CEP Clinical and Experimental Pediatrics [Epub ahead of print]
https://doi.org/10.3345/cep.2025.00759

Supplementary Table 2. Missing values for continuous variables

Variable	Percentage of missing values (%)	Strategy to address missingness
Direct bilirubin (μmol/L)	98.66	Remove variable
Creatine kinase (U/L)	79.93	Remove variable
C-reactive protein (mg/L)	49.83	Multiple imputation
Total bilirubin (μmol/L)	37.46	Multiple imputation
Albumin (d/L)	36.12	Multiple imputation
Alanine transaminase (U/L)	35.12	Multiple imputation
Alkaline phosphatase (U/L)	34.78	Multiple imputation
Absolute lymphocyte count ($\times 10^9$/L)	7.36	Multiple imputation
Platelet ($\times 10^9$/L)	4.68	Multiple imputation
Total white cell count ($\times 10^9$/L)	4.35	Multiple imputation
Duration of invasive ventilation (day)	97.74	Remove variable
Duration of oxygen therapy (day)	64.41	Pairwise deletion

A total of 12 variables contained missing values. The highest proportion of missing data was observed in direct bilirubin (98.7%), duration of invasive ventilation (97.7%), creatine kinase (79.9%). These variables were excluded from the analysis due to extreme missingness (>50%), which impair statistical reliability. The missingness in the duration of oxygen therapy (days) variable was systematically related to patients who did not receive oxygen therapy. Therefore, pairwise deletion was used to handle the missing data. The remaining eight variables with missing values were assessed for randomness using Little Missing Completely at Random (MCAR) test, which yielded a *P* value of 0.09, indicating that the missingness pattern was missing at random. Multiple Imputation by Chained Equations (MICE) was employed to address missing data for the remaining variables. The MICE model was implemented using the Huber Regressor algorithm to ensure robust estimation in the presence of outliers. The model was configured to run for a maximum of 10 iterations, with a convergence tolerance threshold of 0.001 to ensure computational stability. To enhance the accuracy of imputations, the algorithm utilized the 5 most relevant variables (n_nearest_features=5) when estimating missing values. The imputation order was specified as ascending, prioritizing variables with fewer missing values to be imputed first. The codes for imputations are available at the following link: https://github.com/chuinhen/Flu_COVID19/blob/main/code/FluCOVID19.ipynb.