Real-time handling of Missing Predictor Values when implementing and using prediction models in daily practice

Steven WJ Nijman[1], T Katrien J Groenhof[1], Jeroen Hoogland[1], Michiel L Bots[1], Menno Brandjes[2], John JL Jacobs[2], Folkert W Asselbergs[345], Karel GM Moons[1], Thomas PA Debray[14]

[1]*Julius Center for Health Sciences and Primary Care, Utrecht University, Utrecht, The Netherlands;*
[2]*LogiqCare, Ortec B.V. Zoetermeer, The Netherlands;*
[3]*Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;*
[4]*Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom;*
[5]*Health Data Research UK, Institute of Health Informatics, University College London, London, United Kingdom;*

Background – Using prediction models to calculate a patients individual risk in clinical practice, requires complete information on all predictors in the prediction model. Unfortunately, routine care data is often incomplete due to a variety of reasons. Although several methods for real-time imputation of missing predictor values exist, they often require immediate access to data from other similar patients and are therefore not directly suitable for routine care.

Aims – To develop and evaluate methods for real-time imputation of missing predictor values in routine clinical care when applying prediction models to individual patients.

Methods – We describe (i) mean imputation (where missing values are replaced by the sample mean), (ii) joint modeling imputation (JMI, where we use a multivariate normal approximation to generate patient-specific imputations) and (iii) conditional modeling imputation (CMI, where a multivariable imputation model is derived for each predictor from a population). We compared the imputation methods by applying a previously developed prediction model (predicting 10-year risk of recurrent vascular disease) in a dataset with 3,880 participants from the Utrecht Cardiovascular Cohort in which missing predictor values were simulated. Furthermore, comparing true and imputed predictor values, the root mean squared error (RMSE) and coverage of the 95% confidence intervals (i.e. the proportion of confidence intervals that contain the true predictor value) were evaluated.

Results – We found that RMSE was lowest when adopting JMI or CMI, although imputation of individual predictors did not always lead to substantial accuracy improvements with regards to the RMSE, as compared to mean imputation. JMI and CMI appeared particularly useful when the values of multiple predictors of the model were missing. Coverage reached the nominal level (i.e. 95%) for both CMI and JMI.

Conclusions – Multiple imputation using, either CMI or JMI, is recommended when dealing with missing predictor values in real time settings.

Keywords
Missing data, multiple imputation, real-time imputation, prediction, decision support system, electronic health care records