SENSITIVITY AND SOFTWARE

Don McLeish and Cyntha Struthers

University of Waterloo

December 5, 2015

Question: www.socrative.com (calgary)

Which of the following statements are true?

A. When the response Y is MAR, then the conditional distribution of Y given $R = 0$ is identifiable from the observed data.

B. When the response Y is MNAR, then the conditional distribution of Y given $R = 0$ is identifiable from the observed data.

C. When the response Y is MAR, then the joint likelihood for (R, Y_{obs}) is proportional to the likelihood for Y_{obs}

D. When the response Y is MCAR, then the joint likelihood for (R, Y_{obs}) is proportional to the likelihood for Y_{obs}

E. Prior to using the MAR assumption, you should use a formal test of its validity.

Sensitivity[2] analysis

Why be concerned about possible failure in the model[1] for missingness (e.g. MAR)?

- Statisticians prefer models that allow some test of fit.
- The MAR assumption implies that $P(*|R=1)$ and $P(*|R=0)$ are the same.
- This assumption allows properties under $P(*|R=0)$ to be estimated based on the data observed for $R=1$.
- Since there are NO observed data for $R=0$, the equality of $P(*|R=1)$ and $P(*|R=0)$ cannot be tested.

---

[1] "What, me worry?" the famous non-robust statistician, Alfred E. Newman

[2] "TO BOLDLY GO WHERE NO MAN HAS GONE BEFORE..." (only women)

Nonignorable missing data models: R and Y till death do they part...

- MAR implies that the likelihood for the observed data factors:

$$P(R, Y_{obs}|\theta) = P(R|Y_{obs})P(Y_{obs}|\theta)$$

- $P(Y_{obs}|\theta)$ is the "likelihood ignoring the missing data mechanism".
- The observed data likelihood $P(Y_{obs}|\theta)$ is proportional to the joint likelihood $P(R, Y_{obs}|\theta)$ if and only if the data are MAR (includes MCAR).
- **Cure:** Collect sufficient covariates X related to missingness to obtain MAR: $P(R|Y) = P(R|Y_{obs})$.[3]

[3] The price is parsimony

MNAR models

- For a MNAR model, both R and Y_{obs} carry information about θ and we need to model them jointly.
- Possible models for modeling (R, Y) jointly are:
  - Selection models (model Y marginally)
  - Pattern mixture models (model R marginally)
- There is a one-to-one relationship between these two models.

Selection models

In a selection model the probability the data are observed is assumed to depend on the value of the response Y:

$$P(R_i, Y_i|\theta) = P(R_i|Y_i, \gamma)P(Y_i|\theta)$$

(Heckman (1976)) where parameters θ and γ are distinct.

- The response Y_i is generated from the model $P(Y_i|\theta)$ (marginal distribution of Y_i).
- The i'th observation is observed with probability $P(R_i|Y_i, \gamma)$ (conditional probability of observing the i'th observation) which may depend on the response.
- The survivorship distribution discussed in the previous session is a selection model.

Pattern mixture models

In a pattern mixture model the observed data and the unobserved are assumed to follow different distributions:

$$P(R_i, Y_i|\theta) = P(Y_i|R_i, \theta)P(R_i|\gamma)$$

(Glynn et al. (1986)) where parameters θ and γ are distinct.

- The i'th observation is observed with probability $P(R_i|\gamma)$ (marginal distribution of R_i).
- The response Y_i is generated from the model $P(Y_i|R_i, \theta)$ (probability distribution of Y_i for a given missing pattern).
- $P(Y) =$
  $P(Y|R = 1)\,P(R = 1) + P(Y|R = 0)\,P(R = 0)$
  is a mixture of $P(Y|R = 1)$ and $P(Y|R = 0)$.
- Pattern mixture models are useful for sensitivity analyses. (More on this shortly.)

Question: www.socrative.com (calgary)

In which of the following studies would you assume a MAR model, a Selection Model, or Pattern Mixture Model.?

A. Y is the white blood count of patients at a university clinic is collected periodically over a four-year period. Some patients drop out of the study.
B. Y is the blood pressure of patients undergoing treatment for high blood pressure over a 3 year period. Some patients drop out of the study.
C. A study asks individuals about whether they use illegal drugs and Y =the amount used. Some study participants refuse to answer.
D. A Statistics Canada study of Y =foreign income uses Canada Revenue Agency data (anonymized). Many of the forms report zero foreign income.

Question: www.socrative.com (calgary)

In which of the following studies would you assume a MAR model, a Selection Model, or Pattern Mixture Model.?

E. In an HIV/AIDS study, participants are asked about $Y =$ the total number of sexual partners over a five year period. Some participants do not answer. (It is thought that women and men might report differently).

F. In an HIV/AIDS study participants are asked about the total number of sexual partners over a five year period. Some participants do not answer.(It is thought that non-response may be related to sexual orientation which was not recorded in the study).

Question: www.socrative.com (calgary)

In which of the following studies would you assume a MAR model, a Selection Model, or Pattern Mixture Model.?

G. In a poll, individuals with landline phones are asked about proposed CPP legislation. It is believed that more seniors have landline phones and agree to participate. (Age of the respondents is not recorded).

Sensitivity analysis

- Different MNAR models may fit the data equally well but may give different conclusions.
- As we have already noted, the observed data cannot be used to decide if a given MNAR mechanism is valid.
- One approach to this problem is to use a sensitivity analysis in which a number of plausible MNAR models are fitted to the data.
- The degree to which the inferences are stable across the models gives an indication of the confidence that can be placed in them.
- One approach to generating these MNAR models is to use pattern mixture models or selection models.
- We consider some illustrative examples.

Example: Pattern mixture models

Suppose Y_{1i} is always observed and Y_{2i} is sometimes missing. Let $R_i = 1$ if Y_{2i} is observed.

For a pattern mixture model, the likelihood of the observed data is:

$$\prod_{i=1}^{n} \left[f(y_{1i}, y_{2i} | R_i = 1, \theta) P(R_i = 1) \right]^{R_i}$$
$$\times \prod_{i=1}^{n} \left[f(y_{1i} | R_i = 0, \theta) P(R_i = 0) \right]^{1-R_i}$$

- There is essentially NO information on the conditional distribution $f(y_2|y_1, R = 0)$ without the MAR assumption.
- Under MAR, $f(y_2|y_1, R = 0) = f(y_1|y_2, R = 1)$ so we can estimate $f(y_2|y_1, R = 0)$ using the observed data.

Example of pattern mixture model: BVN

Y_{1i} is always observed and Y_{2i} is sometimes missing.

$R_i = 1$ if Y_{2i} is observed.

For $r = 0, 1$, suppose $(Y_1, Y_2)|R = r$ is BVN with mean

$$\mu^{(r)} = \left[\mu_1^{(r)}, \mu_2^{(r)}\right]$$

and variance

$$\Sigma^{(r)} = \begin{bmatrix} \sigma_{11}^{(r)} & \sigma_{12}^{(r)} \\ \sigma_{12}^{(r)} & \sigma_{22}^{(r)} \end{bmatrix}$$

- Let $\pi = P\left(\text{both } Y_1, Y_2 \text{ are observed}\right)$.
- There are 11 parameters in the model.

Example of pattern mixture model: BVN

- Assuming a pattern mixture model, the likelihood based on the observed data is

$$\prod_{i=1}^{n} \left[\pi f(y_{i1}, y_{i2}|R=1)\right]^{R_i} \left[(1-\pi) f(y_{i1}|R=0)\right]^{1-R_i}$$

  which is a function of 8 parameters

$$\mu_1^{(1)}, \mu_2^{(1)}, \sigma_{11}^{(1)}, \sigma_{12}^{(1)}, \sigma_{22}^{(1)}, \mu_1^{(0)}, \sigma_{11}^{(0)}, \pi$$

- These parameters are estimable based on the observed data.
- For example, $\hat{\mu}_1 = \bar{y}_1$ since Y_{1i} is always observed.
- How do we estimate the other 3 parameters: $\mu_2^{(0)}$, $\sigma_{12}^{(0)}$, and $\sigma_{22}^{(0)}$?

Example of pattern mixture model: BVN

- Since Y_2 is MAR, we have $P(R = 1|Y_1, Y_2) = P(R = 1|Y_1)$ (missingness depends only on observed quantities) so

$$f(Y_2|Y_1, R = 1) = \frac{f(Y_2, R = 1|Y_1)}{P(R = 1|Y_1)}$$
$$= \frac{P(R = 1|Y_1, Y_2)f(Y_2|Y_1)}{P(R = 1|Y_1)} = f(Y_2|Y_1).$$

- Similarly $f(Y_2|Y_1, R = 0) = f(Y_2|Y_1)$.
- In other words the unobserved conditional distribution $Y_2|Y_1, R = 0$ can be estimated from the observations $Y_2|Y_1, R = 1$.

Example of pattern mixture model: BVN

- For the BVN model $f(Y_2|Y_1, R=0)$ and $f(Y_2|Y_1, R=1)$ are regression models.
- Under the MAR assumption the parameters (slope, intercept, variance of error) are identical given $R=1$ or $R=0$.
- Therefore there are only 8 unknown parameters in this model.
- Under the pattern mixture model $\mu_2 = \pi\mu_2^{(1)} + (1-\pi)\mu_2^{(0)}$.
- How do we estimate μ_2?

Example of pattern mixture model: BVN

- Let $\hat{\mu}_2^{(1)}$ be the mean of the observed Y_{2i} for which $R_i = 1$.
- Use fully observed data (Y_{1i}, Y_{2i}) for which $R_i = 1$ to estimate β_0, β_1, the coefficients of the regression of Y_2 on Y_1.
- Let $\hat{\mu}_2^{(0)} = \hat{\beta}_0 + \hat{\beta}_1 \bar{y}_1^{(0)}$.
- Note that $\hat{\mu}_2^{(0)} = \hat{\beta}_0 + \hat{\beta}_1 \bar{y}_1^{(0)}$ is the average of the imputed Y_2 values (with $R_i = 0$) using regression imputation.
- The ML estimate of μ_2 is $\hat{\mu}_2 = \hat{\pi}\hat{\mu}_2^{(1)} + (1 - \hat{\pi})\hat{\mu}_2^{(0)}$.
- *Regression imputation provides the ML estimates of the means in Normal models.*

Example of pattern mixture model: BVN

- Suppose the missingness of Y_2 depends on Y_2 but not on Y_1, so Y_2 is NMAR.
- As before, $f\left(Y_1 | Y_2, R=0\right) = f\left(Y_1 | Y_2, R=1\right)$ so there are only 8 parameters, all estimable.
- For the BVN model the conditionals are regression models. For $R=1$ data, regress Y_1 on Y_2 to get regression coefficients $\hat{\beta}_0, \hat{\beta}_1$.
- Since $Y_1 = \hat{\beta}_0 + \hat{\beta}_1 Y_2 + \varepsilon$ for both $R=1$ and $R=0$, we can write the ML estimates as $\hat{\mu}_1^{(1)} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_2^{(1)}$ and $\hat{\mu}_1^{(0)} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}_2^{(0)}$ or $\hat{\mu}_2^{(0)} = \frac{\hat{\mu}_1^{(0)} - \hat{\beta}_0}{\hat{\beta}_1}$ from which we obtain $\hat{\mu}_2 = \hat{\pi}\hat{\mu}_2^{(1)} + \left(1-\hat{\pi}\right)\hat{\mu}_2^{(0)}$.

Sensitivity analysis of brain weight data - MNAR

- Recall the brain weight data that we analyzed previously: $Y_2 = \log(\text{brain weight})$ and $Y_1 = \log(\text{head size})$.
- Y_2 was sometimes missing.
- Suppose Y_2 is MNAR.
- For example, suppose that the mean of the distribution of brainweights for $R = 1$ (observed) is larger (smaller) than the mean of the distribution of brainweights for $R = 0$ (not observed).
- To model this we could assume a pattern mixture in which $E(Y_2|R = 0) = E(Y_2|R = 1) + \Delta$.

Sensitivity analysis of brain weight data - MNAR

- Suppose we assume a pattern mixture model with only the mean changing, that is,

$$E\left(Y_2|R=0\right) = E\left(Y_2|R=1\right) + \Delta$$

- $\Delta = 0$ corresponds to MAR.
- If a value of Δ is assumed then all the parameters are estimable.
- A sensitivity analysis can be done by fitting the model for a range of Δ values (including $\Delta = 0$).
- Interested in how the conclusions change as Δ varies.

Sensitivity analysis of brain weight data: MNAR

- If multiple imputation is used then a sensitivity analysis can be done by **adding** a constant Δ to every imputed value and redoing the analysis.
- Usually more than one value of Δ is used (positive and negative values).
- Values of Δ are chosen which are plausible for the given context.
- Interested in how the conclusions change as Δ varies.

Original analysis of brain weight data Δ=0 (MAR)

Analyze the brainweight data assuming MAR.
Y_2 = log(brain weight), Y_1 = log(head size).
Use multiple imputations and software MICE.

```
imp<- mice(brainweight,method="norm",m=5,maxit=1,seed=1)
fit<- with(imp,lm(lweight~lsize+sex))
round(summary(pool(fit)),3)
```

	est	se	t	df	Pr(>|t|)	nmis
(Intercept)	1.577	0.531	2.971	14.979	0.010	NA
log(headsize)	0.679	0.064	10.592	15.034	0	130
sex	-0.003	0.013	-0.250	16.219	0.805	0

Sensitivity analysis of brain weight data $\Delta = +0.1$

```
d <- 0.1
imp$imp$lsize <- imp$imp$lsize +d
fit2<- with(imp,lm(lweight~lsize+sex))
round(summary(pool(fit2)),3)
```

	est	se	t	df	Pr(>|t|)	nmis
(Intercept)	3.420	0.585	5.844	19.664	0	NA
log(headsize)	0.454	0.070	6.440	19.575	0	130
sex	-0.040	0.014	-2.917	29.515	0.007	0

Does this materially change the conclusions?

"*sex*" coefficient is now significant.

Other conclusions are similar.

Sensitivity analysis of brain weight data Δ = -0.1

```
d <- -0.1
imp$imp$lsize <- imp$imp$lsize +d
fit3<- with(imp,lm(lweight~lsize+fem))
round(summary(pool(fit3)),3)
```

	est	se	t	df	Pr(>|t|)	nmis
(Intercept)	2.711	0.408	6.647	15.961	0	NA
log(headsize)	0.545	0.050	11.006	16.074	0	130
sex	-0.010	0.011	-0.937	35.257	0.355	0

Conclusions are similar to original analysis.

Sensitivity analysis: More general pattern mixture models

For the pattern mixture model

$$E\left(Y_2|R=0\right) = E\left(Y_2|R=1\right) + \Delta$$

a sensitivity analysis can done by **adding** a constant Δ to every imputed value. An equivalent model is

$$E\left(Y_2|Y_1, R=0\right) = E\left(Y_2|Y_1, R=1\right) + \Delta.$$

1. We might wish the shift to depend on Y_1, for example, $E\left(Y_2|Y_1, R=0\right) = E\left(Y_2|Y_1, R=1\right) + d\left(\Delta, Y_1\right)$ where d might be a linear function of Y_1.
2. We might also shift the data using a scale determined by some link function g, that is, $E\left(Y_2|R=0\right) = g^{-1}\left[g(E\left(Y_2|R=1\right)) + \Delta\right]$.
3. We could also combine both 1 and 2.

Sensitivity analysis: Selection Model

- In a selection model the probability the data are observed is assumed to depend on the value of the response:

$$P(R, Y|\theta) = P(R|Y, \gamma)P(Y|\theta)$$

- For example for the brain weight data we might assume

$$\text{logit}\left(P\left[R = 0|\, Y_2 = y_2\right]\right) = \gamma_0 + \gamma_1 y_2$$

- If a value of γ_1 is assumed then estimation is possible.
- A sensitivity analysis can be done by fitting the model for a range of γ_1 values (including $\gamma_1 = 0$).
- Interested in how the conclusions change as γ_1 varies.

Monotone dropout and sensitivity analysis - Diggle and Kenward(1994)

For longitudinal data recall that we have monotone dropout if y_k missing implies $y_{k+1}, y_{k+2}, \ldots$ are missing.

- Full data model: $Y_i | x_i \backsim MVN\left(x_i\beta, \Sigma\right)$
- Model for dropout:

$$\log it\left(\frac{p_k}{1-p_k}\right) = \gamma_0 + \gamma_1 y_k + \gamma_2 y_{k-1}$$

  where $p_k = P\left(\text{dropout at time } k\right)$.
- $\gamma_1 = 0$ corresponds to random dropouts (MAR)
- $\gamma_1 = \gamma_2 = 0$ corresponds to completely random dropouts (MCAR).
- Models can be fit using ML and likelihood ratio tests can be used to test hypotheses: $\gamma_0 = 0$ and $\gamma_1 = \gamma_2 = 0$.

National Research Council Special Report on Missing Data in Clinical Trials

- "There is no easy fix for missing data at the analysis stage. Too many current analyses of clinical trials apply naive methods for missing data adjustment that make unjustifiable assumptions, such as last-observation carried-forward approach."
- Sensitivity analysis is a relatively undeveloped area of statistical analysis and at the moment there are no clear guidelines for defining the appropriate sensitivity analysis. (See: Ware et al. (2012)).

National Research Council Special Report on Missing Data in Clinical Trials

- **Recommendation 15 of the N.R.C. Special Report:** Sensitivity analysis should be part of the primary reporting of findings from clinical trials. Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting.
- "It is important that additional research be carried out so that methods to carry out sensitivity analyses for all of the standard models are available."

R software MICE (Multivariate Imputation by Chained Equations)

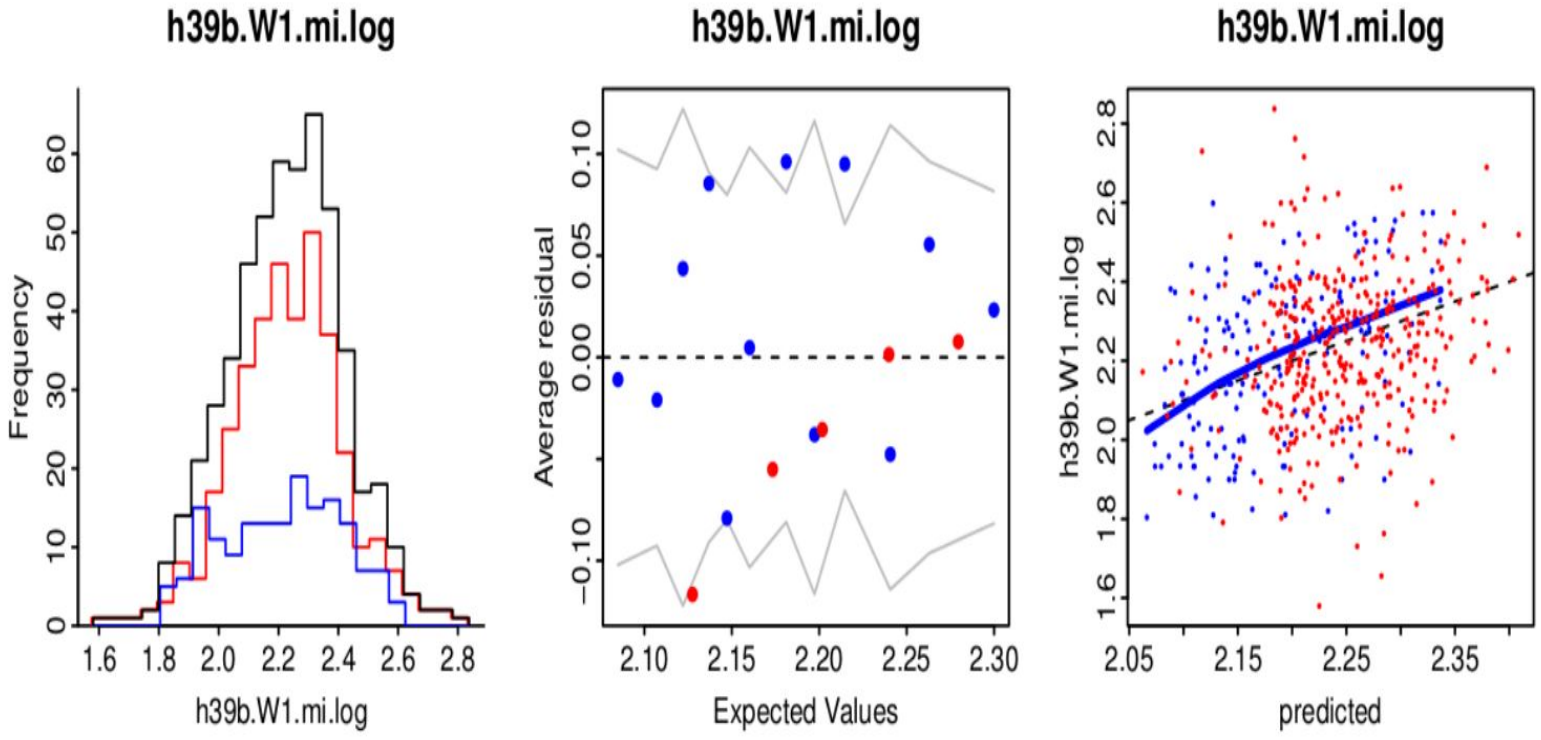
The R package **mice** (Version 2.12, 2015-02-20) imputes incomplete multivariate data by chained equations. The software includes automatic predictor selection, passive imputation, post-processing of imputed values, specialized pooling routines, model selection tools, and diagnostic graphs.

- See *Flexible Imputation of Missing Data* by Stef van Buuren (2012). Book has pseudo-code for many of the algorithms implemented in **MICE** so you can understand what the algorithm does. (useful for writing your own code!)
- van Buuren and Groothuis-Oudshoorn (2011), *MICE: Multivariate Imputation by Chained Equations in R* (Issue 45 of Journal of Statistical Software is devoted to multiple imputation.)

R software mi version 1.0, 2015-04-16: Missing data imputation and model checking

- The **mi** package performs multiple imputation for data with missing values. Iteratively draws imputed values from the conditional distribution for each variable, given observed and imputed values of the other variables. Approximates a Bayesian solution; multiple chains, convergence assessed after a pre-specified number of iterations.
- Allows customization of conditional models and missing values for each variable. Provides graphics to visualize missing data patterns, diagnose the models used to generate the imputations, and to assess convergence. (See: Su et al. (2011).)
- Functions included to run statistical models post-imputation.

Checking the imputation using mi



from Su et. al. (2011)

R software implementing algorithms in Schafer

- **norm** 1.0-9.5: Based on the software **NORM** by Schafer (1999) (available, free). Implements *multiple imputation based on the multivariate normal model* (Chapters 5&6). Includes data augmentation (DA) and MLE for incomplete data using the EM algorithm.
- **cat** 0.0-6.5: Implements *multiple imputation of categorical data* according to log-linear model - Chapters 7&8. MCMC method for simulating posterior draws under a hierarchical loglinear model.
- **pan** 1.3: Multiple imputation for multivariate panel or clustered data. Includes functions for *MLE for generalized linear mixed models* and *imputation of multivariate panel or cluster data*.

Software in SAS

- **PROC MI** (V9 onwards): Includes an implementation of Schafer's **NORM** (*multiple imputation based on the multivariate normal model*).
- **IVEware**: Multiple imputation using chained equations (fully conditional specification models).
- **PROC CALIS**: Fits ML to the observed data if **method(mlmv)** is specified. Assumes data are multivariate normal and MAR.

Software in Stata

- **mi**: Includes an implementation of Schafer's **NORM** as well as many other multiple imputation routines
- **ice**: package by Patrick Royston which implements multiple imputation by chained equations
- **sem**: Fits ML to the observed data if **METHOD=FIML** (full information maximum likelihood) is specified. Assumes data are multivariate normal and MAR.

Splus

The **S-PLUS** library **S+MissingData** is the most extensive implementation of techniques described in Schafer's book.

The library has functions to fit the multivariate normal, log-linear and general location models using EM algorithm and DA. The DA algorithms also produce multiple imputations.

Mplus

Mplus Version 7: "Mplus offers researchers a wide choice of models, estimators, and algorithms in a program that has an easy-to-use interface and graphical displays of data and analysis results. The Mplus modeling framework draws on the unifying theme of latent variables. The generality of the Mplus modeling framework comes from the unique use of both continuous and categorical latent variables."

MATLAB:

ecmmvnrmle: Multivariate normal regression with missing data

ecmnmle: Mean and covariance of incomplete multivariate normal data

ecmnstd: Standard errors for mean and covariance of incomplete data

knnimpute replaces NaNs in DATA with the corresponding value from the nearest-neighbor column using Euclidean distance. If the nearest neighbor column also contains a NaN value, then the next nearest column is used.

The Prevention and Treatment of Missing Data in Clinical Trials: Limiting Missing Data

- Target a population that is not adequately served by current treatments and hence has an incentive to remain in the study.
- Include a run-in period in which all patients are assigned to the active treatment, after which only those who tolerated and adhered to the therapy undergo randomization.
- Allow a flexible treatment regimen that accommodates individual differences in efficacy and side effects in order to reduce the dropout rate.
- Shorten the follow-up period for the primary outcome.
- Avoid outcome measures that are likely to lead to substantial missing data.

The Prevention and Treatment of Missing Data in Clinical Trials: Limiting Missing Data

- Select investigators who have a good track record.
- Set acceptable target rates for missing data and monitor the progress of the trial with respect to these targets.
- Provide incentives to investigators and participants for completeness of data collection.
- Limit the burden and inconvenience of data collection on the participants, and make the study experience as positive as possible.
- Provide continued access to effective treatments after the trial, before treatment approval.

The Prevention and Treatment of Missing Data in Clinical Trials: Limiting Missing Data

- Train investigators and study staff that keeping participants in the trial until the end is important, regardless of whether they continue to receive the assigned treatment. Convey this information to study participants.
- Collect information from participants regarding the likelihood that they will drop out, and use this information to attempt to reduce the incidence of dropout.
- Keep contact information for participants up to date.