

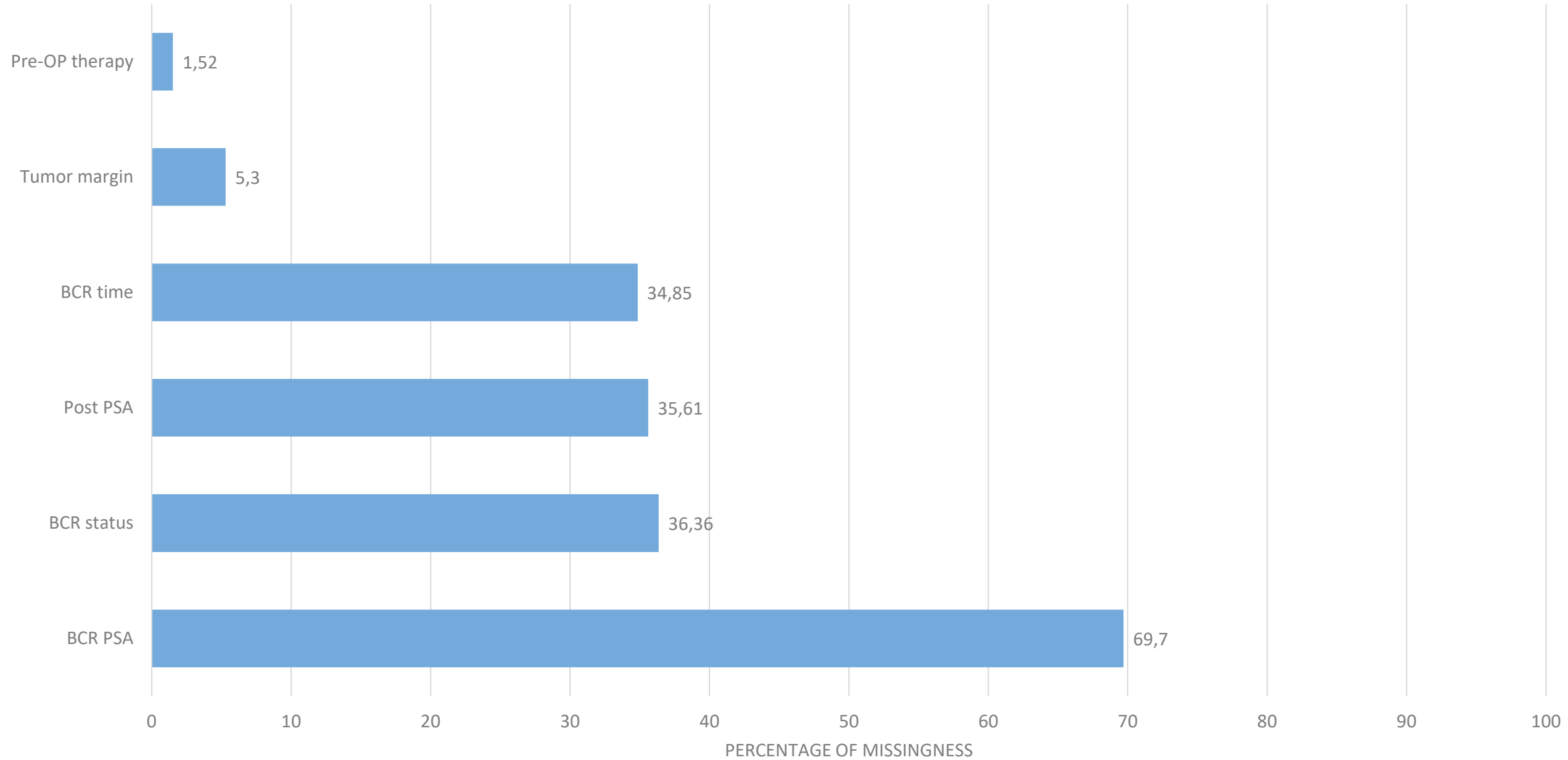
Data distribution and error
scores on imputation methods

Summary

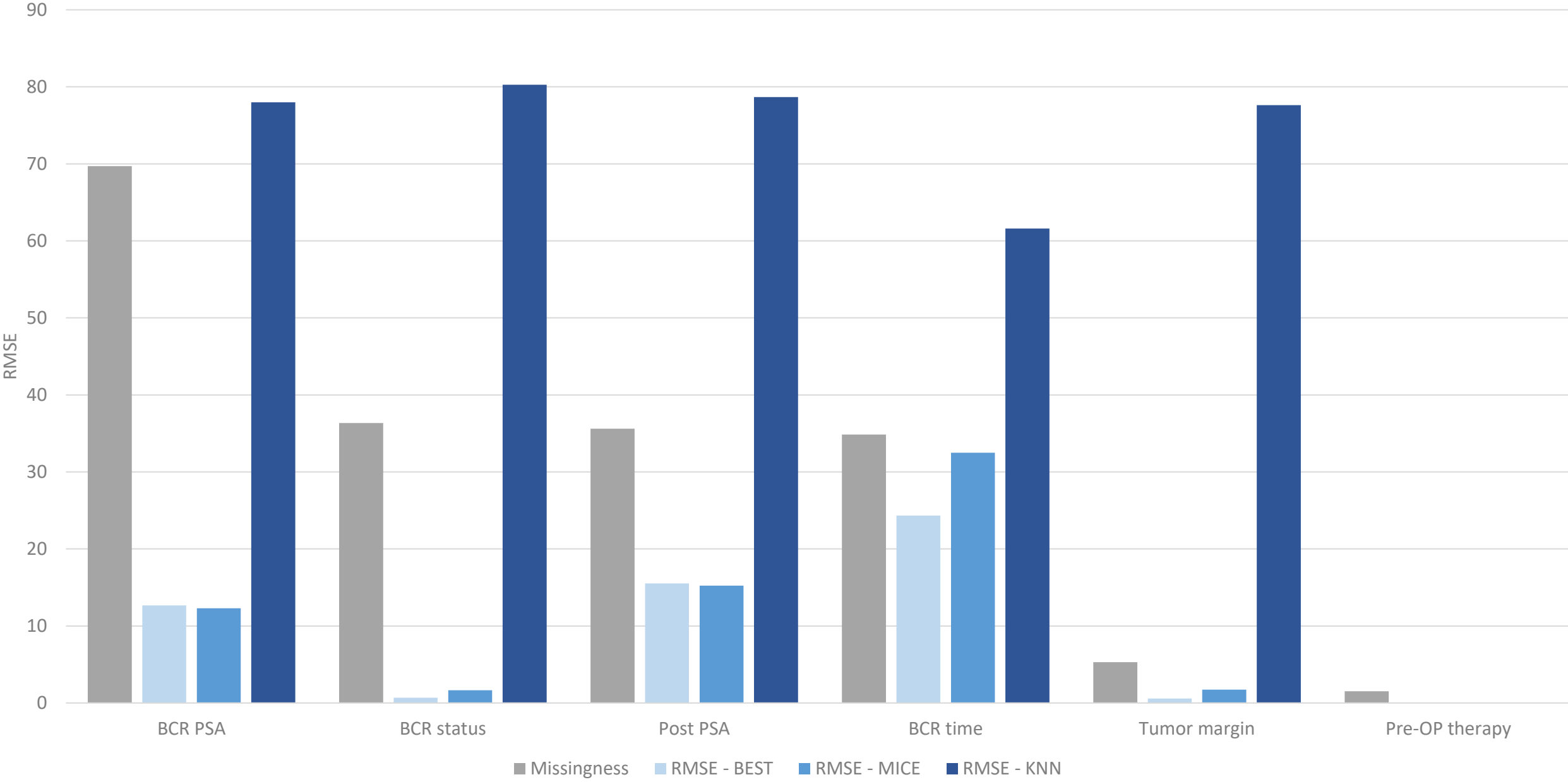
Findings on our assessed prostate cancer data

- Data **distributions** before/after imputing missingness are preserved for most of the cases
- Data with similar **types** (binary, discrete, continuous) behave similarly towards different imputation methods
- The **error scores** of univariate methods are often comparable with multivariate methods

MISSINGNESS IN DATA

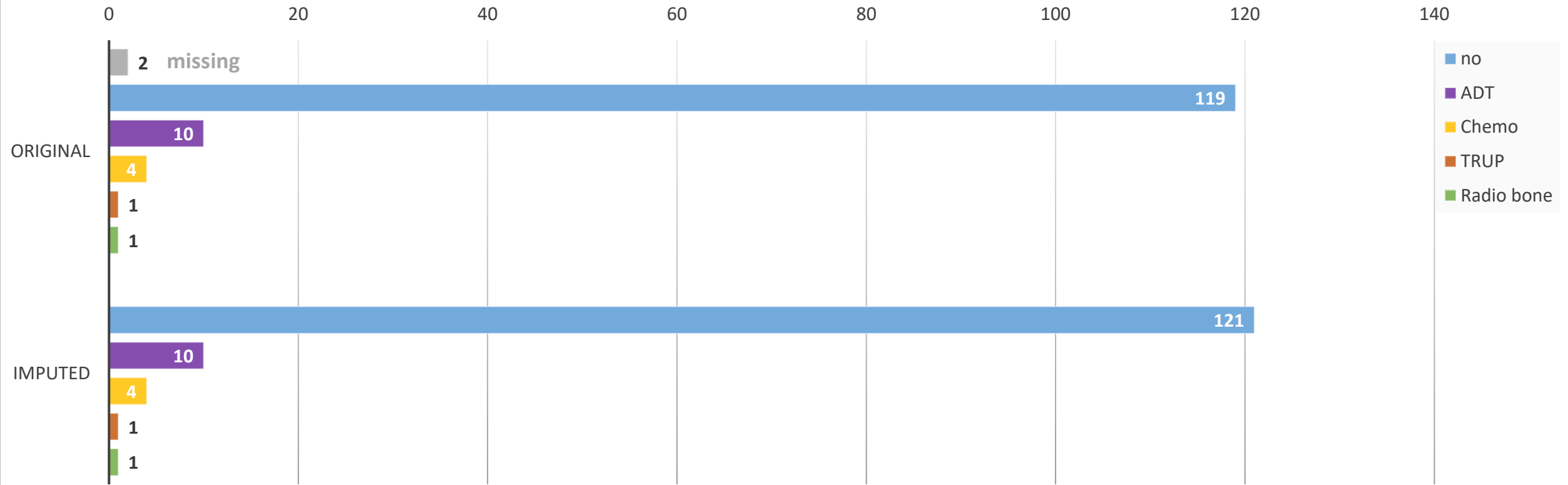


MISSINGNESS AND RMSE PER FEATURE (COMPLETE SET)

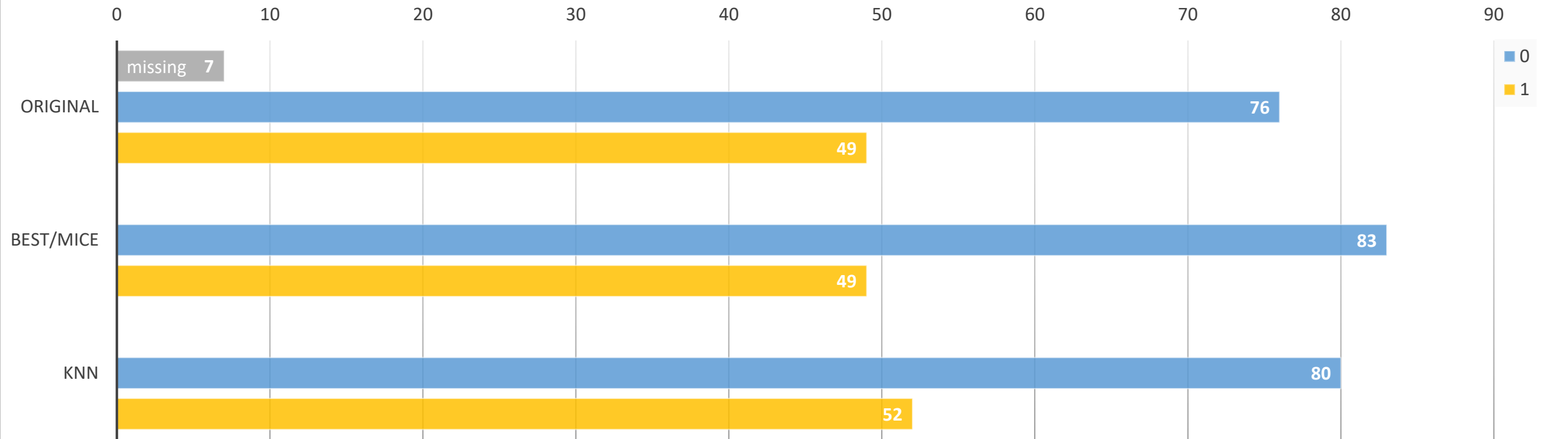


Data Distribution Before/After Imputation

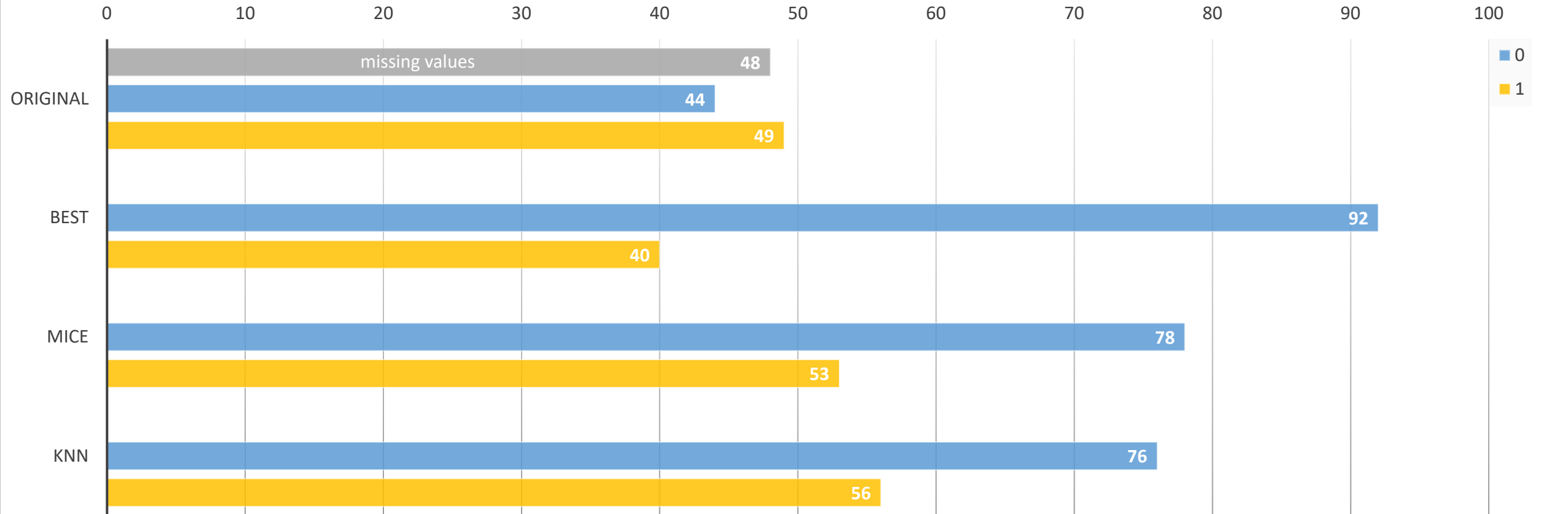
PRE-OP THERAPY – NUMBER OF VALUES PER CATEGORY



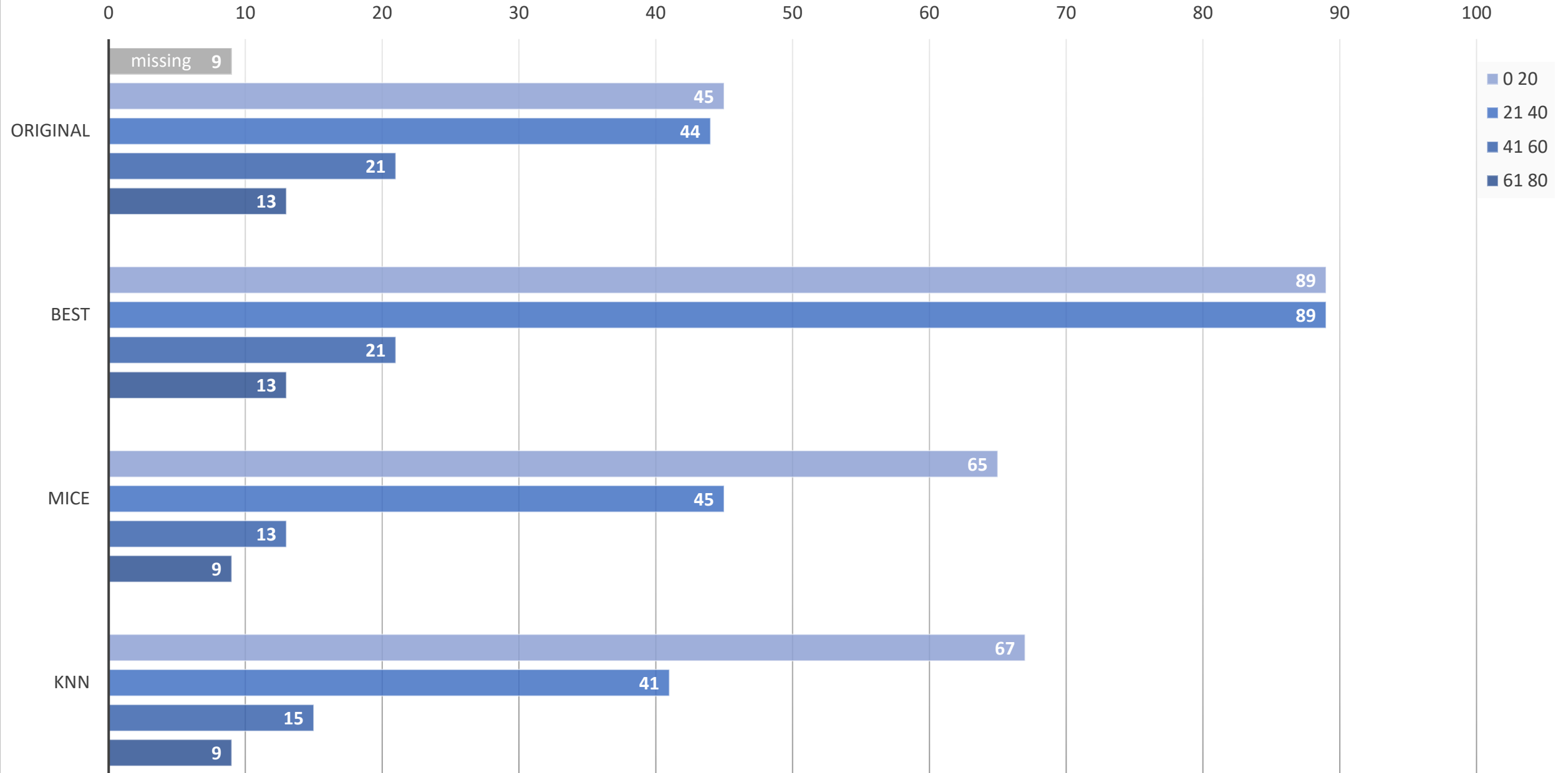
TUMOR MARGIN – NUMBER OF VALUES PER CATEGORY



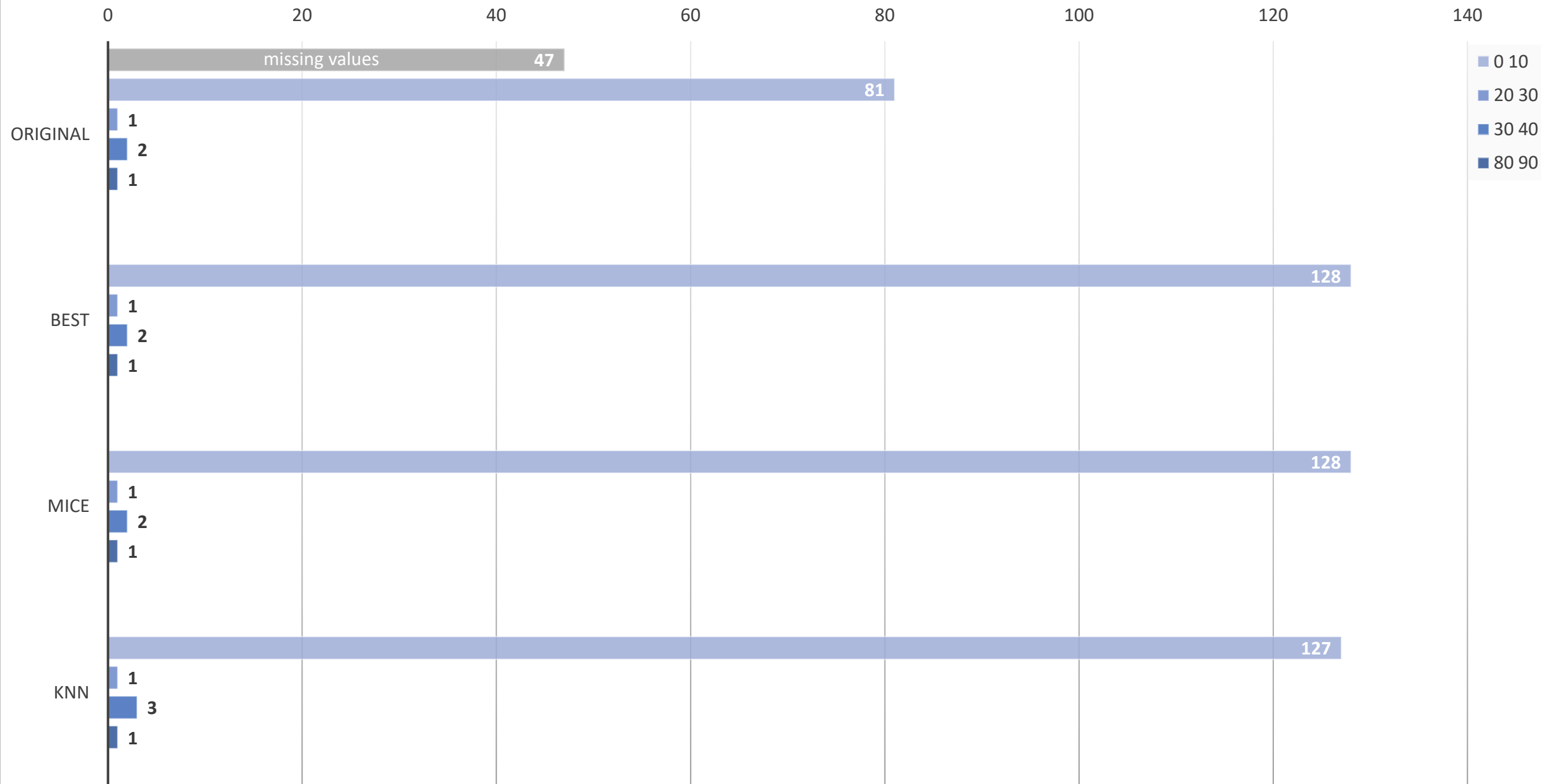
BCR STATUS – NUMBER OF VALUES PER CATEGORY



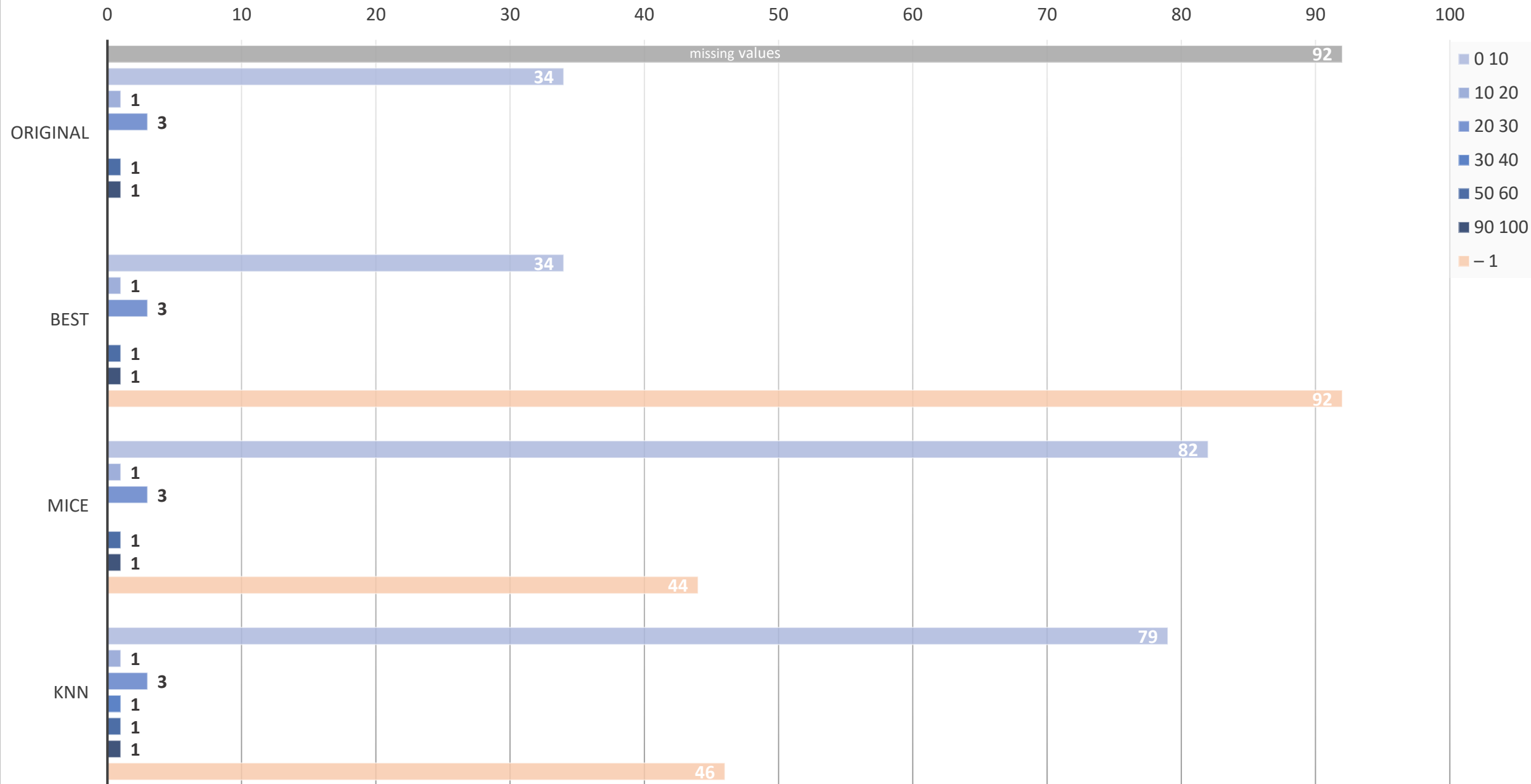
BCR TIME – NUMBER OF VALUES PER INTERVAL



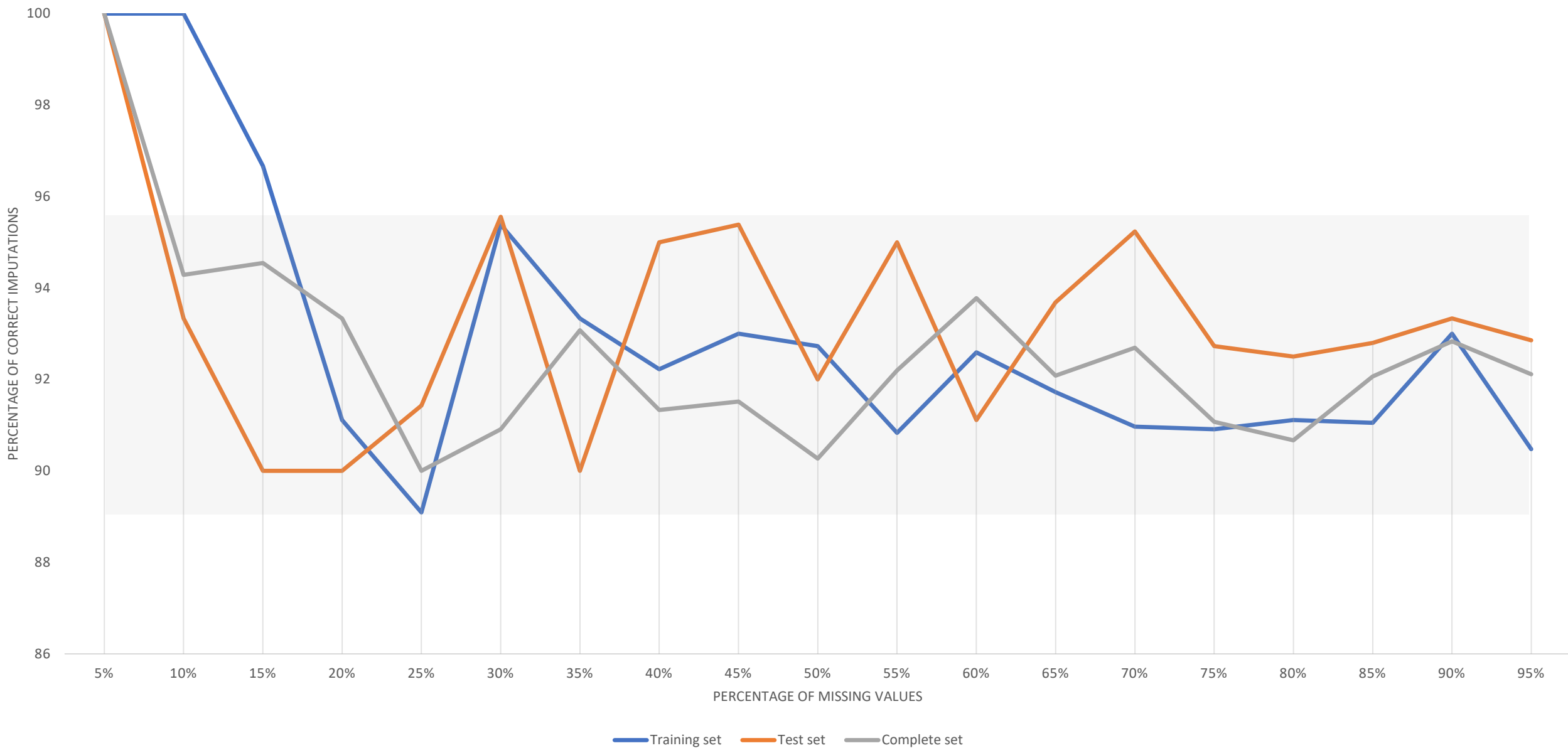
POST PSA – NUMBER OF VALUES PER INTERVAL



BCR PSA – NUMBER OF VALUES PER INTERVAL

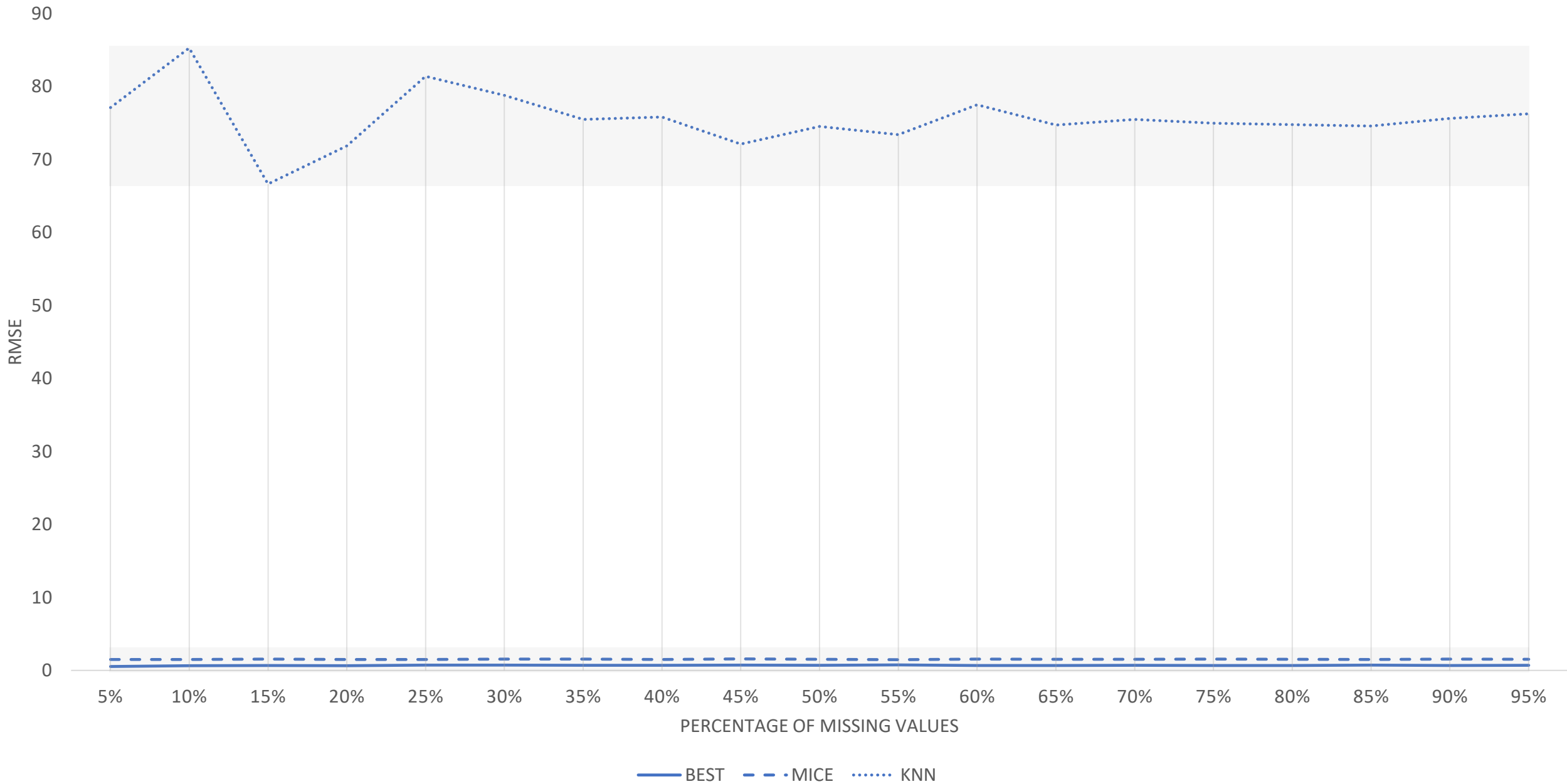


PRE-OP THERAPY – **CORRECT** IMPUTATIONS IN PERCENT (MOST FREQUENT)

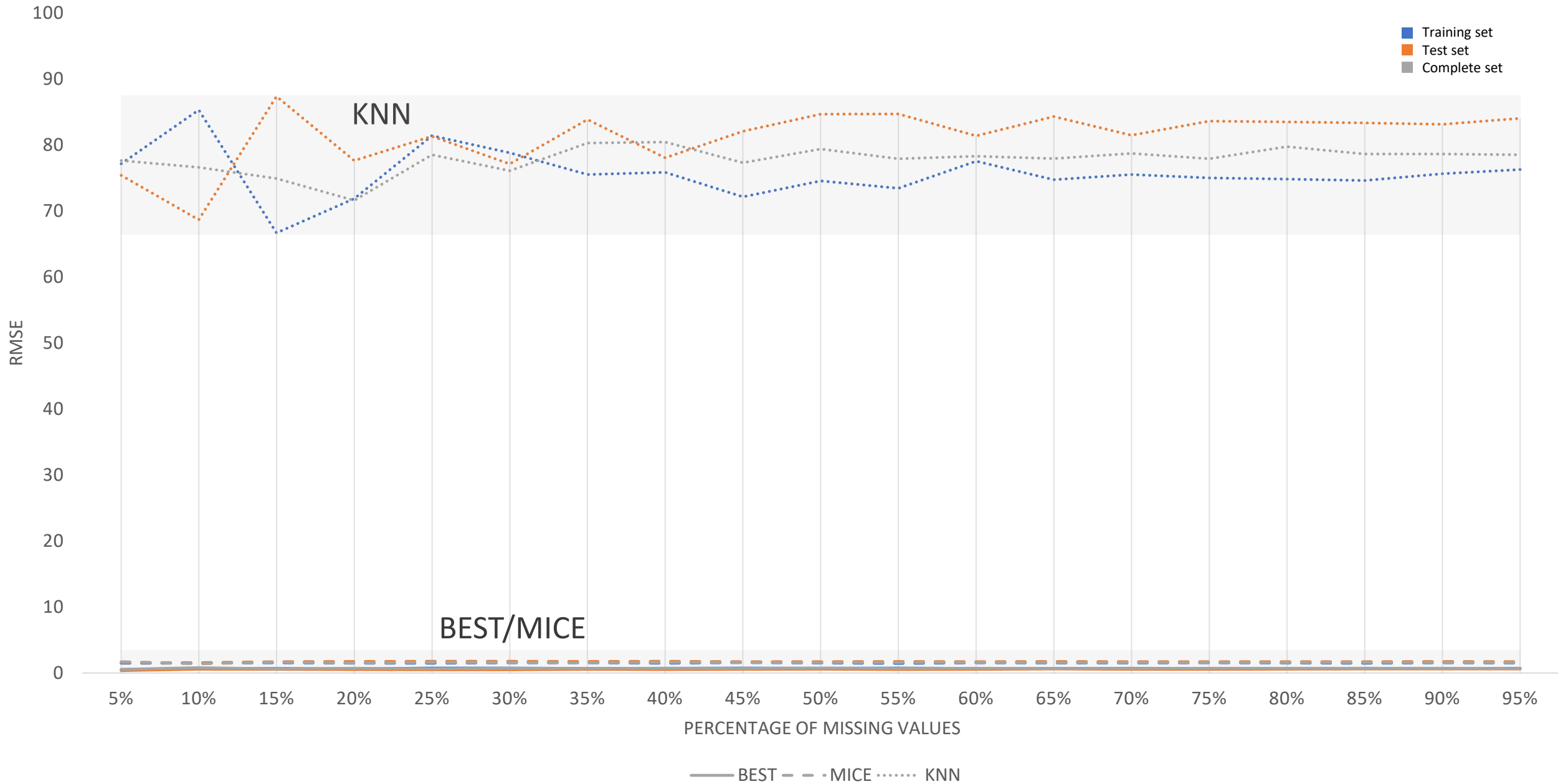


Root Mean Square Error (RMSE) of
imputation methods for simulated
missingness in the data

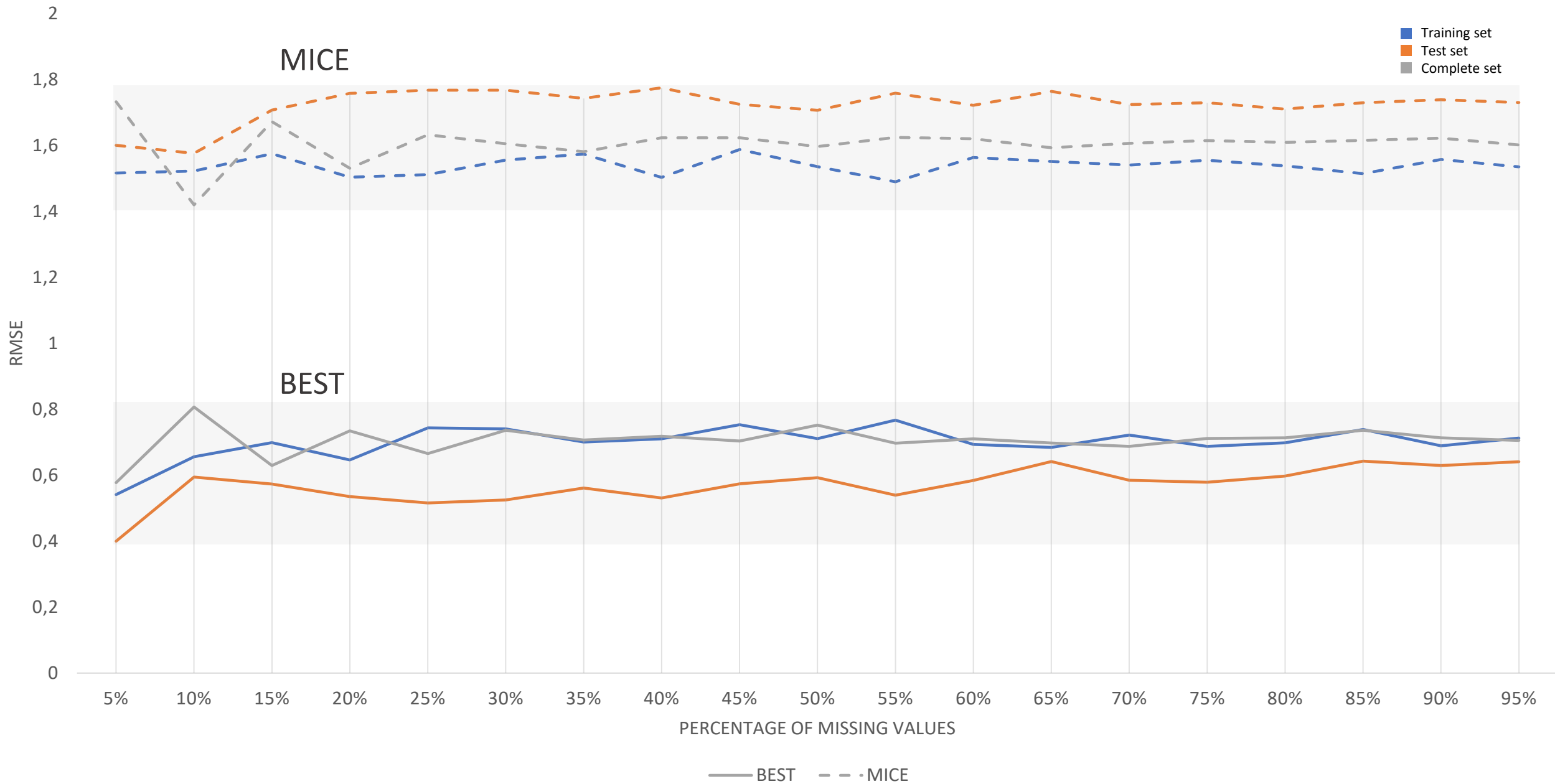
RMSE – TUMOR MARGIN (TRAINING SET)



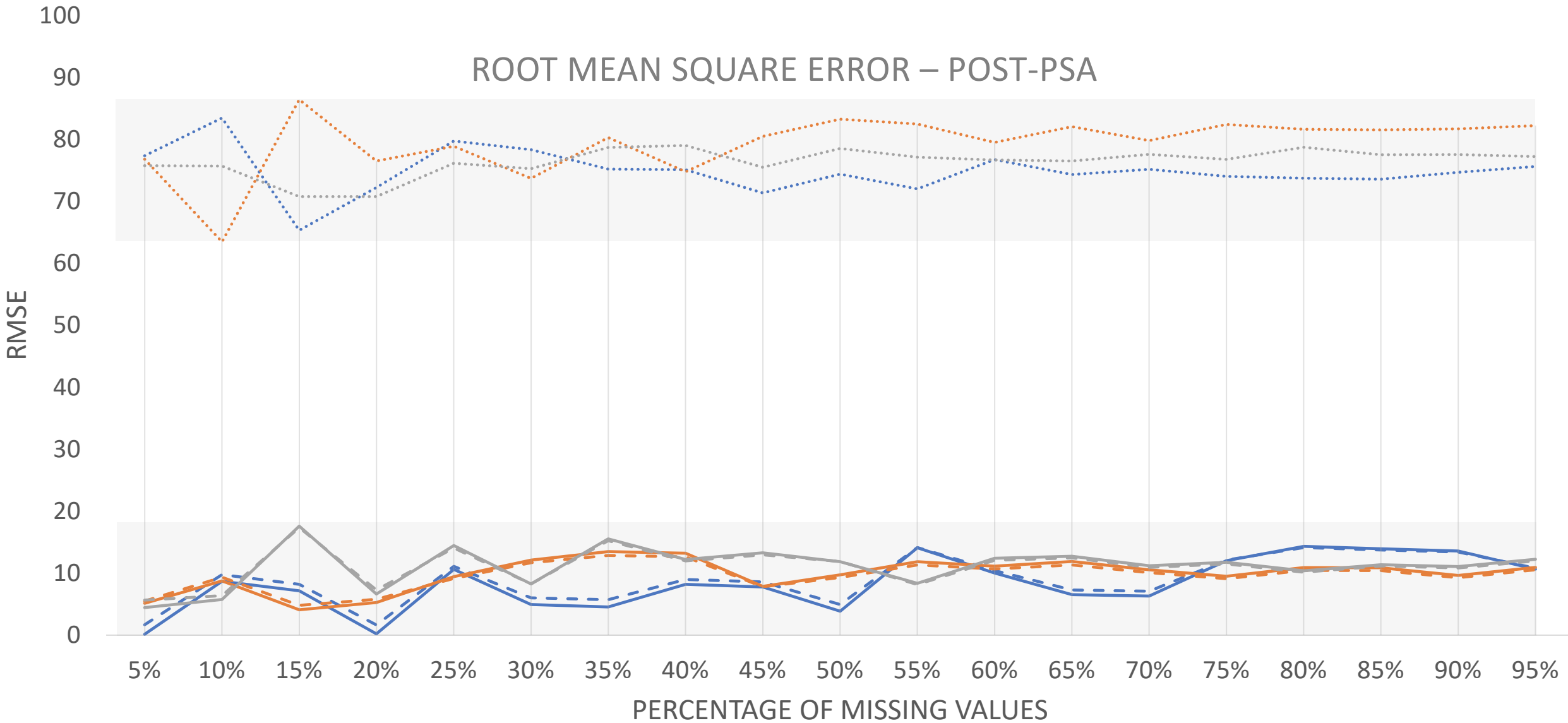
RMSE – TUMOR MARGIN



RMSE – TUMOR MARGIN

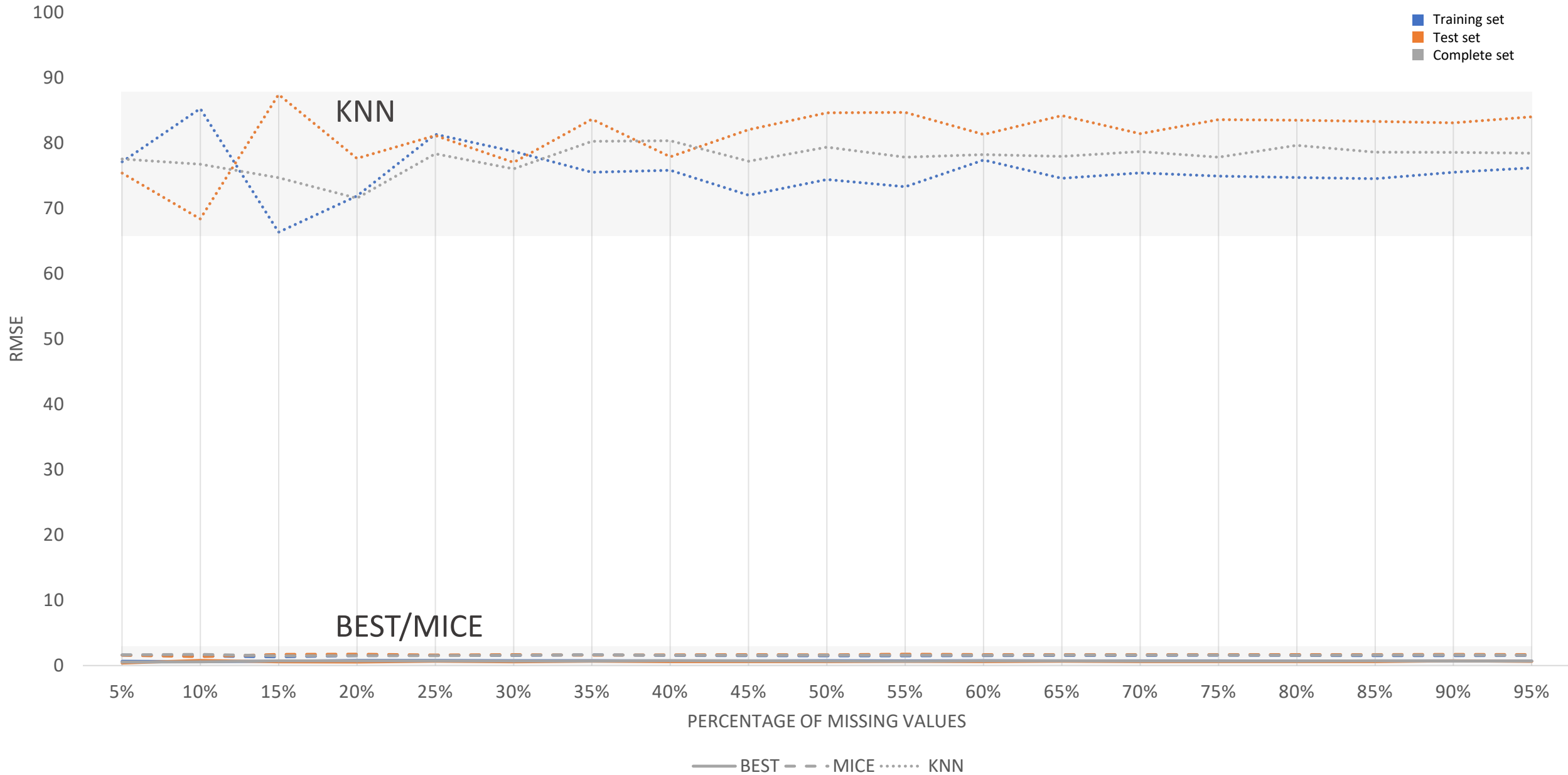


ROOT MEAN SQUARE ERROR – POST-PSA

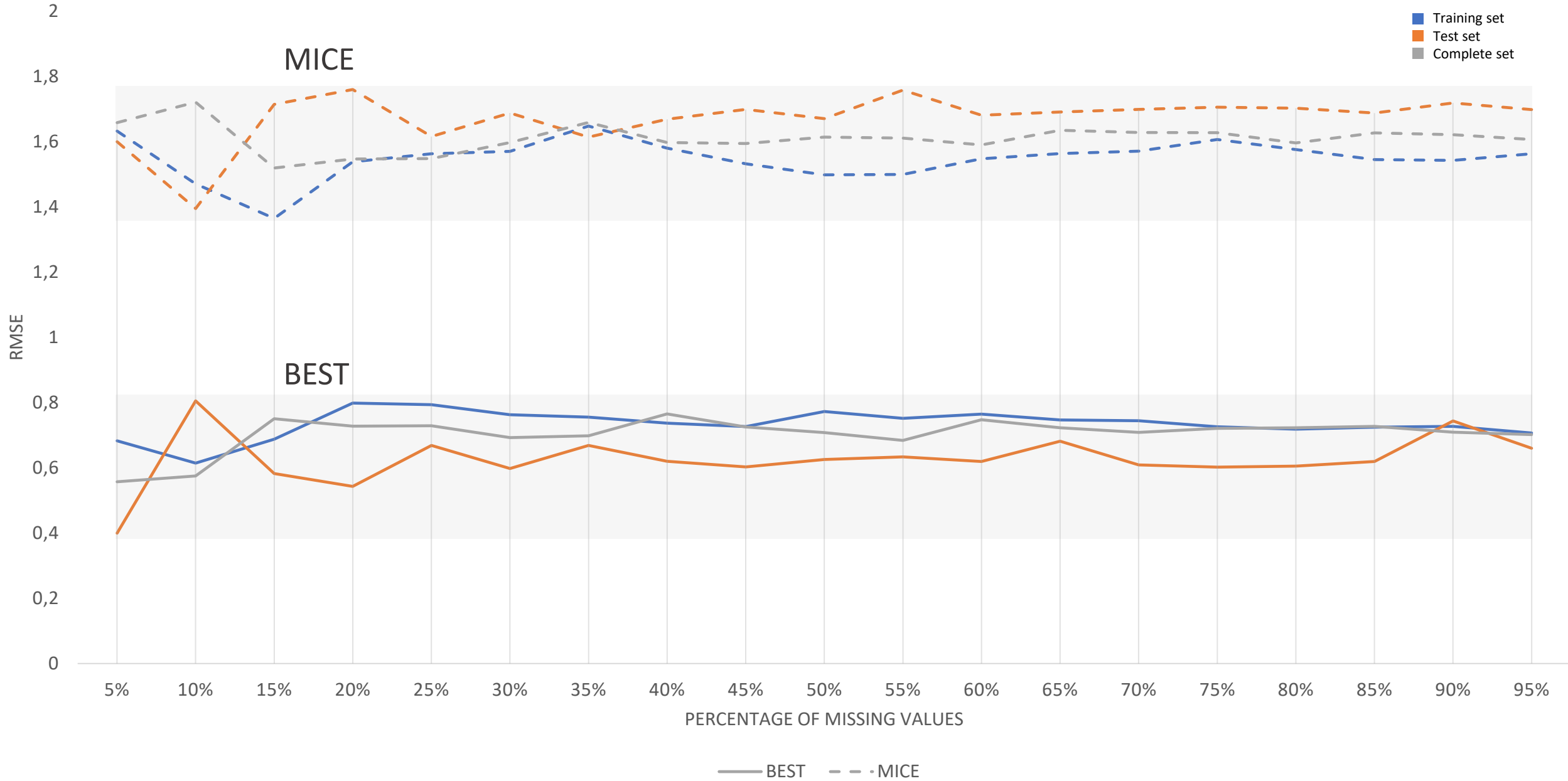


— BEST - - - MICE KNN ■ Training set ■ Test set ■ Complete set

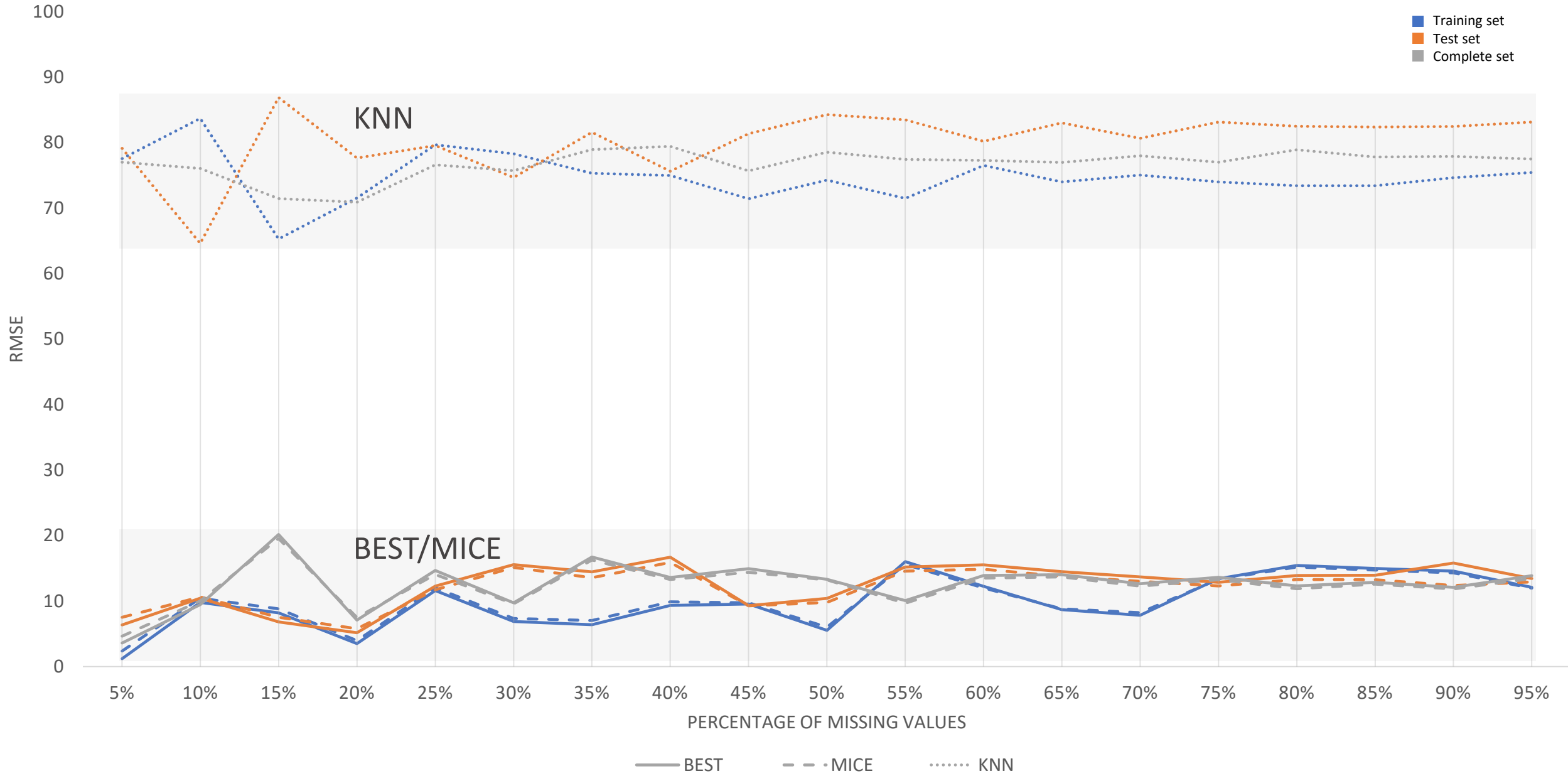
RMSE – BCR STATUS



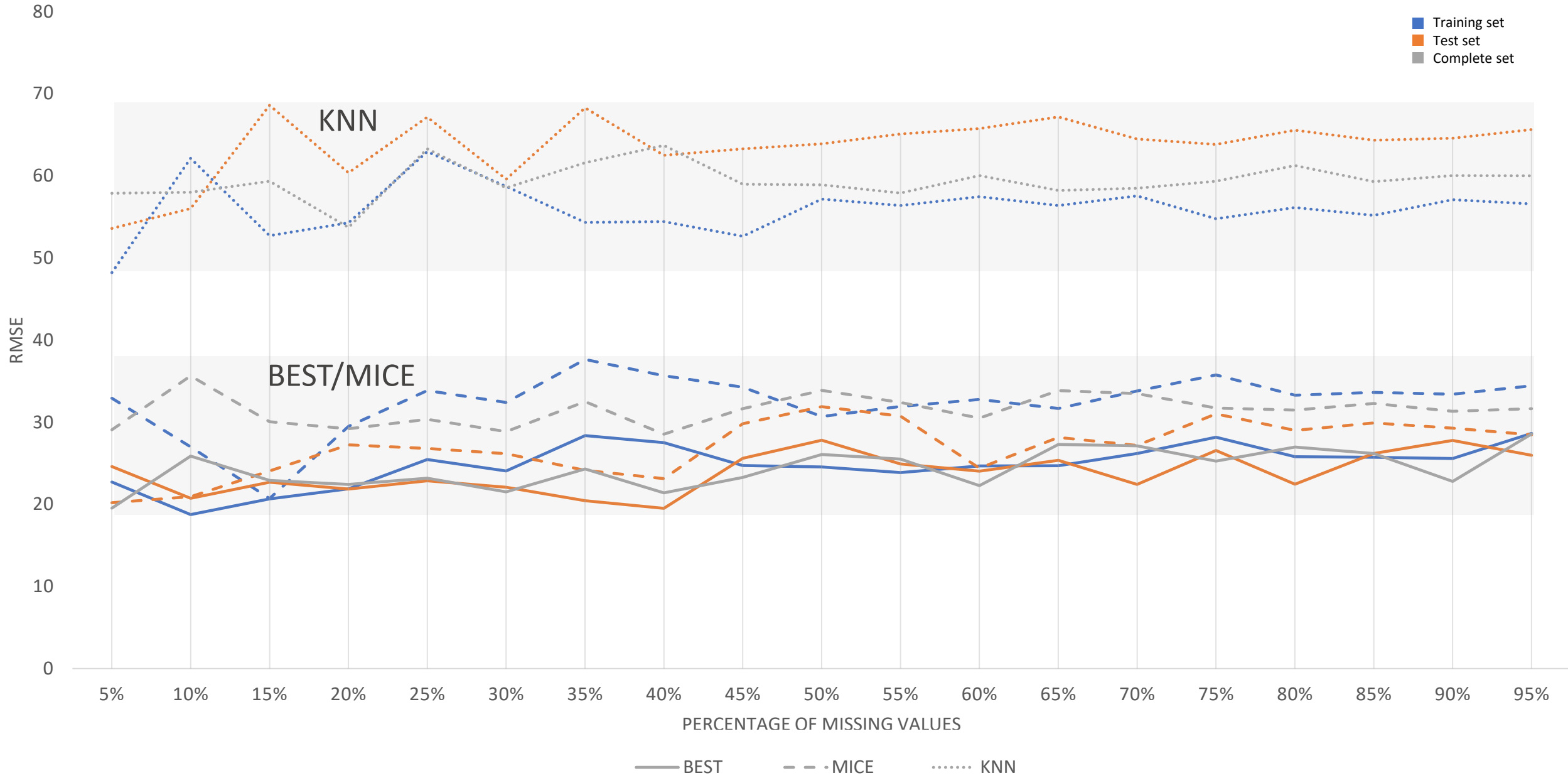
RMSE – BCR STATUS



RMSE – BCR PSA

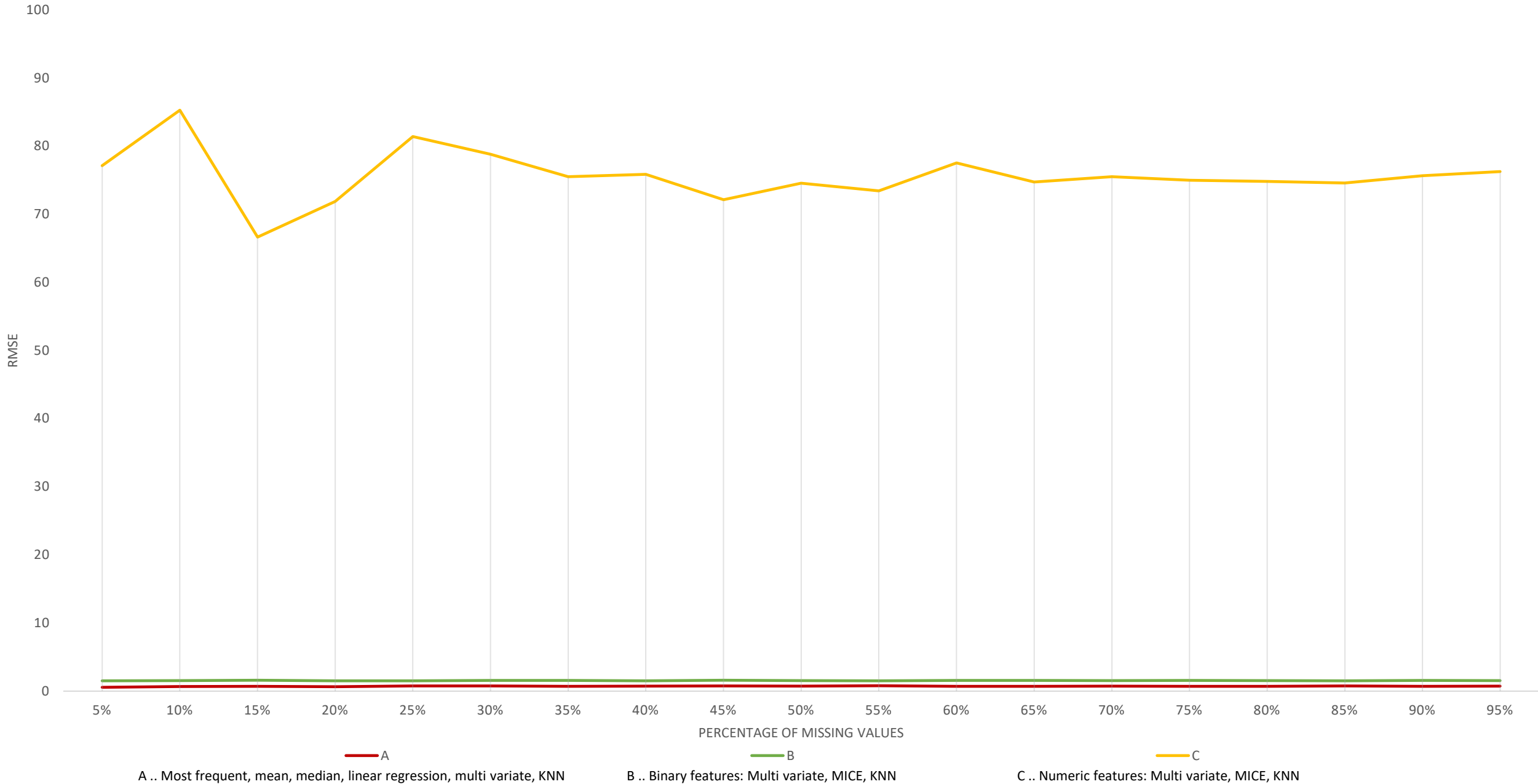


RMSE – BCR TIME

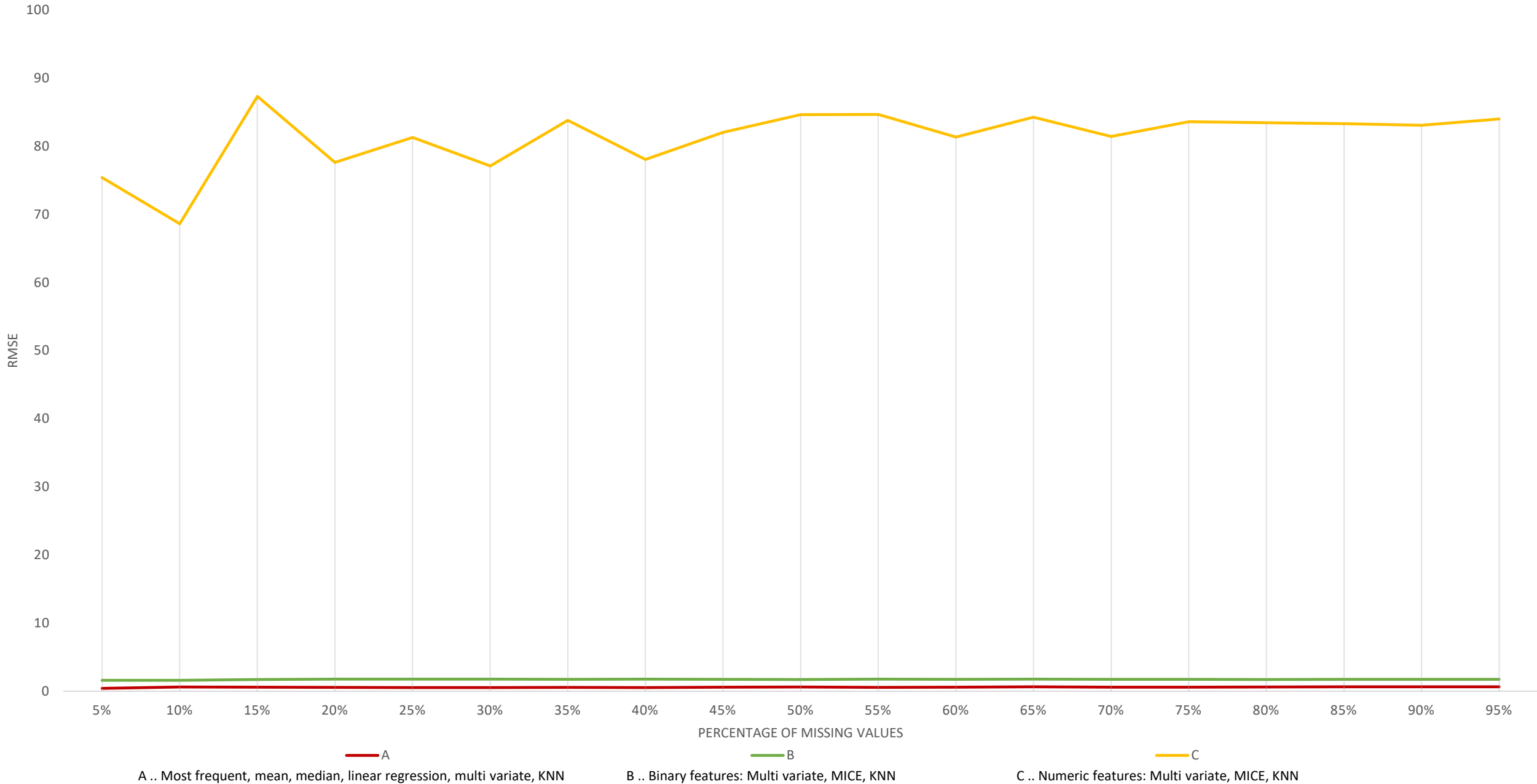


Binary Features

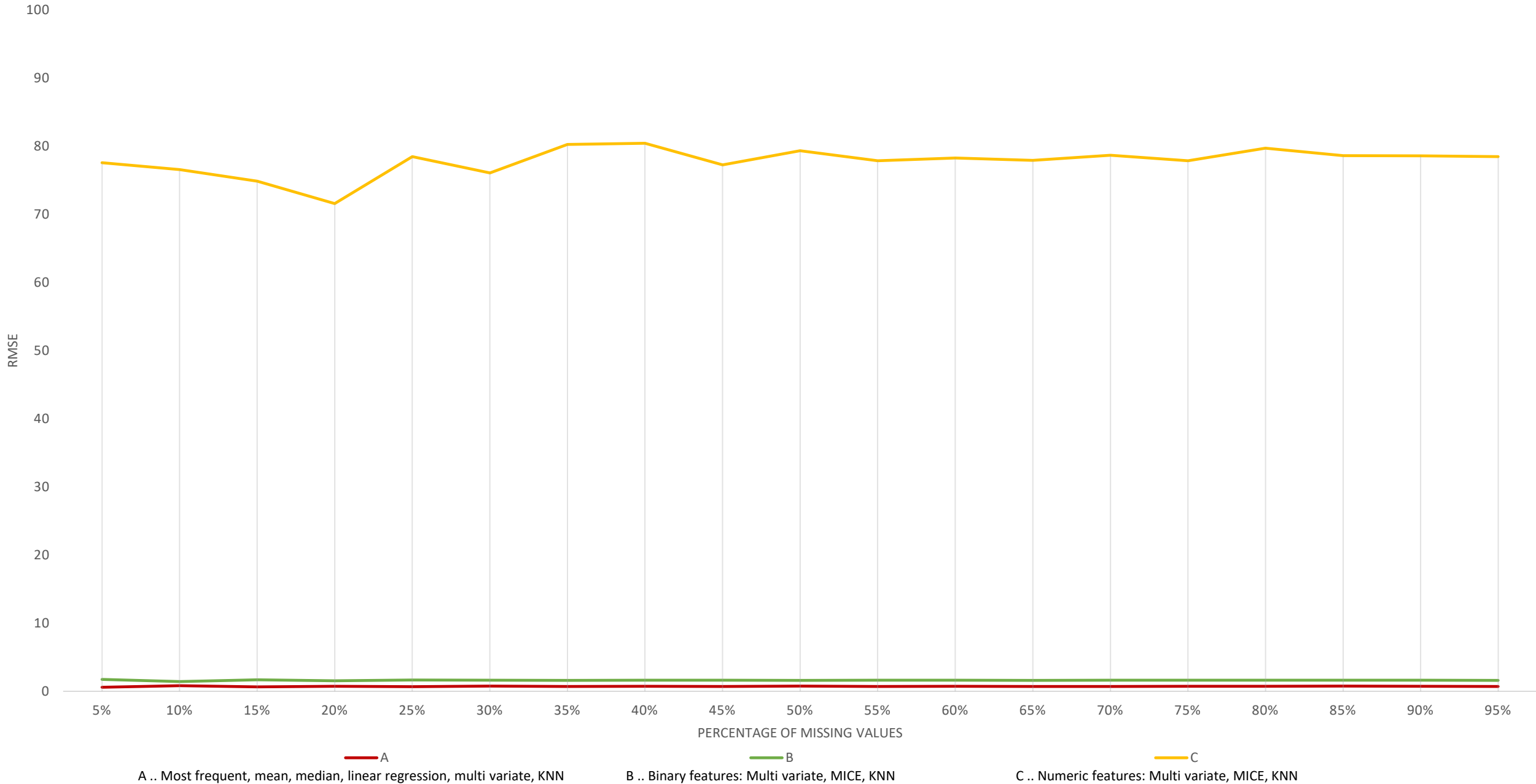
TUMOR MARGIN (TRAINING SET)



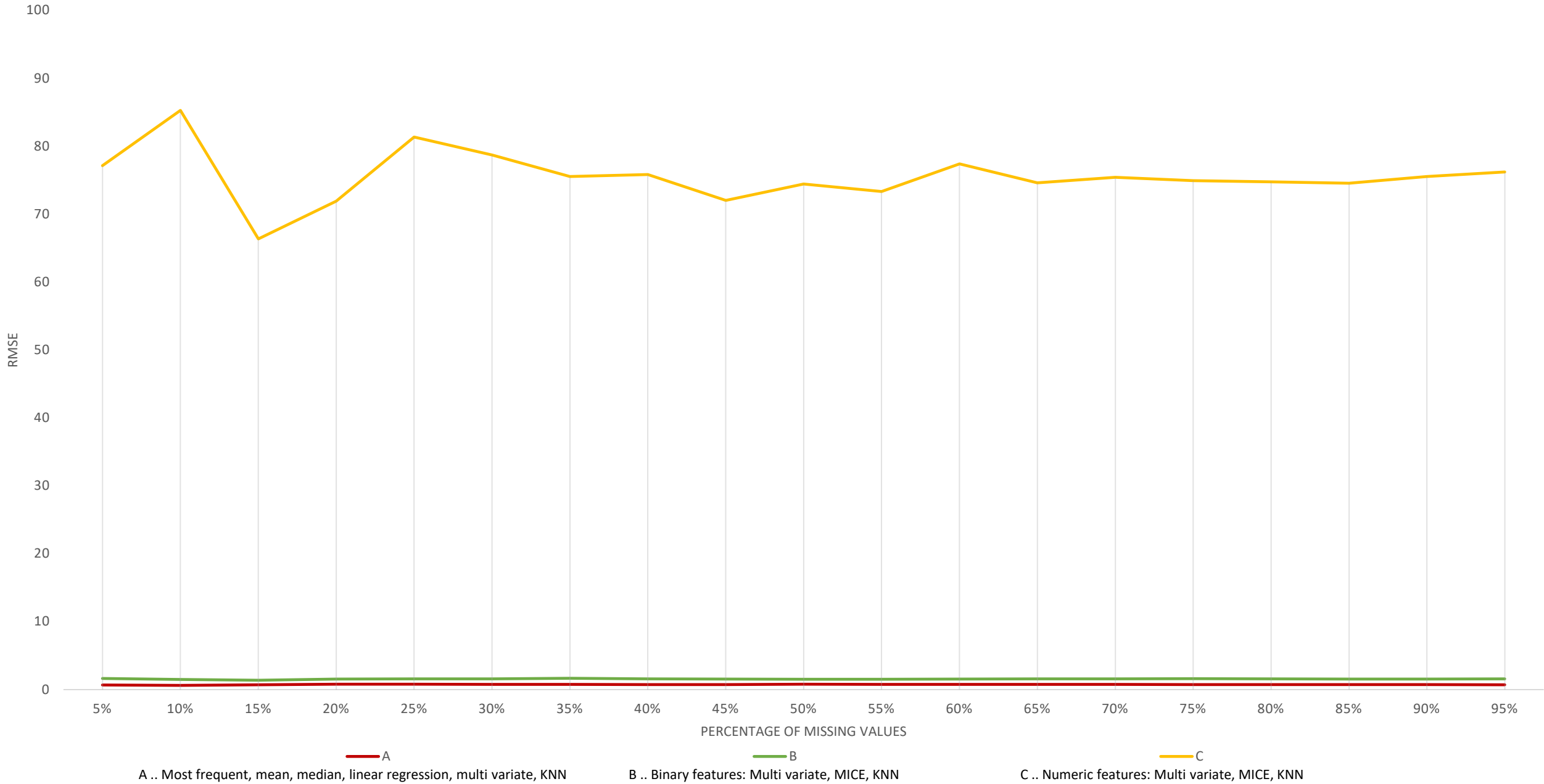
TUMOR MARGIN (TEST SET)



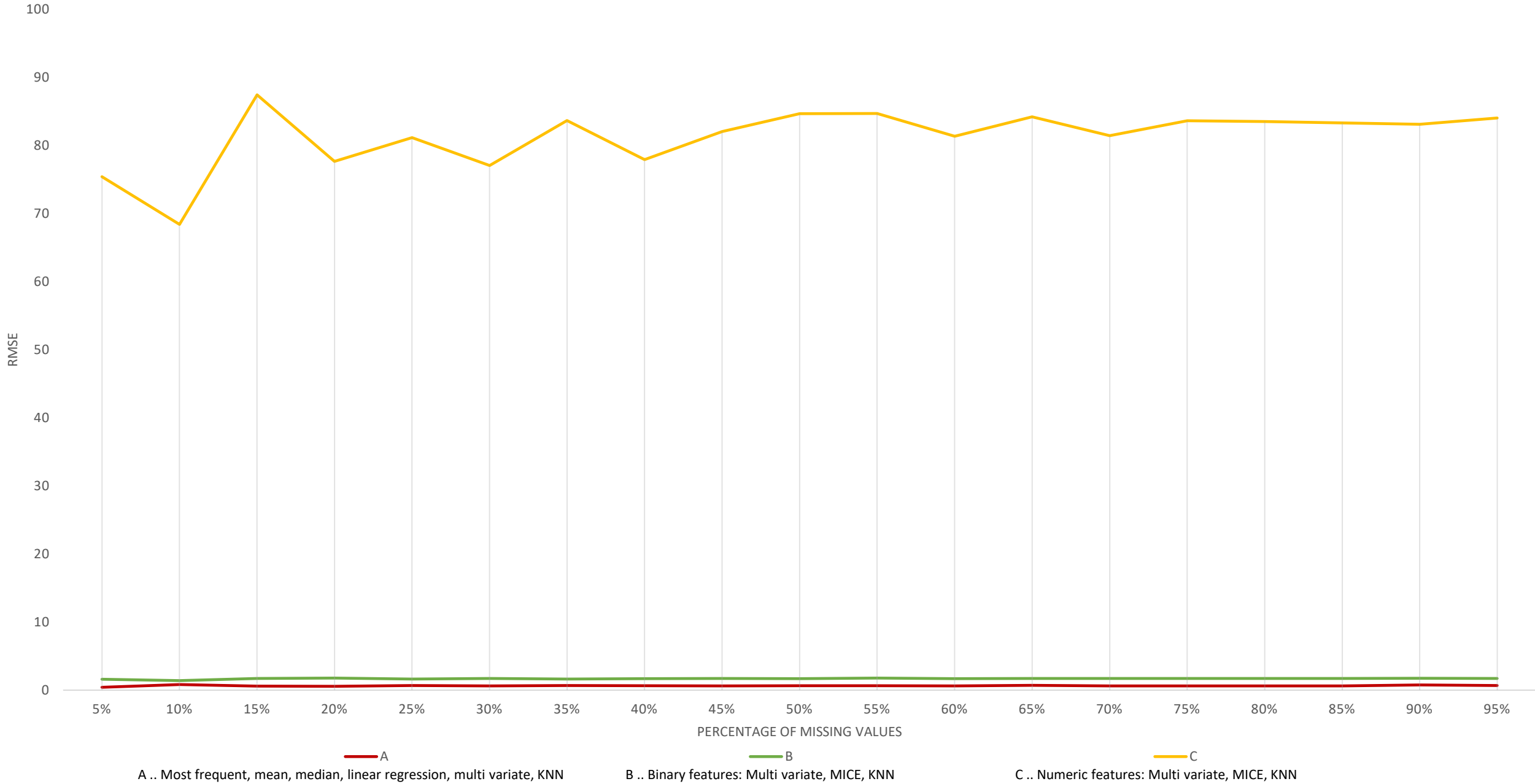
TUMOR MARGIN (COMPLETE SET)



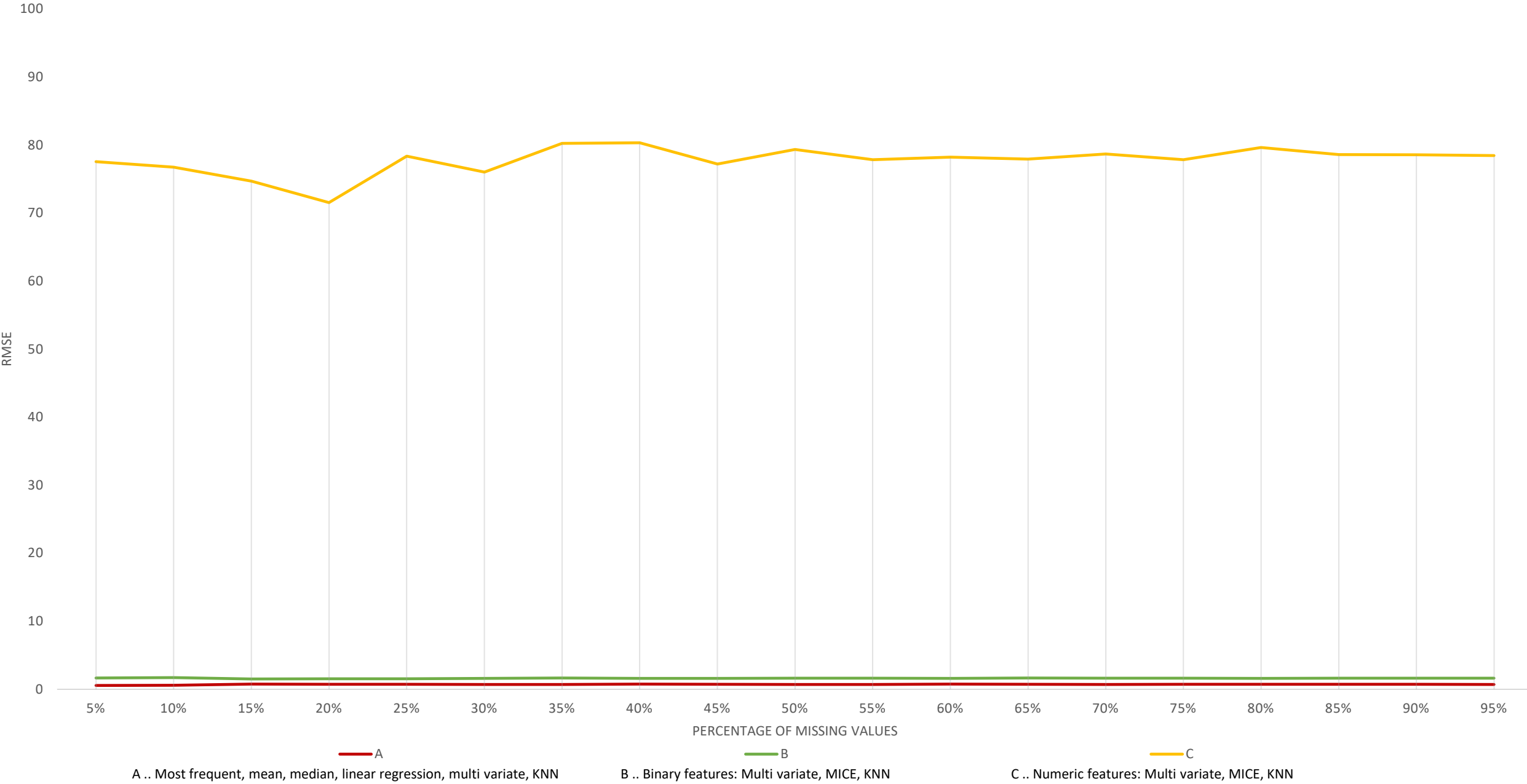
BCR STATUS (TRAINING SET)



BCR STATUS (TEST SET)



BCR STATUS (COMPLETE SET)



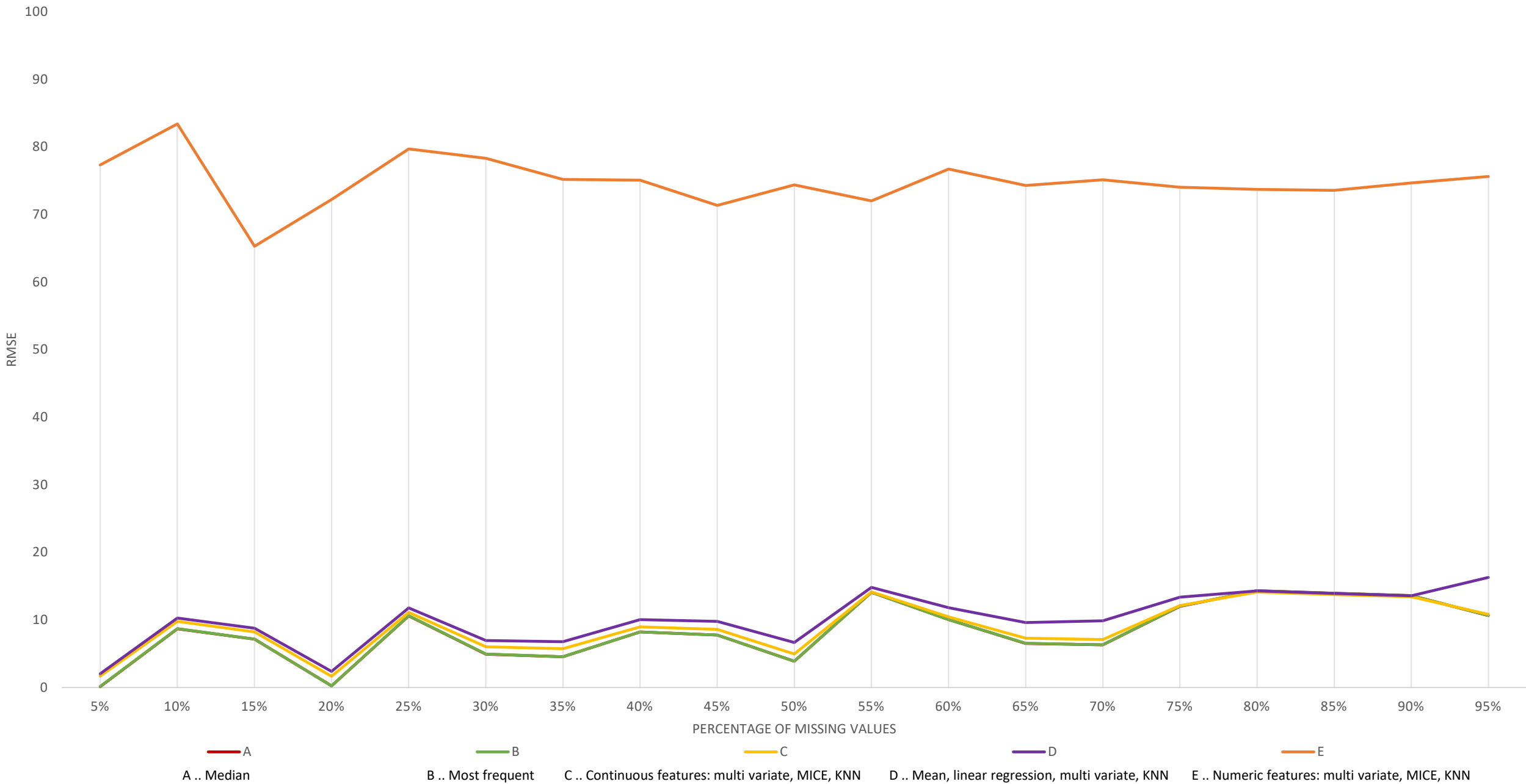
A .. Most frequent, mean, median, linear regression, multi variate, KNN

B .. Binary features: Multi variate, MICE, KNN

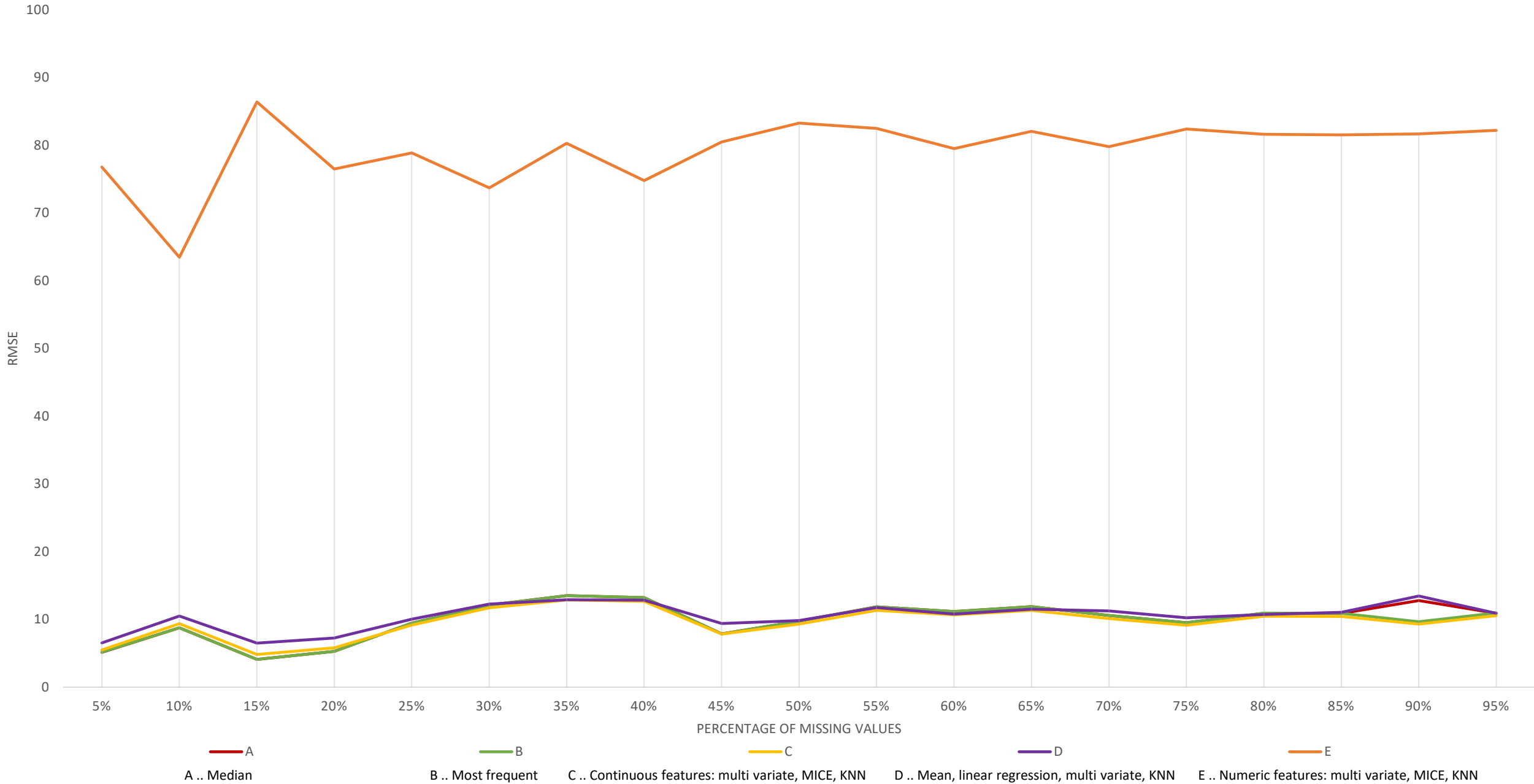
C .. Numeric features: Multi variate, MICE, KNN

Continuous Features

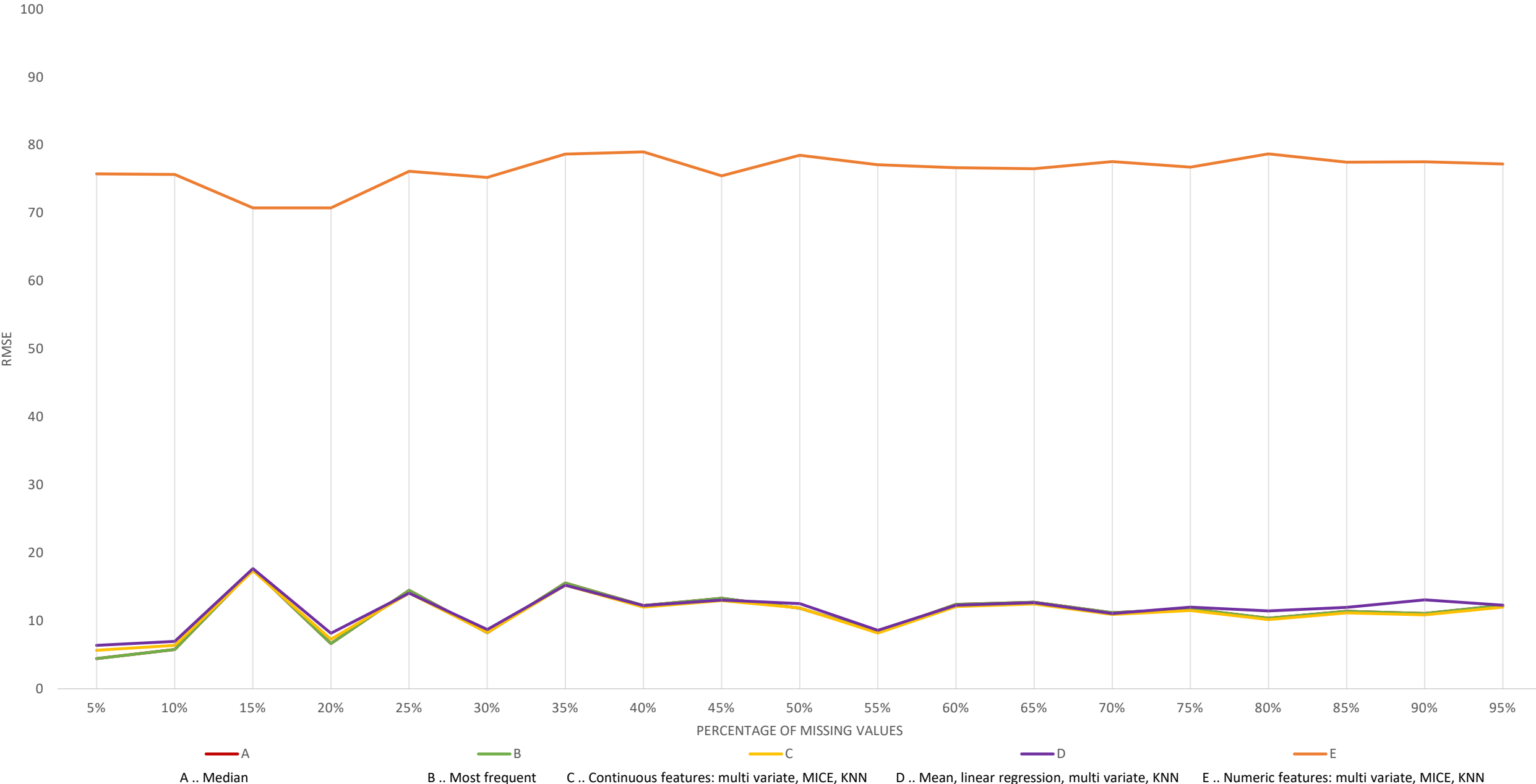
POST PSA (TRAINING SET)



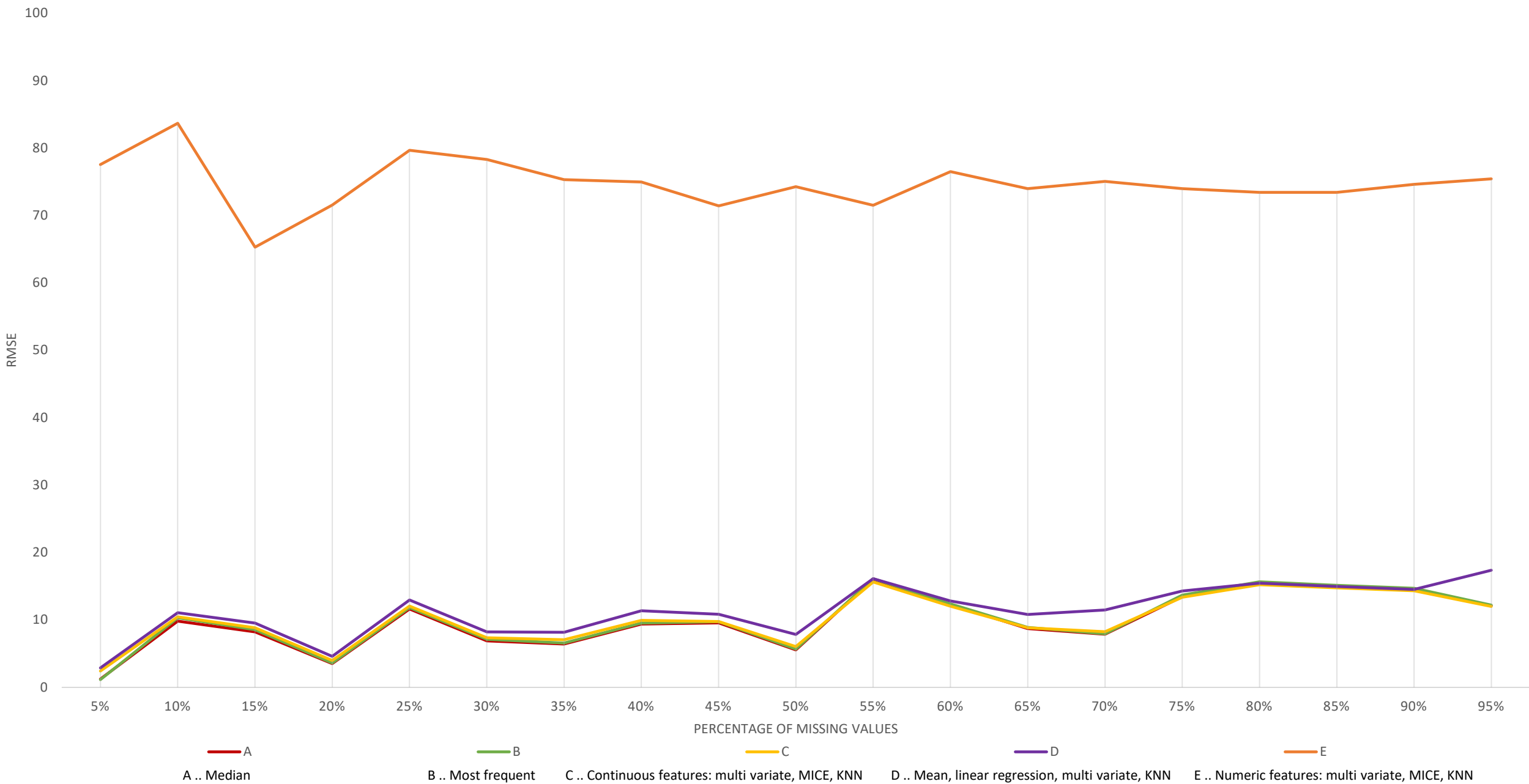
POST PSA (TEST SET)



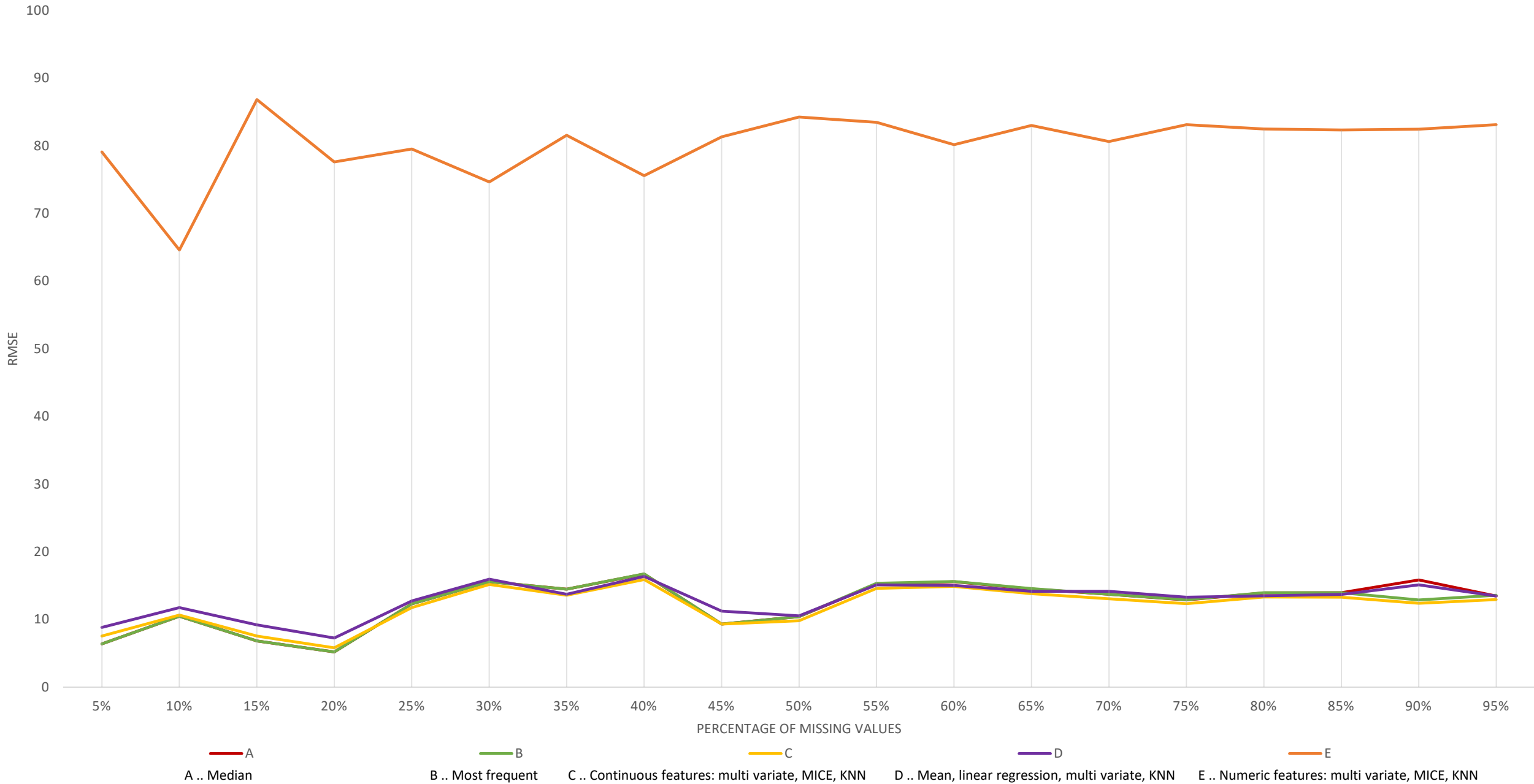
POST PSA (COMPLETE SET)



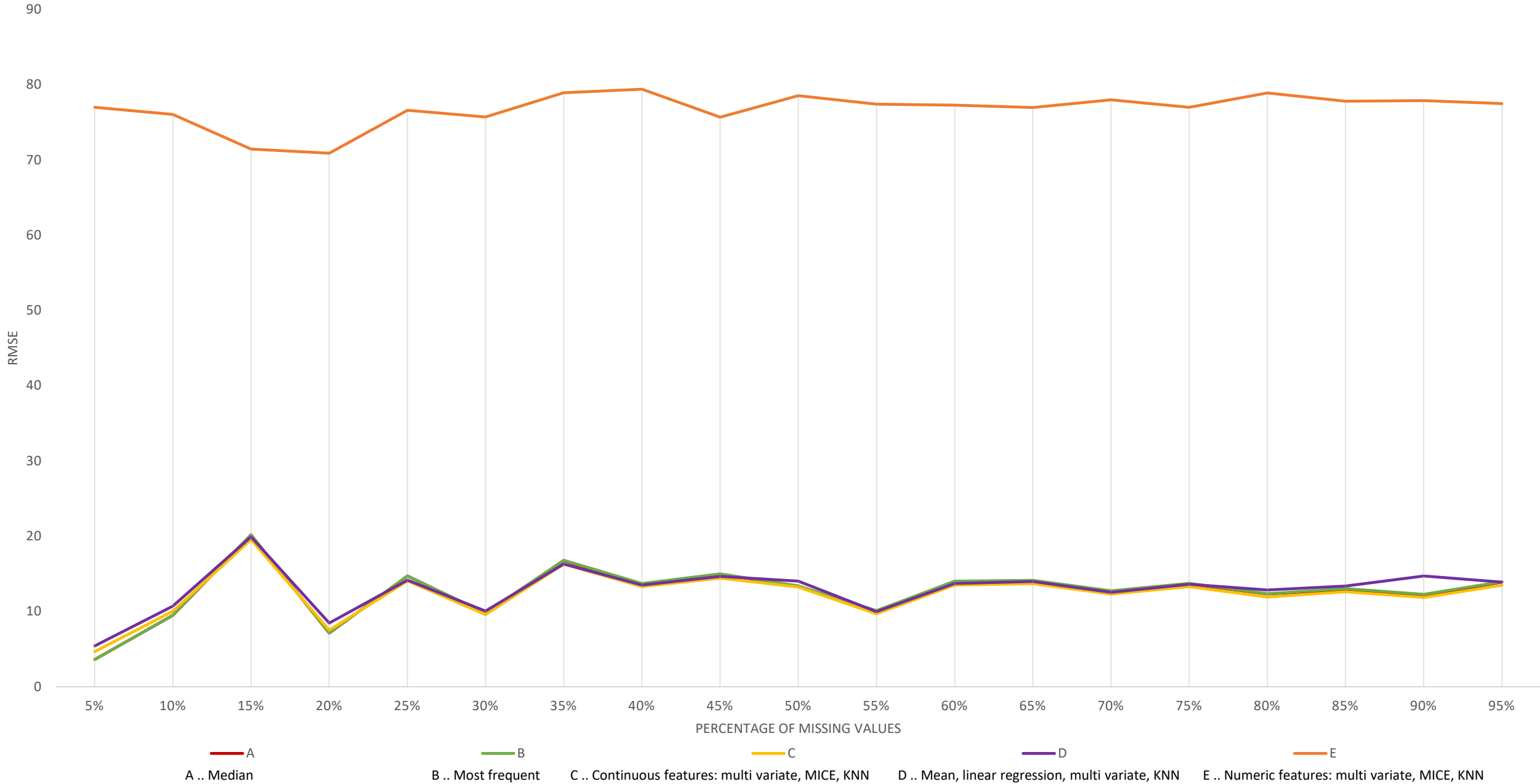
BCR PSA (TRAINING SET)



BCR PSA (TEST SET)

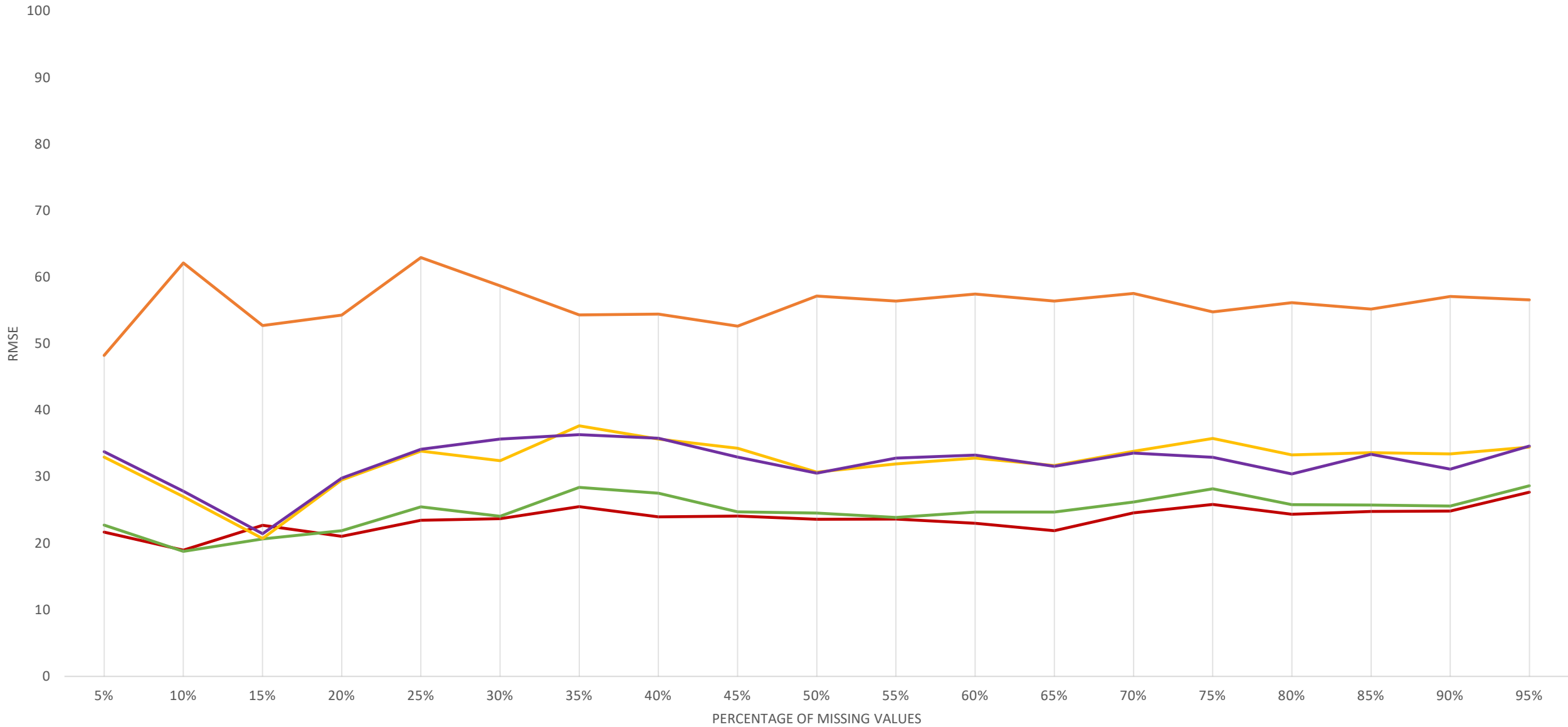


BCR PSA (COMPLETE SET)



Discrete Features

BCR TIME (TRAINING SET)



A

A .. Mean, linear regression, multi variate, KNN

B

B .. Median

C

C .. Discrete features: multi variate, MICE, KNN

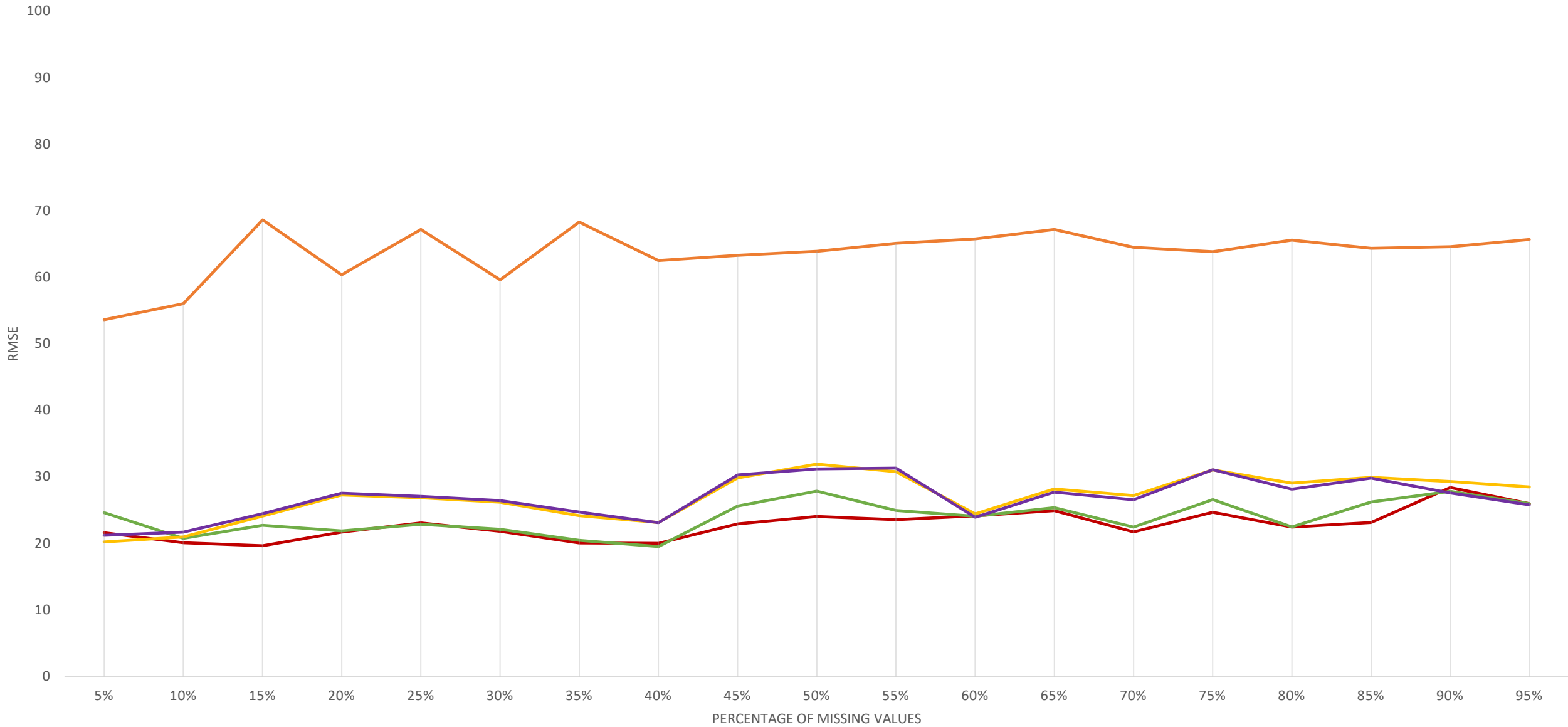
D

D .. Most frequent

E

E .. Numeric features: multi variate, MICE, KNN

BCR TIME (TEST SET)



A

A .. Mean, linear regression, multi variate, KNN

B

B .. Median

C

C .. Discrete features: multi variate, MICE, KNN

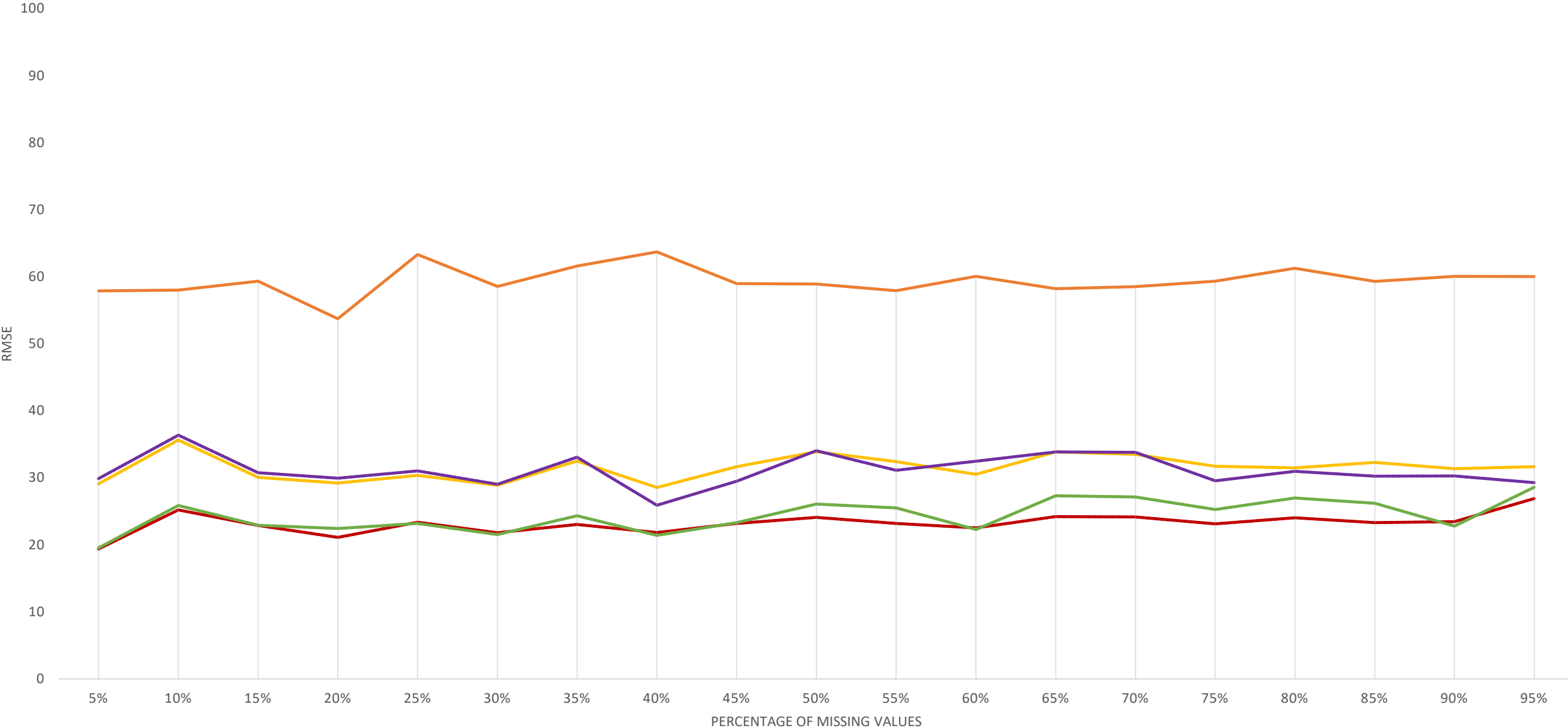
D

D .. Most frequent

E

E .. Numeric features: multi variate, MICE, KNN

BCR TIME (COMPLETE SET)



A

B

C

D

E

A .. Mean, linear regression, multi variate, KNN

B .. Median

C .. Discrete features: multi variate, MICE, KNN

D .. Most frequent

E .. Numeric features: multi variate, MICE, KNN