Multiple Imputation

VGSM, Chapter 11

Dave Glidden

Last Week

- Missing data grouped by relationship between missingness and it's value
- MCAR/MAR and NMAR
- MAR is the assumption where most action is
- Motivates multiple imputation
- Model-based guesses at missing data

UNOS Dataset

- Data on 9775 pediatric kidney transplants from 1990-2002
- Followed for transplant outcomes
  465 deaths
- Predictors: age, age of donor, year, HLA loci, txtype
- Cold ischemia time: missing in 23%

Multiple Imputation

1. Assume data MAR
2. Develop predictive model for missing data
3. Sample from predictive model
4. Make multiple complete datasets
5. Analyze datasets
6. Synthesize results

Selecting MAR Variables

- Bias is the biggest threat
  - More predictors make MAR plausible
  - Include the outcome
- Unimportant covariates: little cost
  - probably not much gained
- Any predictor in model -- include impute
  - possible some that aren't

Modeling Choice

- Specify a parametric model for imputations
- cold_ischemia: continuous variable
  but also has many "0" s
- Many missing values like to be "0"
- Difficult distribution to model
- Try Normal distribution

Predictive Model

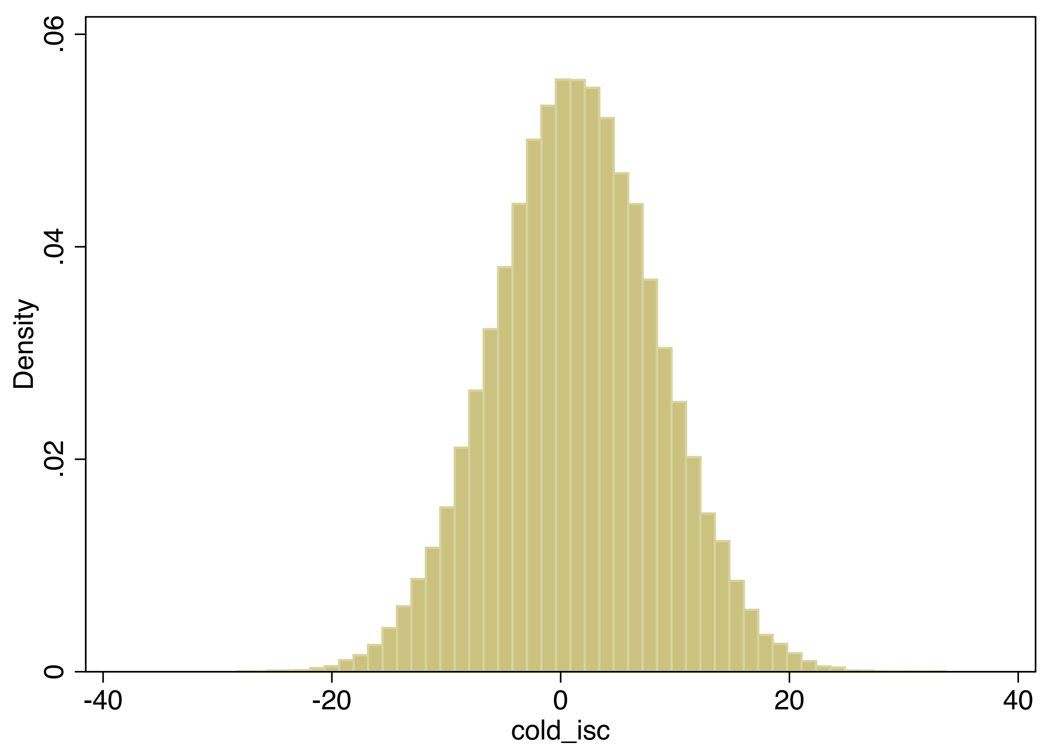
```
reg cold_isc txty death i.hla age_don age i.hlamat
_Ihlamat_0-6        (naturally coded; _Ihlamat_0 omitted)

      Source |       SS       df       MS              Number of obs =    7347
-------------+------------------------------           F( 10,  7336) = 1208.63
       Model |  606123.292    10  60612.3292           Prob > F      =  0.0000
    Residual |  367898.979  7336  50.1498062           R-squared     =  0.6223
-------------+------------------------------           Adj R-squared =  0.6218
       Total |  974022.271  7346  132.592196           Root MSE      =  7.0817

------------------------------------------------------------------------------
    cold_isc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   18.62832   .2249336    82.82   0.000     18.18738    19.06925
       death |   .6670558    .388197     1.72   0.086    -.0939219    1.428034
             |
      hlamat |
          1  |   .9133429   .3598258     2.54   0.011     .2079808    1.618705
          2  |   .7184059   .3568126     2.01   0.044     .0189507    1.417861
          3  |   1.232046   .3520939     3.50   0.000     .5418412    1.922252
          4  |   1.377349    .395934     3.48   0.001     .6012044    2.153493
          5  |   1.955531   .5342044     3.66   0.000     .9083367    3.002725
          6  |   2.253536   .5631313     4.00   0.000     1.149637    3.357436
             |
     age_don |   .0124535   .0068079     1.83   0.067    -.000892     .0257989
         age |    .028418   .0162956     1.74   0.081    -.003526     .0603619
       _cons |  -.5646455   .4347184    -1.30   0.194    -1.416818     .2875275
------------------------------------------------------------------------------
```

Normal Imputations

Mean 1.4. SD =7



Misspecified Distribution

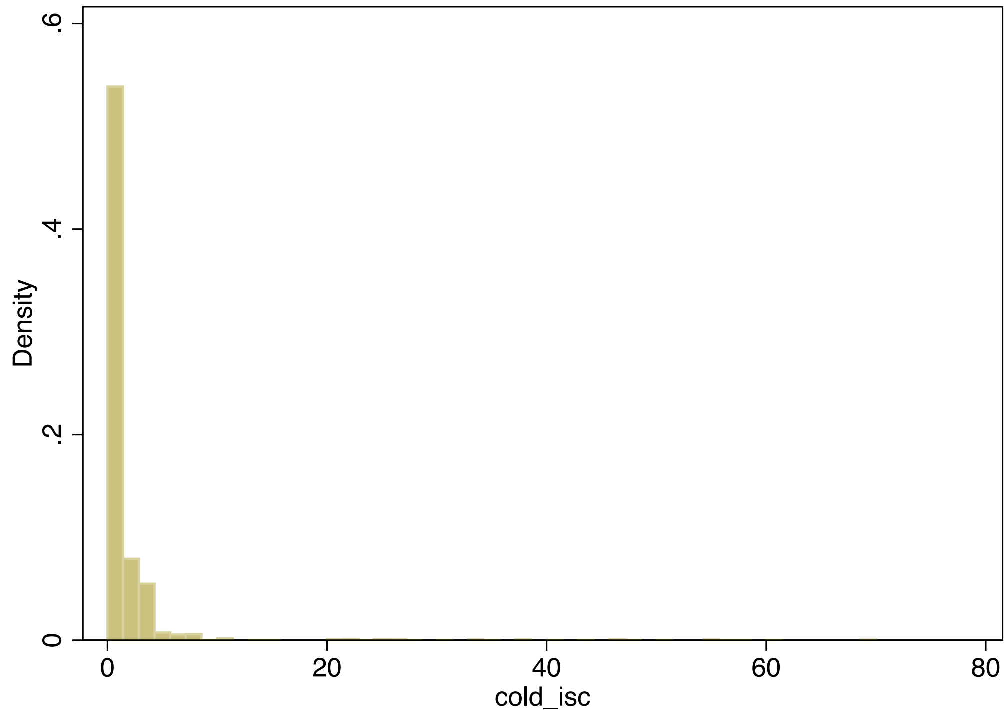
- Minimal effect
- Most important to get moments correct
  means, SD, covariance of imputed data
- Unappealing to get implausible values
- Option to force plausible value

Predictive Means Matching

- Linear regression fit to missing data
  $E(Y_{mis}) = \beta_0 + \beta_1 x_1 + \ldots.\ \beta_p x_p$
- Sample from "posterior" of β | Data
- With each beta calculate predictions
  for all observations
- Impute Y_{mis} from observed values k nearest neighbors
- Forces plausible value
  must specify *k* — *Stata default, k=1*

PMM Imputations

Mean 1.4, SD =4



Missing Data

Variable	Obs	Mean	Std. Dev.	Min	Max
hlamat	9541	2.57363	1.378919	0	6
age_don	9662	31.29952	13.38866	0	73
age	9766	11.64653	5.291385	0	18
cold_isc	7525	10.85967	11.51735	0	72
death	9775	.0475703	.2128662	0	1
prevtx	9702	.1443001	.3514119	0	1
txtype	9775	.4733504	.4993148	0	1

Multiple Missing Variable

- Can be handled by multiple imputation
- Means developing a model for several variable simultaneous
- Two main ways to do this
  - assume variables are normal
    so called multivariate normal
  - "chained" regression equal
    so called iterative chained equations

Multivariate Normal Imputation

- Treats all variables as normal
- Imputes using linear regression
  imputes well for linear
- Can be coaxed to do binary and ordinal
- Can't impute for unordered categorical

Chained Imputation

- Regression model for each missing variable
- “ICE”: iterative chained equations
- Combined for multivariate imputations
- Builds a multivariate distribution
- Separate models: gives big flexibility

Possible Imputation Models

- Continuous: regress, pmm
- Binary: logit
- Ordinal: ologit
- Nominal: mlogit
- Discrete: poisson, nbreg
- Continuous+threshold: truncreg

separate models for each missing variable

3. Make multiple complete datasets

Imputing in Stata

- Extensive suite of capabilities
- Data management
- Creating imputations
- Analyzing and synthesizing

mi package

- Powerful package for missing data
- Extensive capabilities
  lots of choices, appear esoteric
- Can handle multiple missing predictors
- Presents results in easy to read format

mi set

- Must declare data to be "mi set"
- `mi set mlong`
- Stores the imputations in long format
  just governs how imputations are stored
- There is also `mi stset` & `mi xtset`

mi register

- Must declare variable as needing or not needing imputation
- `mi register imputed cold_isc`
- `mi register regular death hlamat age_don age`
- Says cold_isc will be imputed and others not

mi impute

- This command actually does the imputations
- Can impute single or multiple variables
- Specifies:
  - predictors used for imputation
  - statistical model for imputation
  - number of imputations

imputing and analysis happen in separate steps

mi impute

- This command actually does the imputations
- Can impute single or multiple variables
- Specifies:
    - predictors used for imputation
    - statistical model for imputation
    - number of imputations

imputing and analysis happen in separate steps

Imputing Cold Ischemia

- mi impute regress cold_isc = death txtype i.hlamat age_don age, add(100) rseed(123)
- Use linear regression to impute cold_isc
- Predictors: death, txtype, etc
- add(100) = 100 imputations
- rseed( ) = sets a random seed
  means you can replicate imputations
  very very useful

Imputation Output

```
. mi impute regress cold_isc = hlamat age_don age death prevtx txtype
, add(100) rseed(123)

Univariate imputation                       Imputations =      100
Linear regression                                 added =      100
Imputed: m=1 through m=100                      updated =        0

_________________________________
                   |               Observations per m
                   |----------------------------------------------
          Variable |   Complete   Incomplete   Imputed |     Total
-------------------+-----------------------------------+----------
          cold_isc |       7347         2020      2020 |      9367
------------------------------------------------------------------
(complete + incomplete = total; imputed is the minimum across m
 of the number of filled-in observations.)
```


Predictive Mean Matching

```
. mi impute pmm cold_isc = hlamat age_don age death prevtx txtype,
add(100) rseed(123) dots knn(1)

Univariate imputation                       Imputations =      100
Predictive mean matching                         added  =      100
Imputed: m=1 through m=100                      updated =        0

------------------------------------------------------------------
                   |               Observations per m
                   |----------------------------------------------
          Variable |   Complete   Incomplete   Imputed |     Total
-------------------+-----------------------------------+----------
          cold_isc |       7347         2020      2020 |      9367
------------------------------------------------------------------
(complete + incomplete = total; imputed is the minimum across m
 of the number of filled-in observations.)
```

MI Dataset — 20 Imputations

```
tabulate _mi_m

      _mi_m |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      9,367       18.82       18.82
          1 |      2,020        4.06       22.88
          2 |      2,020        4.06       26.94
          3 |      2,020        4.06       31.00
          4 |      2,020        4.06       35.06
          5 |      2,020        4.06       39.12
          6 |      2,020        4.06       43.18
          7 |      2,020        4.06       47.23
          8 |      2,020        4.06       51.29
          9 |      2,020        4.06       55.35
         10 |      2,020        4.06       59.41
         11 |      2,020        4.06       63.47
         12 |      2,020        4.06       67.53
         13 |      2,020        4.06       71.59
         14 |      2,020        4.06       75.65
         15 |      2,020        4.06       79.71
         16 |      2,020        4.06       83.76
         17 |      2,020        4.06       87.82
         18 |      2,020        4.06       91.88
         19 |      2,020        4.06       95.94
         20 |      2,020        4.06      100.00
------------+-----------------------------------
      Total |     49,767      100.00
```

original data

imputation “10”

9,347 record dataset has 49,767 records from 20 impute

Chained Syntax

- mi register imputed hlamat age_don age cold_isc prevtx death txtype
  register everything
- mi impute chained (pmm, knn(1)) cold_isc (ologit) hlamat (regress) age_don (regress) age (logit) prevtx = death txtype, add(100) rseed(897)
- *(method) variable = nonmissing*
- Very flexible syntax

Conditional Models

```
 mi impute chained (pmm, knn(1)) cold_isc (ologit) hlamat (regress) age_don ///
>  (regress) age (logit) prevtx =  death txtype, add(100) rseed(897) replace dots

Conditional models:
               age: regress age i.prevtx age_don i.hlamat cold_isc death txtype
            prevtx: logit prevtx age age_don i.hlamat cold_isc death txtype
           age_don: regress age_don age i.prevtx i.hlamat cold_isc death txtype
            hlamat: ologit hlamat age i.prevtx age_don cold_isc death txtype
          cold_isc: pmm cold_isc age i.prevtx age_don i.hlamat death txtype , knn(1)

Performing chained iterations:

imputing m=1 through m=100
```


Imputations

```
Multivariate imputation                     Imputations =      100
Chained equations                                 added =      100
Imputed: m=1 through m=100                      updated =        0

Initialization: monotone                     Iterations =      100
                                                burn-in =       10

          cold_isc: predictive mean matching
            hlamat: ordered logistic regression
           age_don: linear regression
               age: linear regression
            prevtx: logistic regression

------------------------------------------------------------------
                   |               Observations per m
                   |----------------------------------------------
          Variable |   Complete   Incomplete   Imputed |     Total
-------------------+-----------------------------------+----------
          cold_isc |       7525         2250      2250 |      9775
            hlamat |       9541          234       234 |      9775
           age_don |       9662          113       113 |      9775
               age |       9766            9         9 |      9775
            prevtx |       9702           73        73 |      9775
------------------------------------------------------------------
```

Multiple Imputation

1. Assume data MAR
2. Develop predictive model for missing data
3. Sample from predictive model
4. Make multiple complete datasets
5. Analyze datasets
6. Synthesize results

Analyze and Synthesize

Source of Uncertainty

- Predictive distn of cold_isc
    - $cold\_isc_i = \alpha+\beta_1*x_1+\beta_2*x_2+..\ \beta_p*x_p+\varepsilon_i$
- Sources of uncertainty

1. prediction: ε_i
2. parameters: $\beta_1, \beta_2...,\beta_p$
3. form of model: *predictors, transformation*

- Imputation deals with 1 & 2

Result of 1st Imputation

```
Logistic regression                             Number of obs     =      9,775
                                                LR chi2(11)       =     103.06
                                                Prob > chi2       =     0.0000
Log likelihood = -1818.4095                     Pseudo R2         =     0.0276

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.013205   .0060088     2.21   0.027     1.001497    1.025051
     age_don |   .9901152   .0038156    -2.58   0.010      .982665    .9976219
         age |   .9485558   .0084687    -5.92   0.000      .932102    .9653002
      prevtx |   1.353609   .1697865     2.41   0.016     1.058583    1.730858
      txtype |   1.184901   .2099091     0.96   0.338     .8373152    1.676777
             |
      hlamat |
          1  |   1.039716   .1861719     0.22   0.828     .7319779    1.476832
          2  |   .8916278   .1633056    -0.63   0.531     .6227049    1.276688
          3  |   .8342293   .1559687    -0.97   0.332     .5782872    1.203448
          4  |   .7531256   .1692512    -1.26   0.207     .4848144    1.169929
          5  |   .8627274   .2608962    -0.49   0.625      .476939    1.560574
          6  |    .915812   .2973173    -0.27   0.786     .4846908    1.730405
             |
       _cons |   .1004935   .0226025   -10.22   0.000     .0646681    .1561658
```

Imputations

Imputation #	OR txtype	SE(OR)
1	1.18	0.20
2	1.12	0.20
3	1.17	0.21
4	1.20	0.21
5	1.20	0.21
mean	1.17	0.21
SD	0.03	0.01

Combining Imputations

K = Number of Imputations = 5

MI Estimate = $(OR_1 + OR_2 + \ldots + OR_K)/K$
$= 1.17$

MI Variance = $(1 + 1/K)*\{ SE(OR_k) \}^2 +$
$\{ \text{mean}(\{SE_k(OR)\}^2) \}$

$= (1+1/5)*\{ 0.033 \}^2 +$
$(0.206)^2$
$= .044$

MI SE = sqrt(MI Variance) = 0.21

Imputed Regression

the analysis step

- mi estimate, or : logistic death cold_isc txty i.hlam age_don age
- Fits a logistic regression with imputations based on the last commands
- After ":" usual commands
- Need "or" to get odds ratios

Results

```
mi estimate, eform: logistic death cold_isc age_don age  prevtx txtype i.hlamat

------------------------------------------------------------------------------
       death |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.012858   .0062722     2.06   0.039      1.00062    1.025246
     age_don |   .9904244   .0038148    -2.50   0.012     .9829756    .9979296
         age |   .9485306   .0084643    -5.92   0.000     .9320852    .9652662
      prevtx |   1.357085   .1709674     2.42   0.015     1.060149     1.73719
      txtype |    1.17499   .2110693     0.90   0.369     .8262195    1.670985
             |
      hlamat |
          1  |   1.035961   .1870476     0.20   0.845     .7271112    1.475999
          2  |   .8521283   .1601151    -0.85   0.395     .5894205    1.231926
          3  |     .80732   .1519525    -1.14   0.256      .558221    1.167576
          4  |   .6858873   .1610081    -1.61   0.109     .4326734     1.08729
          5  |   .8363426   .2538196    -0.59   0.556     .4613647    1.516087
          6  |    .913864   .2937117    -0.28   0.779     .4867319    1.715826
             |
       _cons |   .1035627   .0233885   -10.04   0.000      .066521    .1612307
------------------------------------------------------------------------------
```

PMM Results

```
mi estimate, or: logistic death cold_isc age_don age year prevtx txtype i.hlamat

Multiple-imputation estimates                   Imputations     =          5
Logistic regression                             Number of obs   =       9775

------------------------------------------------------------------------------
       death |     exp(b)   Std. Err.       t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.010375   .0071917     1.45   0.154      .9959816    1.024976
     age_don |   .9905783   .0038174    -2.46   0.014      .9831245    .9980887
         age |   .9480711   .0084704    -5.97   0.000      .9316139     .964819
      prevtx |   1.366319    .174341     2.45   0.014      1.063889    1.754719
      txtype |   1.233898   .2356188     1.10   0.272      .8466503    1.798269
             |
      hlamat |
          1  |   1.043439   .1886803     0.24   0.814      .7319728    1.487438
          2  |   .8503009   .1644444    -0.84   0.402      .5813449    1.243688
          3  |   .8090594   .1524061    -1.12   0.261      .5592625    1.170429
          4  |   .6948287    .160795    -1.57   0.116      .4413473    1.093893
          5  |   .8621357   .2653442    -0.48   0.630      .4713913    1.576775
          6  |   .9053687    .294839    -0.31   0.760      .4781864     1.71417
             |
       _cons |   .1033468   .0233963   -10.03   0.000      .0663109    .1610677
------------------------------------------------------------------------------
```


Data Management

- Imputing creates phantom observations
- Data is only good for imputations
- Useful command -- "mi extract" creates a dataset based on one "imputation"
- mi extract 0, clear
  back to regular incomplete data
- mi extract 10, clear
  complete dataset using imputation #10
  allows you to check imputation set

Imputation 10

```
. mi extract 10, clear

. logistic death cold_isc age_don age  prevtx txtype i.hlamat

Logistic regression                               Number of obs   =       9775
                                                  LR chi2(11)     =     109.38
                                                  Prob > chi2     =     0.0000
Log likelihood = -1815.2502                       Pseudo R2       =     0.0292

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.013805   .0059296     2.34   0.019      1.00225    1.025494
     age_don |   .9906357   .0037988    -2.45   0.014     .9832182    .9981092
         age |   .9483539   .0084602    -5.94   0.000     .9319163    .9650814
      prevtx |   1.463773   .1797112     3.10   0.002      1.15072    1.861993
      txtype |   1.124474   .1991213     0.66   0.508     .7947287    1.591035
             |
      hlamat |
          1  |   1.048589   .1855962     0.27   0.789     .7412181    1.483422
          2  |    .835029   .1530105    -0.98   0.325     .5830792    1.195847
          3  |   .7858572   .1462116    -1.30   0.195     .5457258    1.131652
          4  |   .6671839   .1515587    -1.78   0.075     .4274497    1.041372
          5  |   .8654381    .255724    -0.49   0.625     .4849743    1.544377
          6  |    .858816   .2780721    -0.47   0.638     .4552957    1.619969
             |
       _cons |   .1043656   .0233851   -10.09   0.000     .0672712    .1619143
------------------------------------------------------------------------------
```

but this dropped all other imputations

mi xeq

simpler approach

- mi xeq 10: summ cold_isc
  summarize cold_isc from imputation 10

```
m=10 data:
-> summ cold_isc

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    cold_isc |      9,775    10.14958     11.4468          0         72
```

mi xeq

examining results

```
 mi xeq 10: logistic death cold_isc age_don age  prevtx txtype i.hlamat

-> logistic death cold_isc age_don age  prevtx txtype i.hlamat

Logistic regression                               Number of obs   =       9775
                                                  LR chi2(11)     =     109.38
                                                  Prob > chi2     =     0.0000
Log likelihood = -1815.2502                       Pseudo R2       =     0.0292

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.013805   .0059296     2.34   0.019      1.00225    1.025494
     age_don |   .9906357   .0037988    -2.45   0.014     .9832182    .9981092
         age |   .9483539   .0084602    -5.94   0.000     .9319163    .9650814
      prevtx |   1.463773   .1797112     3.10   0.002      1.15072    1.861993
      txtype |   1.124474   .1991213     0.66   0.508     .7947287    1.591035
             |
      hlamat |
          1  |   1.048589   .1855962     0.27   0.789     .7412181    1.483422
          2  |    .835029   .1530105    -0.98   0.325     .5830792    1.195847
          3  |   .7858572   .1462116    -1.30   0.195     .5457258    1.131652
          4  |   .6671839   .1515587    -1.78   0.075     .4274497    1.041372
          5  |   .8654381    .255724    -0.49   0.625     .4849743    1.544377
          6  |    .858816   .2780721    -0.47   0.638     .4552957    1.619969
             |
       _cons |   .1043656   .0233851   -10.09   0.000     .0672712    .1619143
------------------------------------------------------------------------------
```


mi extract

- Useful command
- Analysis not supported in mi estimate:
- Use "mi extract, `k'" looped for k
- Save results
- Analyze using published formulae
  not optimal, but possible

How Many Imputations

- Feasibly many:
  “50% missingness 5 draws is highly efficient"
- True but gives poor standard errors
  better guideline 20-40
- Chained Imputation on UNOS data
  - 20 imputations in 2 min.
  - 50 imputations in 5 min.
  - 500 imputations in 50 min.

50 Imputations

```
. mi estimate, eform: logistic death cold_isc age_don age  prevtx txtype i.hlamat

Multiple-imputation estimates                    Imputations     =         50
Logistic regression                              Number of obs   =       9775
                                                 Average RVI     =     0.0371
                                                 Largest FMI     =     0.1946
DF adjustment:   Large sample                    DF:     min     =    1310.53
                                                         avg     =  894264.52
                                                         max     = 5923449.56
Model F test:       Equal FMI                    F(  11,351411.5)=       9.02
Within VCE type:          OIM                    Prob > F        =     0.0000

------------------------------------------------------------------------------
       death |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.009856   .0066926     1.48   0.139     .9968112    1.023071
     age_don |   .9905272   .0038156    -2.47   0.013      .983077    .9980339
         age |   .9482437   .0084732    -5.95   0.000     .9317811    .9649971
      prevtx |   1.376873   .1752271     2.51   0.012     1.072907    1.766954
      txtype |   1.237584   .2299089     1.15   0.251     .8598254    1.781308
             |
      hlamat |
          1  |   1.048506   .1886219     0.26   0.792     .7369546    1.491768
          2  |   .8519965   .1592486    -0.86   0.391     .5906521    1.228977
          3  |   .8045399   .1525117    -1.15   0.251     .5548634    1.166565
          4  |   .6962644   .1608806    -1.57   0.117     .4426748    1.095125
          5  |   .8605673   .2608053    -0.50   0.620     .4751321    1.558674
          6  |   .9041763   .2940166    -0.31   0.757     .4780357    1.710196
             |
```

500 Imputations

```
. mi estimate, eform: logistic death cold_isc age_don age  prevtx txtype i.hlamat

Multiple-imputation estimates                    Imputations     =        500
Logistic regression                              Number of obs   =       9775
                                                 Average RVI     =     0.0381
                                                 Largest FMI     =     0.1853
DF adjustment:   Large sample                    DF:     min     =   14556.70
                                                         avg     =   1.19e+07
                                                         max     =   7.27e+07
Model F test:       Equal FMI                    F(  11, 3.4e+06)=       9.02
Within VCE type:          OIM                    Prob > F        =     0.0000

------------------------------------------------------------------------------
       death |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    cold_isc |   1.010071   .0066465     1.52   0.128     .9971269    1.023184
     age_don |   .9905008   .0038143    -2.48   0.013     .9830531    .9980049
         age |   .9482384   .0084729    -5.95   0.000     .9317765    .9649913
      prevtx |   1.373944   .1765813     2.47   0.013     1.067999    1.767533
      txtype |     1.2328   .2287713     1.13   0.259      .856902    1.773593
             |
      hlamat |
          1  |   1.049783   .1892282     0.27   0.788     .7373369    1.494627
          2  |   .8522425   .1586011    -0.86   0.390     .5917734    1.227357
          3  |   .8053358   .1526196    -1.14   0.253     .5554756    1.167587
          4  |   .6966721    .159863    -1.58   0.115       .44433    1.092323
          5  |   .8545295   .2592478    -0.52   0.604     .4715075    1.548694
          6  |   .9090107   .2962998    -0.29   0.770     .4798573    1.721971
```


Approach to Imputation Model

- Inclusivity over parsimony
  more predictor: MAR more plausible
- Will remove biases -- primary objective
- Two things to include: outcome, predictor in your analytic model
- Functional form tends not to be critical

Making Up Data?

- No.
- Making a model for missingness
- Making a guess at missing data
- Fully representing uncertainty
  - this is the key to it's validity

Maximum Likelihood (ML)

Based on implication #1

- Maximum likelihood valid for MAR data
  as long as CMAR variables incorporated
- Works especially well for missing outcome
  especially in longitudinal analysis
  where attrition depends on previous values
- ML often fit with the EM algorithm
  EM also iterative fills in missing data
- Imputation & analysis models consistent
  MI and maximum like very similar

Disadvantages of ML

- Cumbersome for missing predictors
  don't usual model their distribution
- Awkward if CMAR variables are not a part of desired
- Lacks flexibility of MI

Inverse Weighting

Based on implication #1

- Model pr(non missingness|C_{MAR})
  weight by chance of being observed
- Two step approach
  not implemented in software
  easy to mess up (my experience)
- Clear connection to complete data analysis
- Inefficient

3 Methods

- Same MAR assumption
- Inverse weight:
  model missingness, data model flexible
- MI/ML:
  model data, missingness model unspecified