# Supplemental materials for “Patient preferences for frontline therapies for Ph-positive acute lymphoblastic leukemia: a discrete choice experiment”

## Model selection

The initial specification of the MNL model was based on a system of dummy-coded variables (0/1) to measure participants’ preferences for discrete changes in the treatment attributes. The utility that a respondent (n) received from a treatment (j) in a given choice task (t) was defined as:

Where ‘CV’ corresponds to ‘major Cardiovascular events,’ ‘MYELO’ corresponds to the ‘suppressed immune system,’ ‘OS’ corresponds to ‘overall survival’ and ‘DOR’ corresponds to the ‘duration of remission.’ The numbers in brackets indicate the level of the attribute that each parameter relates to (OS and DOR in months).

Each parameter measures the value participants perceive for changes between levels of one attribute. For example, measures the value of change from the reference level 50% risk of a major CV event to 25% risk of a major CV event. The model specification also included an interaction term (measured with ) between OS and DOR.

Each choice within the discrete choice experiment (DCE) was followed by the question: “*If these treatments had been offered to you by your doctor when you were diagnosed, would you have: a) Taken the treatment that you chose above? b) Taken neither of these treatments?*”, to check participants would be actually willing to use the treatment chosen. To check whether initial and follow-up choices should be combined for the estimation of preferences, two dummy-coded multinomial logit (MNL) models were estimated, one based on the initial choices only and the other based on the combined (initial + follow-up) choices. Predictive validity of each model was calculated using a 10-fold cross-validation procedure. The average predictive validity for the MNL model based on initial choices was 66.0% vs. 55.8% for the model based on combined choices. Therefore, initial and follow-up choices were not combined.

Next, two MNL models were estimated to determine whether an interaction effect should be included between overall survival (OS) and duration of survival (DoR) since participants’ valuation of OS may be influenced by the length of time in remission. The model including the interaction effect did not significantly outperform the model without the interaction effect in a log-likelihood ratio (LR) test (Deviance = 0.2; *P* = 0.655). Therefore, the model without the interaction effect was selected to decrease the risk of overfitting the choice data.

A heteroscedastic MNL model was estimated to determine whether choice data from the different versions of the DCE survey could be combined. This model did not outperform the MNL model in an LR test (Deviance = 2.8; *P* = 0.094). Therefore, it was decided to pool the data together.

We next checked whether the specification of the MNL model could be further refined by treating the attributes as continuous rather than discrete variables. A linear regression was fitted through the dummy-coded estimates of the benchmark MNL model and the coefficient of determination (R2) was used to determine whether the data supported the assumption of linearity. The R2 of each of the four attributes was greater than the accepted 0.7 threshold to assume linearity in preferences (OS = 0.82; DoR = 0.83; cardiovascular risk = 0.88; myelosuppression risk = 0.99), therefore the specification of the MNL model was simplified by using linear coding of the attributes’ levels, providing a new benchmark MNL model.

Finally, we checked whether it was necessary to account for the panel nature of the choice data (i.e., same respondent providing multiple observations/choices). For this, an error component MNL model was estimated and its performance tested against the new linear benchmark MNL model. An LR test showed no significant difference between the two models (Deviance = 1.99; DF = 1; *P* = 0.158).

Therefore, the final model used to analyze the preference data incorporated initial choices only, no interaction effect between OS and DoR, combined DCE survey variants, linear effects, and no individual-level error component:

Where the model parameters are:

* () measures the influence of a 1% decrease in risk of major cardiovascular events
* ) measures the influence of a 1% decrease in risk of myelosuppression
* () measures the influence of a 1-month increase in OS
* ) measures the influence of a 1-month increase in duration of remission

The linear coding of attributes was needed to derive meaningful maximum acceptable risk and maximum acceptable benefit measures. When the utility function is linear and additive, marginal rates of substitution are computed as a ratio of estimated preferences:

This specification has the lowest BIC (3082.7), indicating the best statistical fit, out of the eight specifications tested (**Table S1**).

**Table S1. Comparison of statistical performance across models**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Choices** | **Coding** | **Interaction effect** | **Observations** | **Parameters** | **Log-likelihood** | **BIC** | **PV (%)** |
| Initial | Dummy | Yes | 2412 | 30 | -1490.2 | 3214.1 | 66 |
| Initial | Dummy | No | 2412 | 29 | -1490.3 | 3206.4 | 65.8 |
| Combined | Dummy | Yes | 2412 | 31 | -2287 | 4815.4 | 55.8 |
| Combined | Dummy | No | 2412 | 30 | -2287 | 4807.7 | 55.8 |
| Initial | Linear | Yes | 2412 | 6 | -1521.1 | 3089 | 64.8 |
| ***Initial*** | ***Linear*** | ***No*** | ***2412*** | ***5*** | ***-1521.9*** | ***3082.7*** | ***64.7*** |
| Combined | Linear | Yes | 2412 | 7 | -2327.7 | 4709.9 | 54.7 |
| Combined | Linear | No | 2412 | 6 | -2328 | 4702.7 | 54.3 |

The specification with the lowest BIC was used as the reference model (italicized).   
Abbreviations: BIC, Bayesian information criteria; PV, Predictive validity