**Program specification document**

*We recommend that completing in this document should be an early step for unit members in writing a new methodology program.*

**Name of program**: **mimix**.

**Brief description**:

R port of the mimix stata program – Reference based imputation for sensitivity analysis of a longitudinal trial with protocol deviation

**Software package to be used (incl. version):** R (>v 3.5.0)

**Names of the contributors to this specification document:** Kevin McGrath, Ian White

**Developer:** Kevin McGrath.

**Proposed reviewer(s):** Ian White, James Carpenter, Matteo Quartagno, Suzie Cro

**Date**: 12/06/2020.

|  |
| --- |
| Aim of program: |
| To emulate the functionality of the mimix program written in Stata with some differences :- the Stata version has an interim option which the R version does not and the R version includes a Causal option and Delta adjustment which the Stata version does not. Imputation method 6  imputation options meth=  Missing at Random MAR,  last mean carried forward, LMCF  jump to reference J2R, requires reference treatment group  copy increments in reference CIR, requires reference treatment group  copy reference CR, requires reference treatment group    Additional to Stata version :  Causal method requires K0,K1 values  Delta adjustment requires delta ,dlag values  The MCMC method chosen uses functions emNorm and mcmcNorm from the norm2 package by Schafer (norm2 user guide 2016).  Warnings from the norm2 package are suppressed and are explained in the norm2 user guide as indicating that the EM solution causes the covariance matrix to be nearly singular. The prior may be changed from to the ridge prior instead of the default jeffreys if this is thought to be an issue. |
| Input data format(s): |
| *Typically a data set: in this case specify variables required, e.g. confidence intervals or standard errors, and whether numeric/string data. Include screenshots of example input data sets.*  *Longitudinal data-set required in long format (one record per individual per time-point) from the data-set,*  *user to specify the data arguments;*  *the response/dependent variable, ( numeric)*  *treatment group, (numeric)*  *time variable , (numeric)*  *individual id, (numeric or character)*  *plus any additional fully observed baseline covariates, (numeric)*  *covar (vector specifying numeric data types)*    *For the analysis arguments, the user must specify,*  *meth, the imputation method (from the 6 methods) (char)*  *M , the number of imputations required, (numeric)*  *refer, the reference treatment group (applicable for meth = J2R, CIR, CR, Causal) (char)*  *seedl set a random-number seed value to enable replication of results (numeric)*  *prior, change prior from jeffreys to ridge or uniform (char)*  *burnin, number of iterations for the burn-in period in the Markov chain Monte Carlo procedure (default 1000)*  *bbetween, number of iterations between imputations in the MCMC procedure*  *methodvar, 2 element vector indicating individual-specific columns of method and reference group.*  *delta vector of numeric values same length as number of time-points for Delta adjustment of imputed values , corresponding to a values in Roger’s 5 macros paper*  *dlag vector of numeric values same length as number of time-points for Delta adjustment corresponfing to corresponding to b values in Roger’s 5 macros paper*  *K0 numeric value if Causal method specified*  *K1 numeric value for the exponential constant in Causal method from White, Royes, Best paper.*  Figure 1 Asthma data-set example      Figure 2 Antidepressant data set showing individual-specific cols |
| Example syntax statement(s): |
| *impdata<-*  *mimix("antidepressant",c("basval","PATIENT.SEX"),"HAMD17.TOTAL","TREATMENT.NAME","PATIENT.NUMBER","VISIT.NUMBER",5,1,NULL,101,c("jeffreys"),1000,NULL,c("methodvar","referencevar"),c(0.5,0.5,1,1 ))*  *impdata*<-  *mimix("antidepressant",c("basval","PATIENT.SEX"),"HAMD17.TOTAL","TREATMENT.NAME","PATIENT.NUMBER","VISIT.NUMBER",5,1,"Causal",101,c("jeffreys"),1000,NULL,NULL,c(0.5,0.5,1,1 ),2)* |
| Outputs: |
| *e.g. graphs, text, data etc., and/or how the data set is changed.*  *Save output to user specified data-set containing the M imputed data sets appended to the original data-set in wide format*    *mpantiIndivDt <- mimix("antidepressant",c("basval","PATIENT.SEX"),"HAMD17.TOTAL","TREATMENT.NAME","PATIENT.NUMBER","VISIT.NUMBER",5,1,NULL,101,c("jeffreys"),1000,NULL,c("methodvar","referencevar"),c(0.5,0.5,1,1*    During running, *the program displays missing pattern lookup table plus summary table plus possible warning message originating from the norm2 package*  *Note: Eigen power method failed to converge*  *OCCURRED IN: estimate\_worst\_frac in MOD norm\_engine*  *(this may disappear if the prior is changed to ridge)*    *Plus summary table of patients in sub-grouping*    *To find regression output on imputed data sets*    *library(mice)*  fit*<-with(data= as.mids(impantiIndivDt,.id="SNO",.imp="II"),expr=lm(HAMD17.TOTAL.7~TREATMENT.NAME+basval+PATIENT.SEX))*  *summary(pool(fit))* |
| Flowchart/pseudo-code of multiple program files with commands/data (if applicable): |
| Runmimix program  *mimix*  *Read data set ,*  *Check inputs,*  *If meth then call preprodata*  *Convert longitudinal long format data to wide format creating depvar.time variables*  *and find missing data patterns*  *Output list including finaldat and mg missing pattern lookup ,*  *(also pass other values back)*  *Recode meth from character value to numbers 1-6*  *else if methodindiv then call preproIndivdata*  *Convert longitudinal long format data to wide format creating depvar.time variables*  *and find missing data patterns*  *Output list including finaldat and mg missing pattern lookup ,*  *(also pass other values back)*  *end*  *some processing …. then*  *loop over each treatment*  *from finaldat create subsets for each treatment group with selected vars (contains missing)*  *loop over M imputations*  *perform mcmcNorm on a regression (multivariate normal)*  *drawing beta and Sigma parameters from the Bayesian posterior*  *(warning msgs may be created as below) see norm2 user guide for explanation*  *Note: Eigen power method failed to converge*  *OCCURRED IN: estimate\_worst\_frac in MOD norm\_engine*    *Save the* M parameters.Beta and parameters.Sigma,  *in paramBiglist*  *end M loop*  *end treatment loop*  *Loop over every row from the missing pattern lookup table mg*  *Loop over 1:M imputations*  *If meth then*  *Create vectors signifying missing columns according to the missing pattern group*  *create a suitable parameter means vector and Sigma matrix depending on which method*  , - J2R, CIR, MAR…  *else if methodindiv*  *call ifmethodindiv to create suitable means and Sigma*    *Use the parameter means, and Sigmas so created in the m’th imputation in the*  *algorithm to calculate vector of estimates (amongst other feature using Cholesky transformation*  *and inverse normal generator).*    *End loop M*  *End Loop row pattern group*  *If Delta adjustment required call AddDelta*  *To run regression on imputed datasets*  *convert imputed dataset to mids format (from mice package)*  *impdata= as.mids(impdatasetJ2R, .id="SNO",.imp="II")*  *fit<-with(impdata, lm(fev.12~treat+base))*  *summary(pool(fit))* |
| Initial error checking list: |
| *e.g. if no data, if some variables are missing/incorrectly specified etc.*   1. *Check meth or methodvar specified but not both* 2. *Check reference group specified when meth = J2R,CIR,CR,Causal* 3. *Check program runs when no covars specified* 4. *Check covariates all numeric* 5. *Check baseline covariates all complete* 6. *Check Causal constants K0,K1 specified when meth=Causal* 7. *Check results consistent with Stata output , run m=200, one method is use analyselist() to take individual ids with different missing patterns and compare estimates with Stata output. Another method is to compare regression results under Rubin’s rules.*     1. *MAR*    2. *J2R*    3. *CIR*    4. *CR*    5. *LMCF* 8. *Check Causal method with White, Royes, Best paper* 9. *Check Causal Kd=0 , 1 agrees with J2R, CIR estimates,* 10. *Check Delta method with?* 11. *Check mice method of pooling regression results* 12. *Check norm 2 method of mcmc generation* |