Reassessment of Progression-Free Survival as a Surrogate End Point of Overall Survival

Hsuan-Hui Grace Liu

Honors Thesis
UNC Eshelman School of Pharmacy
University of North Carolina at Chapel Hill

3.2023

Approved: Yanguang Cao

Abstract

Introduction: Progression-free survival (PFS) has been increasingly used as a surrogate endpoint for overall survival (OS) by FDA in cancer drug approval. However, whether PFS can fully represent OS has remained uncertain. Breast cancer is the most prevalent cancer among females, and PFS has been the most common endpoint in drug approval trials for decades. Nevertheless, studies have shown conflicting results on whether PFS correlates with OS among subtypes of breast cancer and across different treatments. This study aims to reassess the correlation between PFS and OS in breast cancer, and evaluate under what circumstances can we consider using PFS as a surrogate of OS.

Methods: We conducted a meta-analysis of randomized clinical trials that assessed both PFS and OS for drug efficacy in patients with breast cancer using individual patients' data extracted from Project Data Sphere. Trials were included if both PFS and OS were assessed and provided in the datasets. Descriptive analyses were conducted to assess the heterogenicity across studies. Pearson and other correlation analyses were performed in individual trials to evaluate the correlation between PFS and OS in each study, and the iQWIG guideline was used to determine the strength of correlation. Cox proportional hazards regression (Cox) was performed to detect predictors of survival and ultimately build a survival prediction model. The analysis for subgroup correlation evaluation is still to be determined to find a higher correlation in certain subgroups of patients.

Results: Two out of 13 (dat200, dat158) breast cancer studies that were available through the Data Sphere Project met inclusion criteria. The r values of correlation were 0.70, and 0.66, for dat200 and dat158 respectively. Cox analysis results from both studies showed that PFS was a positive prognostic indicator for survival, while older age was a negative prognostic predictor. Other covariates such as pretreatment tumor size, HER2, receptor status and race also played a role in predicting survival. Correlation subgroup analyses of dat200 showed higher correlation coefficients in several subgroups of patients

Conclusions: We concluded that PFS is poorly correlated with OS according to the correlation coefficient, r value $\leq$ 0.7 based on iQWIG, and thus is not an ideal surrogate of OS in evaluating drug efficacy in breast cancer. PFS and age, along with receptor status, HER2, pre-treatment tumor size and race may be used to predict survival. Further assessments are required to validate the subgroup analyses and determine which population would benefit from using PFS as a surrogate for OS.

Introduction

OS has been the gold standard to assess the clinical benefit of cancer therapies and has been used by the FDA for cancer drug approval. OS is defined as the time from randomization until death from any cause. It is considered the most reliable endpoint that directly assesses survival. It is easy to measure without bias in the measurement.[1] However, there are some limitations of evaluating OS in clinical trials. One is that the multiple subsequent therapies patients receive after the study can confound survival analysis. Another challenge is the difficulty to follow up and obtain the results with patients after an extended survival time.

Recently, PFS, one of the tumor-based surrogates, has been used by FDA for cancer drug approval. PFS is defined as the time from randomization till objective tumor progression or death, whichever occurs first. It reflects tumor growth on top of the survival analysis, which is not affected by subsequent therapies. It can be obtained before the determination of survival and thus can potentially accelerates the drug approval process and benefits patients who urgently need new medications.[2] However, one of the drawbacks is the bias in evaluating tumor progression. Another huge issue that remained unsolved is whether PFS correlates with OS and can serve as a surrogate of OS for drug approval.

Several studies have been performed to evaluate the correlation between PFS and OS. A review of several meta-analyses of level 1 RCT in oncology has found a poor correlation between PFS and OS.[3] They included various types of cancer types and therapies. Among 36 studies with 65 individual trials, they found that more than half of trial-level correlations were of low strength; less than one-fourth were highly correlated. Notably, among 15 correlations of high strength, 6 occurred in the adjuvant setting and 6 in the metastatic setting; 3 high correlations occurred in adjuvant colorectal cancer. These findings indicated that a higher correlation might be associated with a certain cancer type or drug therapy. Another study evaluated superiority-design oncologic RCTs from the ClinicalTrials.gov database for patients with metastatic solid tumors. Among 82 RCTs with both PFS and OS as endpoints, only 31 (37.8%) trials with a positive PFS signal showed a subsequent OS benefit.[4] In a systematic review of trial-level meta-analyses that reported individual trial-level correlations, a low correlation between PFS-OS was found in 48%.[5] Despite a generally poor correlation across all cancer types and treatments, several reviews of meta-analyses have found a better correlation in colorectal cancer,[6,7] indicating that some cancer types may be associated with a higher correlation.

PFS has been approved by FDA as a surrogate endpoint to evaluate the efficacy of drug candidates in clinical trials for the treatment of several cancers, which includes breast cancer.[8] Breast cancer is the most common cancer type among females in the US, ranking 1st in new cases and 2nd in cancer death rate.[8] The good news is with proper treatments, the survival can be long, with the 5-yr survival rate of 90% in localized breast cancer. However, the correlation between PFS and OS in breast cancer remains unclear and varies across different subtypes, treatments, and statistical analyses.[7,9] Plus, most meta-analyses that assess the correlation between PFS and OS in breast cancer patients used aggregate patient data for analyses instead of extracting individual patients' data.[10,11,12] Using aggregate data is less time-consuming and is expected to yield similar results as using individual patients' data, yet the latter method holds several advantages with regard to data control and checking, such as allowing adjustment for the same variables across studies, data and analyses checking, and permits the use of time-to-event data which is specifically helpful for estimating survival. Using individual patients' data also facilitates exploration of heterogeneity at the patient level and subgroup analyses of patient level data, and provides the opportunity to address questions not addressed in the original publication.[13] In sum, whether PFS can be used as a surrogate for OS in breast cancer is still unclear and should be reevaluated using individual patients' data.

As a result, our research aims to (1) Re-evaluate the correlation between PFS and OS in breast cancer using individual patient data. (2) Determine which types of patients benefit the most from using PFS as a predictor for OS. (3) Create a model that predicts overall survival using various covariates.

Method

We conducted a meta-analysis of randomized clinical trials that assessed both individual-level PFS and OS for drug efficacy in patients with breast cancer. This study followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines for systematic reviews and meta-analyses.[14]

Data Source and Study Selection

We searched studies on Project Data Sphere, an open-access platform that aggregate trial data from pharmaceutical companies, medical centers, and government organizations. We used the search term of "breast cancer" to look for trials that were performed in a breast cancer population. We included studies meeting the following criteria: (1) both PFS and OS were study endpoints and were assessed in the trials; (2) the definition of PFS and OS followed FDA's guidance;[12] (3) both PFS and OS values were provided in the datasets. Trials that only one of the treatment arms were available were also included for further analyses. We excluded studies that were not focused on survival as the endpoints, and if PFS and OS were assessed but data were not accessible through Project Data Sphere.

Data Extraction and Study Endpoints

The study designs, inclusion/exclusion criteria and definition of endpoints were documented to allow for assessments of the heterogeneity across different trials. A number of covariates applicable to predictive survival were extracted from studies that met inclusion criteria for further analyses, including demography (gender, age, race), diseases (ER/PgR, HER2, stage, pretreatment tumor size, metastatic sites, etc), prior and trial treatments, and outcomes (PFS, OS). Based on the FDA's guidance, OS was defined as the time from randomization until death from any cause; PFS was defined as the time from randomization till objective tumor progression or death, whichever occurs first.

The primary endpoint was to re-evaluate the correlation between PFS and OS in breast cancer patients. The secondary endpoints were to assess the subgroup of patients that have a higher correlation between PFS and OS, and thus can benefit from using PFS as a surrogate of OS, and to build a model that can predict survival of breast cancer patients.

Data Synthesis and Analysis

Analyses were performed using RStudio. Descriptive statistics were performed on each individual dataset to assess patients' characteristics, cancer status/stage, treatment, and outcomes. Data are reported as counts (% of total population), mean, or median.
In one of the studies, more than 89% of patients had no event (death or disease progression) at the end of the study. They inform no correlation between PFS and OS and can potentially cause immortal time bias. To obtain more accurate results, we decided to remove these patients from further analyses.
Correlation analyses were performed on the dataset from each individual cohort. The r values were used to interpret the validation of the strength of correlation according to the guidance published by the Institute for Quality and Efficiency in Health Care (iQWIG),[15] an independent institute that assesses the quality and efficiency of medical treatments. Based on the guidance, the strength of correlation is classified as low ($r \leq 0.7$), medium ($r > 0.7$ to $r < 0.85$) or high ($r \geq 0.85$). Cox proportional hazards regression (Cox) was used to build a predictive model for survival analysis as well as to detect prognostic

indicators for survival. The analysis to evaluate which subgroups of patients would have a higher correlation and would benefit from using PFS as OS is still to be determined. In this study, we performed two analyses, including one that used a correlation analysis in patients grouped by 1-2 variables and the second using a mixed effect model to assess the relationship between post-progression analysis and covariates.

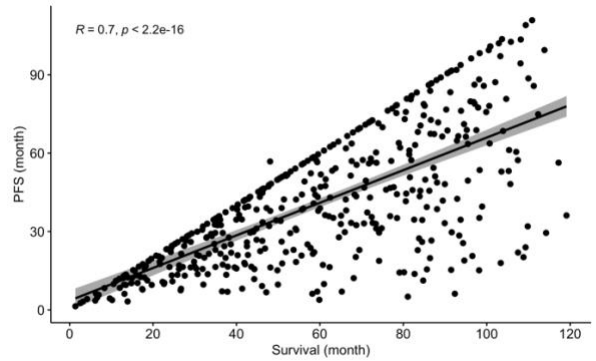
Result

Two out of 13 breast cancer studies met inclusion criteria and thus were included in our study for further analyses (dat200[16] and dat158[17]). The study design, patient and disease characteristics and study outcomes were summarized in Table 1. The major differences we noticed was that dat200 included patients from early stage of breast cancer (no prior chemo, longer PFS and OS) while patients in dat158 had metastatic disease and a shorter survival.

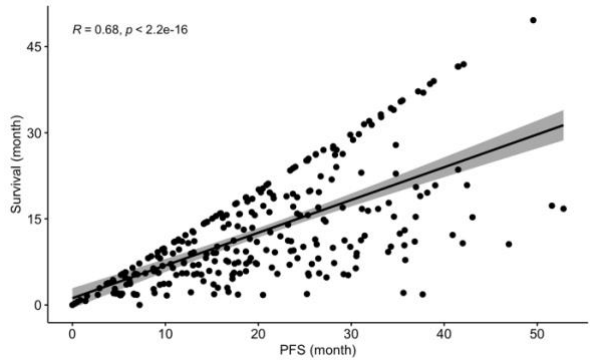
	Dat200*	**Dat158**
Study Population	women with 0-3 positive axillary lymph nodes with disease progression or death	chemotherapy-naïve, metastatic breast cancer
Treatment	2*2 factorial study comparing adjuvant paclitaxel/cyclophosphamide vs doxorubicin hydrochloride	Received either nab-paclitaxel or ixabepilone or paclitaxel; +/-bevacizumab
Patient population	437	283 from paclitaxel arm
Gender (female %)	100%	98%
Race (white %)	83%	78%
Age (years)	86% between 40~70	57 (median)
ER/PgR (either positive %)	55%	71%
HER2 (positive %)	6.9%	<1%
Stage	N/A	80% stage IV
Histologic grade	1: 9.1% 2: 34% 3: 57%	N/A
Tumor size	<<2cm: 54% 2~5 cm: 43% >5 cm: 3%	N/A
Prior chemo	0%	Prior taxane: 44%
Prior hormone therapy	6.5%	N/A
Progress, Death (%)	100%, 61%	79%, 50%
PFS (month, mean)	36 (20~57)	10 (6,17)
OS (month, mean)	59 (36,83)	19 (12,28)

Table 1. Baseline characteristics of each study.
*89% of patients who had no death or disease progressed were excluded.

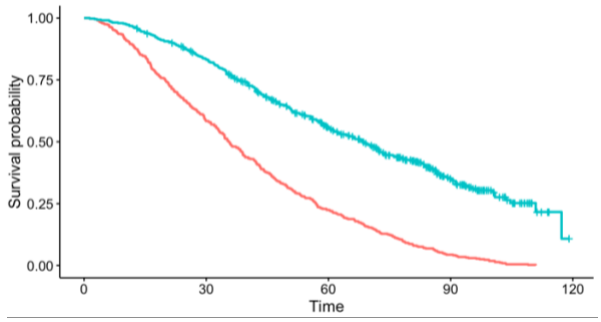
Shapiro-Wilk Test showed that PFS and OS in dat200 were not normally distributed. Spearman correlation was performed, and the correlation R values were 0.70. For dat158, both PFS and OS were not normally distributed, and the Spearman correlation R was 0.68. The correlation and survival curves for each dataset were listed in **Graph 1**.



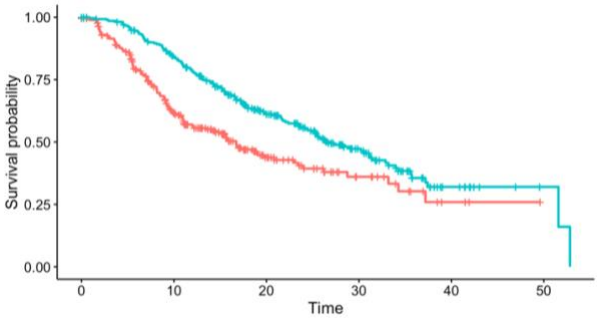
Graph A. Correlation of dat200; r=0.70



Graph B. Correlation of dat158; r=0.68



Graph C. Survival curve for dat200.



Graph D. Survival curve for dat158.

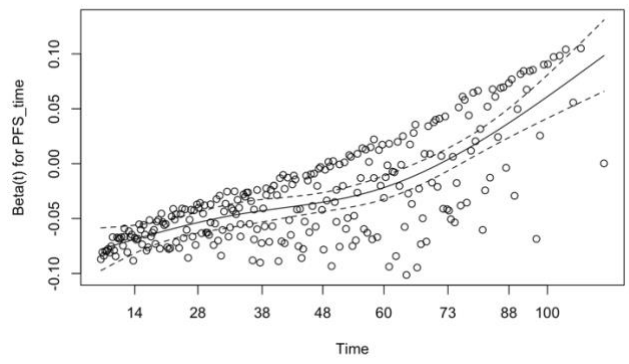
Graph 1. Correlation and survival curves for dat200 and dat158, respectively. PFS is the red line and OS is the teel line.

Some potential predictive variables were included in the Cox analyses if available in the datasets (trial treatment, histologic grade, stage of disease, number of positive nodes, tumor size, age, receptor status, HER2 status and prior hormone therapy). The results of dat200 and dat158 with variables showing significance were listed in Table 2. Pr < 0.05 indicated the covariates were significant. Coef is the regression coefficients, with a positive sign indicating a higher hazard (risk of death) and thus worse prognosis. For example, the positive coef of tumor size > 5 cm and age ≥ 60 yo of dat200 suggested that they were poor prognosis for survival than their comparison group, while negative coef of PFS time indicated it as a positive prognostic factor for OS. Exp(coef) is hazard ratios and give the effect size of covariates. Age ≥ 60 yo in dat200 had a hazard ratio of 1.577, indicating a 1.577 more chance of dying than comparison group (age < 60 yo). Dat158 also showed similar prognostic results in PFS time and age. ER/PgR negative, older age, race of black were associated with higher hazard while HER2 positive was a positive prognostic factor of OS.

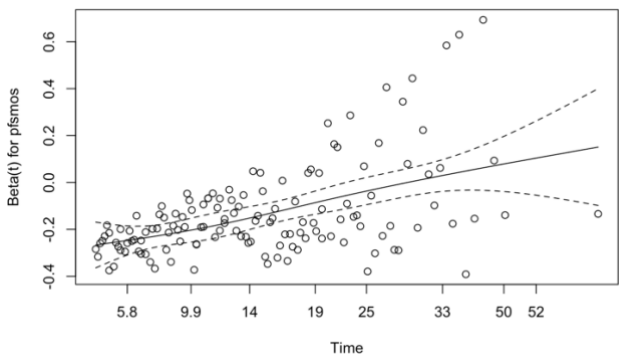
Dat200	**Coef**	**exp(coef)**	**Pr(>|z|)**
Tumor size > 5cm[&]	0.755	2.127	0.0174
Age ≥ 60 yo[&&]	0.456	1.577	0.005
PFS time	-0.029	0.972	<2e-16
Dat158	**Coef**	**exp(coef)**	**Pr(>|z|)**
ER/PgR both negative*	0.784	2.190	0.0002
Age 50-69 yo	0.905	2.472	0.001
Age => 70 yo	1.270	3.560	0.0006
Race = black**	0.456	1.577	0.005
HER2 positive***	-0.029	0.972	<2e-16
PFS time	-0.135	0.874	8.56e-13

Table 2. Summary of Cox analysis
[&]Compare with size < 2 cm. [&&]Compare with age < 60 yo
*Compare with either positive **Compare with white ***Compare with negative

We checked the hazard of tumor size, age, and PFS for dat200. The hazard of tumor size and age remained constant, whereas the hazard for PFS increased over time. The hazards for dat158 were similar: the hazard of hormone receptor, HER 2, race and age remained constant while PFS trended up gradually over time.



Graph A. Hazard of dat200.



Graph B. Hazard of dat158.

Graph 2. Hazard of PFS for dat200 and dat158, respectively.

The subgroup analysis using proposal 1 (running correlation among 1-2 variables) in dat200 are listed in Table 3. We included covariates with correlation higher than baseline (0.7) and number of patients in each subgroup > 20. The findings for dat158 and results of the mixed effect model are still underway. Due to a small sample size in dat158, we assumed the number of patients in each group would be very few and the correlation would be falsely high due to the small number.

Variable/group	COR	Number of patients
HER positive	0.88	30
Prior hormone treatment	0.81	28
Age 50~<60 yo	0.77	137
*Treatment 1	0.74	103
*Treatment 2	0.75	89
Tumor > 5 cm, age 50~<60 yo	0.82	58
Tumor < 2 cm, age 40~<50 yo	0.73	71
Tumor < 2 cm, age 50~<60 yo	0.72	75
Receptor and HER2 negative, no prior hormone therapy	0.77	45

Table 3. Subgroup analysis using correlation in dat200
*Treatment 1: doxorubicin and cyclophosphamide, 4 cycles. Treatment 2: doxorubicin and cyclophosphamide, 6 cycles

Discussion

First, our study found that the correlations between PFS and OS in breast cancer were classified as low according to iQWIG. All patients in dat200 received no prior chemotherapy, which showed that PFS was not an appropriate surrogate in females with breast cancer receiving chemotherapy for the first time. In dat158 which patients had metastatic breast cancer, the correlation was low as well. In sum, PFS and OS were poorly correlated in the setting of early or advanced breast cancer.

Despite that PFS's low correlation with OS, PFS was a statistically significant predictor of OS according to cox analyses by both datasets. The result showed that an extended PFS was a good prognostic factor

for longer survival. On the other hand, older ages, larger tumor size, HER2 negative, ER/PgR both negative, race of black were associated with higher hazard and thus were negative prognostic factors of survival. The result was consistent with prior research findings that tumor size remained a prognostic factor independent of other factors in patients with breast cancer. [18,19] Patients at older ages are also prone to death of all causes and have a poorer prognosis compared to younger subjects.[20]

In our preliminary subgroup analysis of dat200, higher correlations compared to the baseline of 0.7 were found in patients with certain characteristics e.g., age, chemotherapies, receptor status, tumor size and combinations of those factors. The findings indicate that in these specific patient populations, PFS can serve as a more reliable predictor of OS and thus facilitate clinical decisions and policymaking. Further analyses are required to solidify the significance of the results.

Limitations

First, our study was limited by the small number of studies that provided individual patients' data. The results may not be representative and applicable to other subtypes of breast cancer or treatments. However, using individual patients' data allowed evaluate the impact of each variable on our outcomes of interest and perform subgroup analyses. Second, we couldn't perform the collective analyses of two studies due to the heterogenicity in study design, trial medications, studied population, data collection that prevents us from analyzing them together. Third, the hazards of PFS didn't remain constant overtime, which violated the assumptions of Cox. This was expected because the definition of PFS included survival and survival decreases overtime. Also, hazards of other covariates remained constant overtime, indicating Cox is still appropriate to use for our study. Another limitation was the small number of patients in each subgroup correlation analysis that may reduce the power of the study and cause type II error. No validation analyses were performed at this point to validate the results. Further post-hoc analyses or other methods are still underway to verify the correlations in subgroups of patients.

Reference

1. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. *Biotechnol Law Rep*. 2015;26(4):375-386. doi:10.1089/blr.2007.9941
2. Chen EY, Joshi SK, Tran A, Prasad V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. *JAMA Intern Med*. 2019;179(5):642-647. doi:10.1001/jamainternmed.2018.8351
3. Prasad V, Kim C, Burotto M, Vandross A. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Intern Med*. 2015;175(8):1389-1398. doi:10.1001/jamainternmed.2015.2829
4. Pasalic D, McGinnis GJ, Fuller CD, et al. Progression-free survival is a suboptimal predictor for overall survival among metastatic solid tumour clinical trials. *Eur J Cancer*. 2020;136:176-185. doi:10.1016/j.ejca.2020.06.015
5. Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer*. 2019;106:196-211. doi:10.1016/j.ejca.2018.11.012
6. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. 2014;30(3):312-324. doi:10.1017/S0266462314000300
7. Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther*. 2012;5:287-296. doi:10.2147/OTT.S36683

8. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure | FDA. Accessed January 18, 2021. https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure
9. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. 2014;30(3):312-324. doi:10.1017/S0266462314000300
10. Sherrill B, Amonkar M, Wu Y, et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. *Br J Cancer*. 2008;99(10):1572-1578. doi:10.1038/sj.bjc.6604759
11. Miksad RA, Zietemann V, Gothe R, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. *Int J Technol Assess Health Care*. 2008;24(4):371-383. doi:10.1017/S0266462308080495
12. Hirai T, Nemoto A, Ito Y, Matsuura M. Meta-analyses on progression-free survival as a surrogate endpoint for overall survival in triple-negative breast cancer. *Breast Cancer Res Treat*. 2020;181(1):189-198. doi:10.1007/s10549-020-05615-4
13. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol*. 2005;5:14. doi:10.1186/1471-2288-5-14
14. BMJ (OPEN ACCESS) Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160. doi: 10.1136/bmj.n160
15. IQWiG. Validity of surrogate endpoints in oncology executive summary of rapid report A10-05, Version 1.1. Cologne, Germany: Institute for Quality and Efficiency in Health Care: Executive Summaries; 2005.
16. Shulman LN, Berry DA, Cirrincione CT, et al. Comparison of doxorubicin and cyclophosphamide versus single-agent paclitaxel as adjuvant therapy for breast cancer

Addendum

Dissemination Plan

There are three key audience of this research, these are:

- The RASP Oversight Committee
- Pharmacy professional conference
- Academia

Challenge Encountered

The analysis and validation method for the second aim, determine which types of patients benefit the most from using PFS as a predictor for OS, has been challenged. We first proposed fixed mixed effect model, but it didn't provide the answers to the question. We then reached out to the UNC Statistical / Data Science Consulting Center for advice. A few potential methods were discussed. However, neither could offer the information we were aiming for. Ultimately, we decided to perform correlation analysis in patients grouped by 1-2 variables. We were acknowledged that the sample size could be small and the incidence of type II error may increase.