We tried to use another indicator: △area under curve (AUC) to measure the goodness-of-fit (A method supported by PMID: 31511177). However, we found that AUC was not suitable for our study. Please find the following evidence:

In NIV-PFS group, calculated results of AUC between fitted models and the original KM values of 6.5-year data are presented in the following table:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| exp | weibull | gamma | lnorm | gompertz | llogis | gengamma | FP1 | FP2 | RCS | RP-hazard | RP-odds | RP-normal | GAM | param-mix | mix-cure |
| -6.83 | 396.4 | 583.73 | -102.2 | -100.86 | -268.27 | -101.03 | 261.98 | 97.52 | -26.6 | -193.86 | -199.77 | -208.11 | -26.38 | -105.05 | -177.7 |

It can be found that exponential model has the smallest absolute AUC, which indicated that exponential model might have the best goodness-of-fit. However, after we checked the survival plot (as shown in the following figure), we found that exponential model was far from a model with the best goodness-of-fit. One possible reason is that the positive and negative differences in AUC offset finally.

