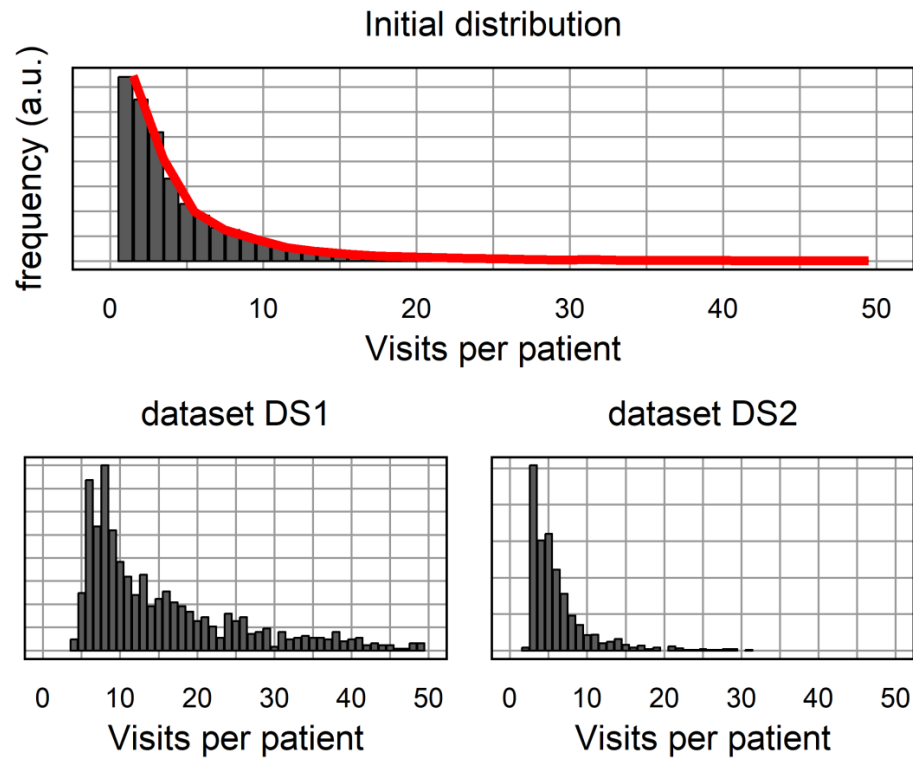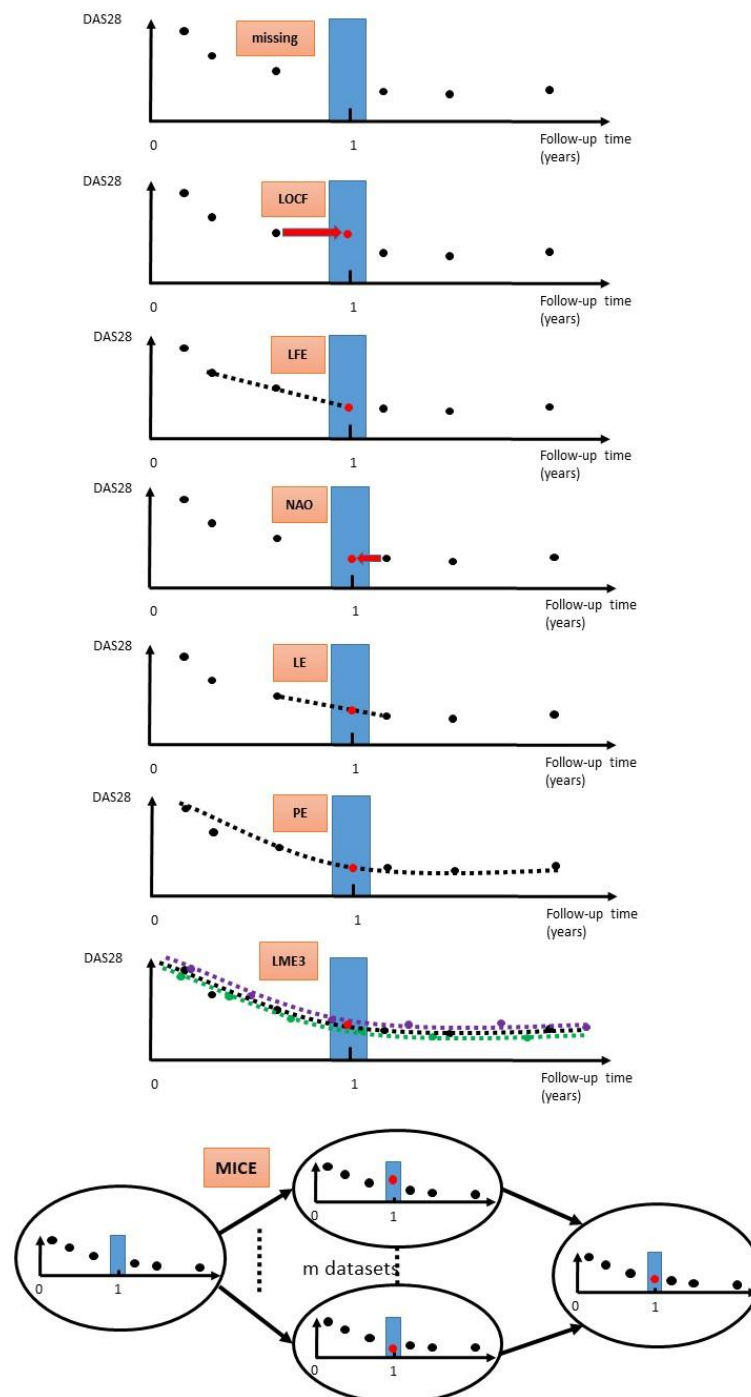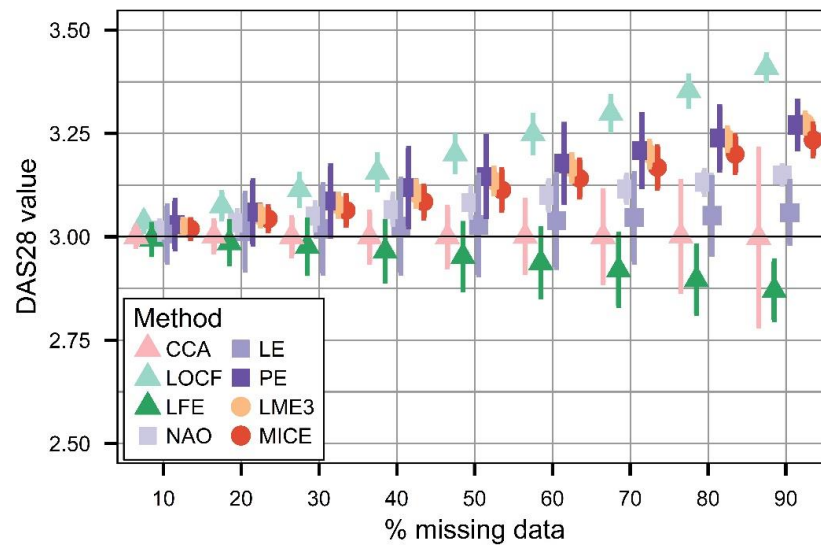## Supplementary Figures

Supplementary Figure S1

Supplementary Figure S1: Distribution of number of visits per patient for the initial collection of registers (top), and the two selected datasets for the simulation study (bottom). The red curve represents the probability distribution function used to sample DS1 in order to obtain DS2.
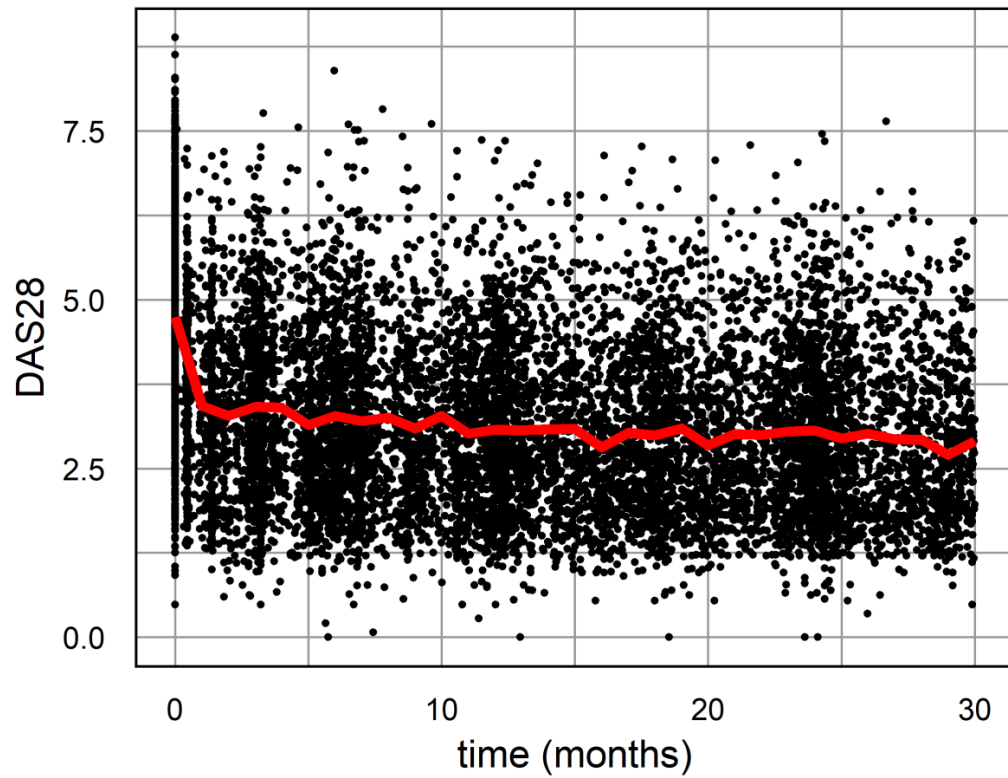
Supplementary Figure S2

Supplementary Figure S2: Schematic of imputations methods for missing value of DAS28 at 1 year follow-up. The blue zone is the 3 month centered time windows used to pick a value at follow-up. First schematic present the situation of missing value. The imputation methods presented are Last Observation Carried Forward (LOCF), Linear Forward Extrapolation (LFE), Nearest Available Observation (NAO), Linear Extrapolation (LE), Polynomial Extrapolation (PE), linear mixed effect with cubic time dependence (LME3), and multiple imputation by chained equation (MICE).
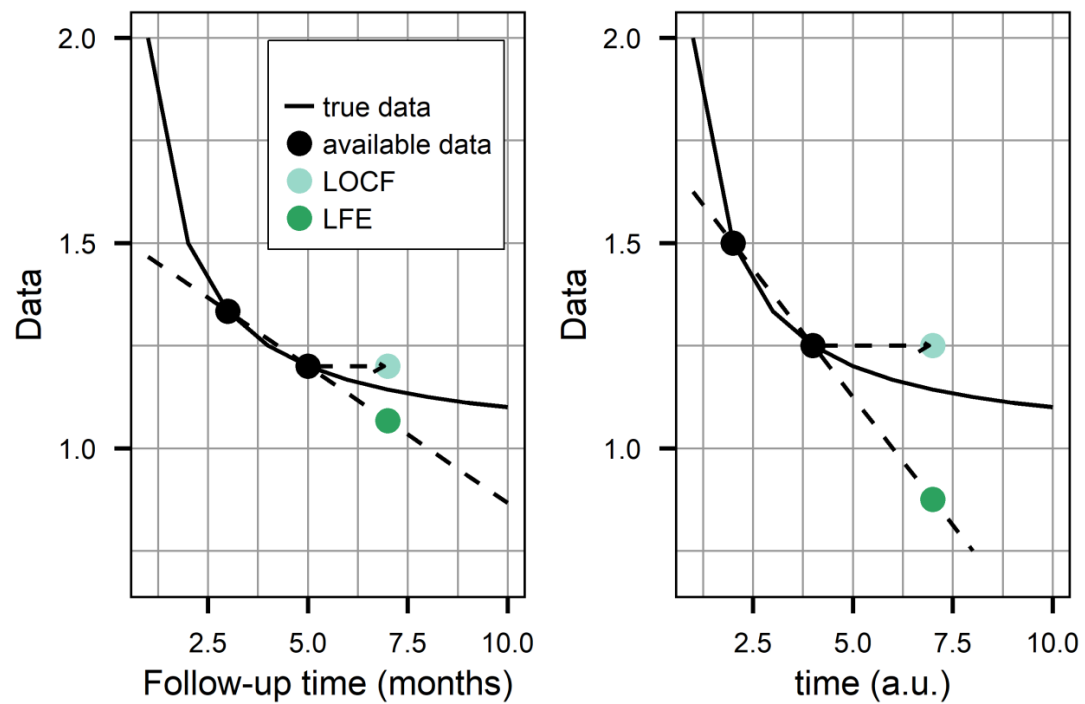
Supplementary Figure S3

Supplementary Figure S3: Evolution of the mean DAS28 values at 12 months, for data missing at random in function of the increase of percentage of missing data at imputation time. Imputation methods are Complete Case Analysis (CCA), Last Observation Carried Forward (LOCF), Linear Forward Extrapolation (LFE), Nearest Available Observation (NAO), Linear Extrapolation (LE), Polynomial Extrapolation (PE), Linear Mixed Effect cubic regression (LME3) and Multiple Imputation by Chained Equation (MICE).
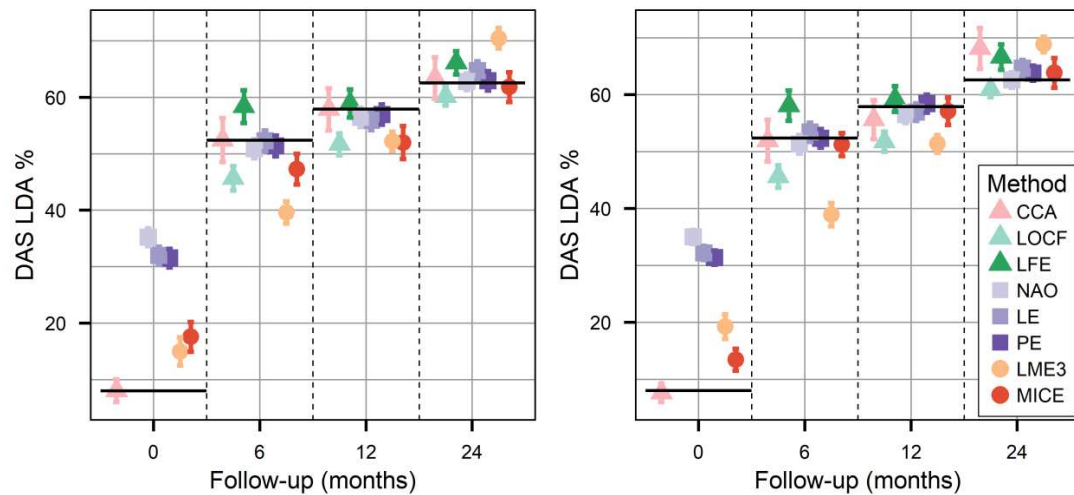
Supplementary Figure S4

Supplementary Figure S4: DAS28 measures for all patient in DS1 (points) and the average of these values calculating with a moving average over 1 month (red curve).
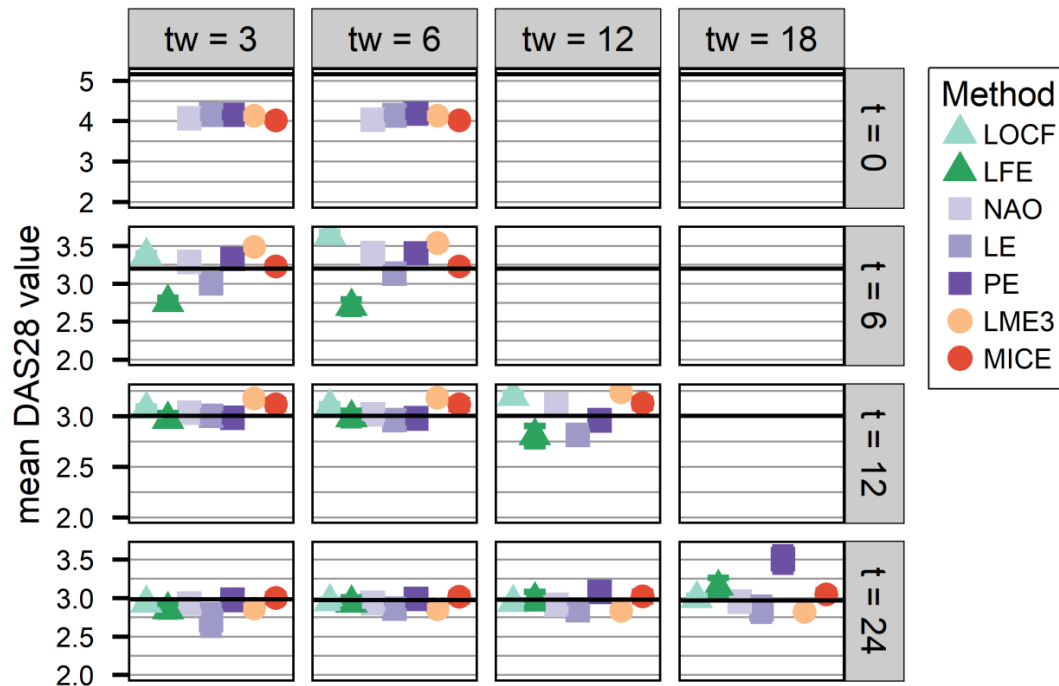
Supplementary Figure S5

Supplementary Figure S5: Illustration of the overestimation of Last Observation Carried Forward (LOCF) imputation and underestimation of Linear Extrapolation (LE) imputation for a variable having a convex decrease in time (solid line), with data available (black dots) close (left) to the imputation time and far (right) from it.
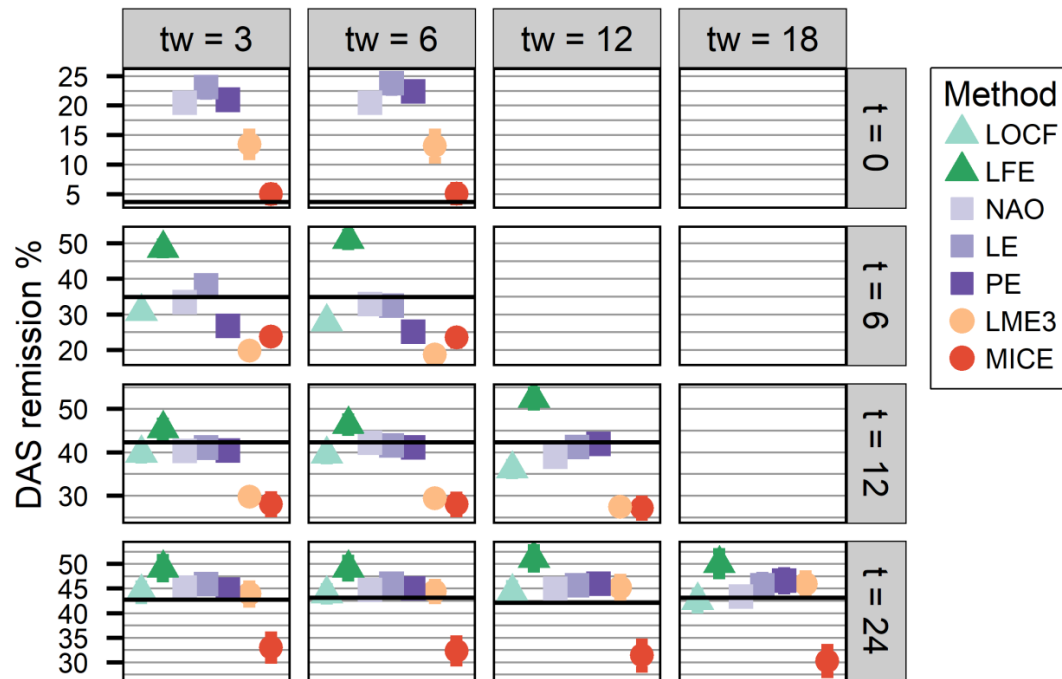
Supplementary Figure S6

Supplementary Figure S6: Mean low disease activity rate calculated over the 1000 patients of DS2 for the different imputation methods, a percentage of missing value of 60% at the imputed time, a deletion time windows of 3 months with data missing completely at random data (MCAR) (left) following an attrition bias (right). Error bars are the standard errors calculated over the 1000 iteration of the imputations. Black horizontal lines represent the true low disease activity rate for each imputation time. Imputation methods are Complete Case Analysis (CCA), Last Observation Carried Forward (LOCF), Linear Forward Extrapolation (LFE), Nearest Available Observation (NAO), Linear Extrapolation (LE), Polynomial Extrapolation (PE), Linear Mixed Effect cubic regression (LME3) and Multiple Imputation by Chained Equation (MICE).

Supplementary Figure S7

Supplementary Figure S7: Mean relative bias of the mean imputed value of DAS28 on DS1 dataset when data is missing at random, for a percentage of missing value of 60% at follow-up t and a deletion time windows of t_w months. Error bars are the standard errors calculated over the 1000 iteration of the imputations. LFE and LOCF estimation are missing for t = 0 since this methods use previous values, which for t = 0 do not exist. Imputation methods are Last Observation Carried Forward (LOCF), Linear Forward Extrapolation (LFE), Nearest Available Observation (NAO), Linear Extrapolation (LE), Polynomial Extrapolation (PE), Linear Mixed Effect cubic regression (LME3) and Multiple Imputation by Chained Equation (MICE).

Supplementary Figure S8

Supplementary Figure S8: Mean DAS28 remission rate calculated over the 1000 patients of DS1 when data is missing at random for a percentage of missing value of 60% at the imputed time t and a variable deletion time windows of t_w months. Error bars are the standard errors calculated over the 1000 iteration of the imputations. Black horizontal lines represent the true remission rate for each imputation time. Imputation methods are Last Observation Carried Forward (LOCF), Linear Forward Extrapolation (LFE), Nearest Available Observation (NAO), Linear Extrapolation (LE), Polynomial Extrapolation (PE), Linear Mixed Effect cubic regression (LME3) and Multiple Imputation by Chained Equation (MICE).