### Goodness-of-fit

1. **Goodness-of-fit**

The goodness-of-fit of the model prediction to the observed individual patient data (IPD) was measured by computing the posterior mean residual deviance, $Dbar$ [33]. The deviance information criterion ($DIC$) was used to compare different models and provided a measure of model fit that penalized model complexity according to Spiegelhalter et al. (2002) [29]:

$$ \left\{\begin{array}{c}DIC=Dbar+pD\\ pD=Dbar-Dhat\end{array} \right.$$

$pD$ is the effective number of parameters and $Dhat$ is the deviance evaluated at the posterior mean of the model parameters. The model with the lowest $DIC$ provided the best data fit. The model fits were visually inspected against original published KM curves.

Model fit statistics are presented in Table 3 and Table 4, for OS and PFS respectively. In both cases, the log-normal fixed-effects model provided the best model fit with the lowest DIC.