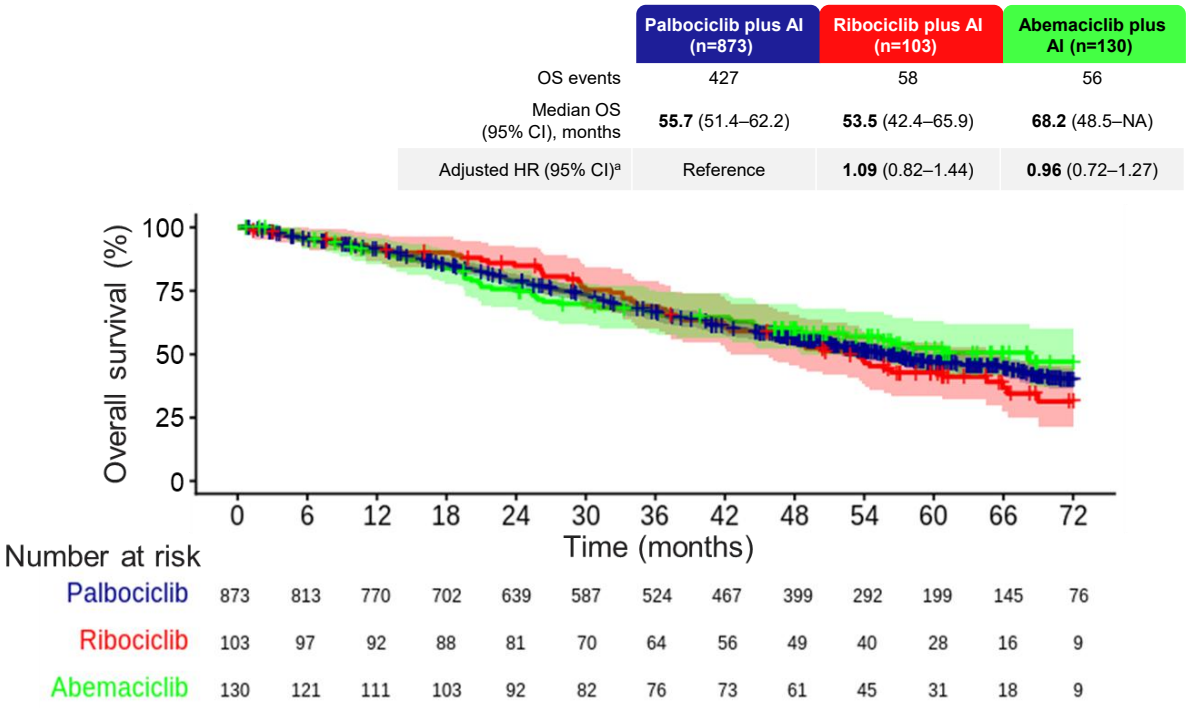
The probability of a missing socioeconomic status was imputed to the lowest possible value (1). The sensitivity parameter $\boldsymbol{\delta}$ is shown against the estimated hazard ratio for ribociclib, as obtained through a MICE approach.

Supplementary Fig. 2 Adjusted risk ratio between ribociclib and palbociclib arms as a function of varying premenopausal prevalence and premenopausal status-death association



RR, risk ratio; RR_adjusted, adjusted risk ratio (true effect size [ribociclib vs palbociclib] on the risk ratio scale that is being masked as a result of the unmeasured confounding); RR_CD, risk ratio between confounder and death (association on the risk ratio scale between the confounder [premenopausal status] and outcome (death)

Supplementary Fig. 3 Unweighted overall survival after restricting the cohort to patients who initiated first-line therapy prior to 2021

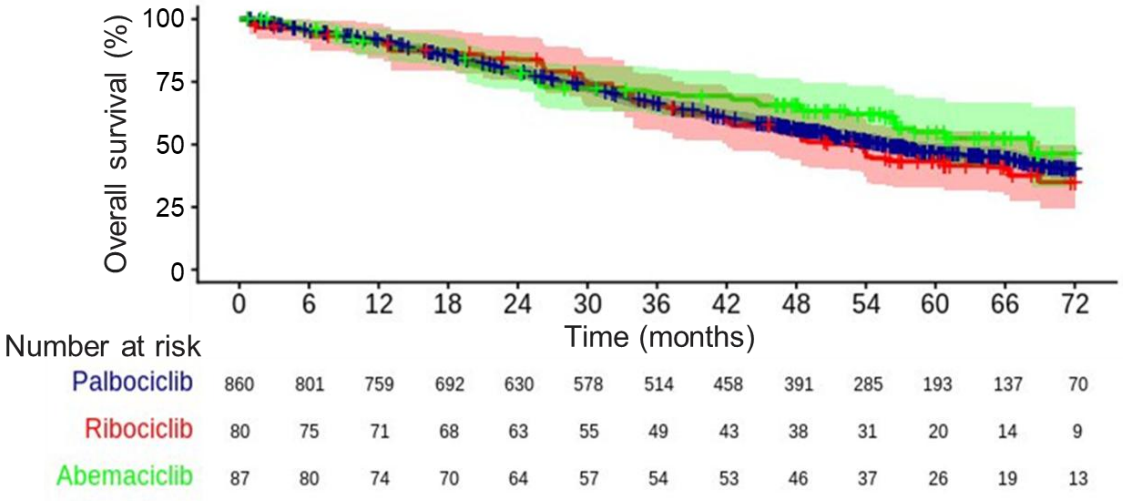


AI, aromatase inhibitor; ECOG PS, Eastern Cooperative Oncology Group performance status; ET, endocrine-based therapy; HR, hazard ratio; mBC, metastatic breast cancer; NA, not available; OS, overall survival

HRs were adjusted for age (< 65 years/≥ 65 years), race (white/black/Asian/other), socioeconomic status (ranked from 1 [lowest] to 5 [highest]), ECOG PS (0/1), *de novo* mBC (yes/no), prior adjuvant AI received (yes/no), prior adjuvant ET received (yes/no), first-line AI partner (anastrozole/letrozole/exemestane), metastatic disease site (liver/non-liver, visceral/non-visceral), and year of first-line therapy initiation (2018/2019/2020).

Supplementary Fig. 4 sIPTW-weighted overall survival after restricting the cohort to patients who initiated first-line therapy prior to 2021

	Palbociclib plus AI (n=860)	Ribociclib plus AI (n=80)	Abemaciclib plus AI (n=87)
OS events	422	45	37
Median OS (95% CI), months	**55.4** (51.2–62)	**52.6** (42–NA)	**68.2** (56.5–NA)
Adjusted HR (95% CI)[a]	Reference	**1.09** (0.81–1.47)	**0.86** (0.61–1.22)



AI, aromatase inhibitor; ECOG PS, Eastern Cooperative Oncology Group performance status; ET, endocrine-based therapy; HR, hazard ratio; mBC, metastatic breast cancer; NA, not available; OS, overall survival; sIPTW, stabilized inverse-probability weighting.

HRs were adjusted for age (< 65 years/≥ 65 years), race (white/black/Asian/other), socioeconomic status (ranked from 1 [lowest] to 5 [highest]), ECOG PS (0/1), *de novo* mBC (yes/no), prior adjuvant AI received (yes/no), prior adjuvant ET received (yes/no), first-line AI partner (anastrozole/letrozole/exemestane), metastatic disease site (liver/non-liver, visceral/non-visceral), and year of first-line therapy initiation (2018/2019/2020).

REFERENCES

1. Jindai K NC, Vorderstrasse BA, Quiñones AR. Multimorbidity and functional limitations among adults 65 or older, NHANES 2005–2012. Prev Chronic Dis. 2016;13:160174. doi: 10.5888/pcd13.160174.

2. Sedrak MS, Freedman RA, Cohen HJ, Muss HB, Jatoi A, Klepin HD, et al. Older adult participation in cancer clinical trials: A systematic review of barriers and interventions. CA Cancer J Clin. 2021;71(1):78-92. doi: 10.3322/caac.21638.

3. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67. doi: 10.18637/jss.v045.i03.

4. Weberpals J. smdi: Perform Structural Missing Data Investigations. R package version 0.3.1, https://janickweberpals.gitlab-pages.partners.org/smdi/. 2024.

5. Novartis Pharmaceuticals Corporation: Kisqali® (ribociclib) prescribing information. https://www.novartis.com/us-en/sites/novartis_us/files/kisqali.pdf (2024). Accessed 6 March 2025.

6. Haine D. The episensr package: basic sensitivity analysis of epidemiological results. R package version 1.3.0. https://dhaine.github.io/episensr/. 2023. doi: 10.5281/zenodo.8299430.

7. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med. 2017;167(4):268-74. doi: 10.7326/M16-2607.

8. Finn RS, Martin M, Rugo HS, Jones S, Im SA, Gelmon K, et al. Palbociclib and letrozole in advanced breast cancer. N Engl J Med. 2016;375(20):1925-36. doi: 10.1056/NEJMoa1607303.

9. Hortobagyi GN, Stemmer SM, Burris HA, Yap YS, Sonke GS, Paluch-Shimon S, et al. Ribociclib as first-line therapy for HR-positive, advanced breast cancer. N Engl J Med. 2016;375(18):1738-48. doi: 10.1056/NEJMoa1609709.

10. Goetz MP, Toi M, Campone M, Sohn J, Paluch-Shimon S, Huober J, et al. MONARCH 3: abemaciclib as initial therapy for advanced breast cancer. J Clin Oncol. 2017;35(32):3638-46. doi: 10.1200/JCO.2017.75.6155.

11. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006;15(5):291-303. doi: 10.1002/pds.1200.