**Statistical Analysis**

Prevalence estimate for child marriage in Andhra Pradesh and Telangana will be calculated from the total sample of women age between 20 to 24 years and the estimates will be provided separately. Differences in percentages of child marriage with respect to socio demographic and marital characteristics will be presented using percentages and frequencies. Bivariate analysis will be performed using chi-square test with 5% level of significance and p value will be reported.

The Bayesian binary logistic regression model will be used to predict the probability of child marriage as a function of the factors. The potential factors influencing the outcome will be selected by bivariate analysis using chi-square test of association between child marriage and each of the factors. Those factors that show p-value less than 0.25 will be taken to the model. The cut-off value of 0.25 is supported by literature since more traditional levels such as 0.05 can fail in identifying factors known to be important. The proceed with Bayesian logistic regression it is necessary to provide a likelihood function for the data and prior distribution. The data from NFHS 4 of individual participants which are assumed to be independent of each other will be considered the likelihood function. The most common prior choice for logistic regression parameter is non-informative prior normal distribution with mean 0 and large variance (1000).

Markov chain Monte Carlo (MCMC) method by Gibbs sampling will be used to simulate samples from the posterior distribution and all the analysis will be implemented on STATA 14 software. Convergence will be assessed by Convergence assessment methods; time series(history) plot, density plot, Monte Carlo standard errors and Gelman-Rubin statistic.

Summary statistics (posterior mean, standard error of posterior mean, odds ratio of posterior mean and credible interval of posterior mean) will be computed for each parameter. The importance of each of the factors will be assessed by carrying out statistical tests of the significance of the regression coefficients (posterior mean) via 95% Bayesian credible interval of the posterior mean.

Two Bayesian logistic models will be provided identifying the factors for Andhra Pradesh and Telangana.