Modelling Informative Dropout using NONMEM

Nick Holford
University of Auckland

General Principles (Abowd)

- The basic insight is that missing data should be modeled using the same probability and statistical tools that are the basis of all data analysis
- Missing data are not an anomaly to be swept under the carpet
- They are an integral part of every analysis

From John Abowd http://instruct1.cit.cornell.edu/courses/cis440/8 (no longer available)

Missing Data Mechanisms

Little & Rubin distinguish between:

- Missing Completely at Random (MCAR)
- Missing at Random (MAR)
- Not Missing at Random (NMAR)

Shafer's Taxonomy of Methods

Shafer has identified a series of methods that have been developed for analysing missing data. NONMEM provides a likelihood based approach.

Modelling Informative Missingness

Model for the Data (PKPD and Disease Progress)

Model for the Missingness Mechanism

= Maximum Likelihood Joint Model

This figure shows an example of viral load observations in patients with human immunodeficiency virus (HIV) infection. The left hand panel is for patients treated with amprenavir and the right hand panel for patients treated with indinavir.

A linear disease progress model might be used to describe the time course of viral load. The model for the data has two fixed effect parameters (intercept (a) and slope (b)). These parameters may be affected by a PKPD drug action model for the effect of treatment. Each parameter has an associated random effect to account for between subject variability. There is a residual error parameter to describe measurement error and model misspecification.
**Disease Progression and Last Observation Carried Forward**

Statistical Madness

- **LOCF**

- **Dropout**

- **S(t)**

- **Placebo**

- **Active**

- **T-test**

**Better Models for Missing Data**

**Hu & Sale Terminology**

Variables are dropout time $T$, observed ($Y_O$) and unobserved values ($Y_U$) of disease progress state (e.g. HIV viral load)

(a) **completely random** (CRD), if $T_i$ is independent of $\eta$, and therefore ($Y_O$, $Y_U$);

(b) **random** (RD), if $T_i$ (given $Y_O$) is independent of $Y_U$ but may depend on $Y_O$. In addition, any dependence of $T_i$ on $\eta$ is only through $Y_O$;

(c) **informative** (ID), if $T_i$ (given $Y_O$) depends on $Y_U$, or explicitly depends on $\eta$ other than through $Y_O$.

Hu & Sale have described how to model missing data and have proposed a modified terminology for missingness mechanism. The three categories are essentially the same as those defined by Little & Rubin but the names are a shorter.

They illustrate their terminology with the example of predicting the probability of a patient dropping out of a trial given the time of dropout and observed and unobserved values of viral load. The unobserved values are predicted from a PKPD and disease progress model.
Hazard, Risk, Survival, Dropout

\[ \text{hazard} = \beta_0 \cdot e^{\beta X} \]
\[ \text{Risk}_i = \int_0^t \text{hazard} \]
\[ \text{Survival}_i = e^{-\text{Risk}_i} \]
\[ \text{Dropout}_i = 1 - e^{-\text{Risk}_i} \]

The probability of a dropping out at a particular time can be predicted by describing the hazard for dropout. Hazard is the instantaneous risk of dropping out. As time passes the cumulative hazard predicts the risk of dropping out at a particular time.

The hazard model shown here is a proportional hazard model which is widely used for survival analysis. \( \beta_0 \) is the baseline hazard. \( \beta \) is a parameter describing how the hazard depends on some other value, \( X \).

The risk is the cumulative hazard. It is obtained by integrating hazard with respect to time.

The probability of survival (not dropping out) can be predicted from the risk by assuming an exponential distribution for not dropping out. The dropout probability can be predicted from the survival probability.

Using NONMEM nomenclature, \( DV \) is a the observed drop-out state. If it is 0 it means the subject has not dropped out. If it is 1 then the subject has dropped out at some time between the last observation when they had not dropped out and the current time when it is known they have dropped out.

The probability of a drop-out event is conditional on whether drop-out has not occurred at the time of the probability prediction. This model shows the probability of not dropping out depends on the hazard accumulated from time 0 until the time of the current observation given that the subject has not dropped out (\( DV=0 \)). In this example the hazard uses the predicted disease state value rather than the observed disease state value.

At the time a subject is known to have dropped out (\( DV=1 \)) the probability of this happening is the product of the probability they the subject had not dropped out at the time of the last observation (when it was known they had not dropped out) and the probability of dropping out at some time between the last observation and the current time when it known the subject has dropped out.
-2 Log Likelihood

Disease Progress

\[ ELS = \sum_{i=1}^{NSUB} \sum_{j=1}^{NOBS} \left( \frac{Y_{OBS_{ij}} - Y_{PRED_{ij}}}{\text{Var}_y} \right)^2 + \ln(\text{Var}_y) \]

\[ CCONTR_{-2LL} = \left( \frac{Y_{OBS_{ij}} - Y_{PRED_{ij}}}{SD_y} \right)^2 + 2 \cdot \ln(\text{SD}_y) \]

Dropout Probability

\[ CCONTR_{-2LL} = -2 \cdot \ln(\text{Like}) \]

Hu & Sale Code

$MODEL$ COMP=CUMHAZ ; compartment for integration of hazard
COMP=[HAZLAST, INITIALOFF] ; comp for LAST PERIOD hazard

INTERC=THETA(1) - THETA(2)*(TRT-1))+ETA(1)
SLOPE=THETA(3)+ETA(2)
BETA=THETA(4)
BET2=THETA(5)
SD2=SD2
VIRL=INTERC+SLOPE*(T-12)
TEMP=BETA*LOCF+BET2*VIRL
DADT(1)=EXP(TEMP)
DADT(2)=EXP(TEMP)

$ERROR$ CMHZ=BSHZ*A(1)
HZLA=BSHZ*A(2)
IF (DVID.EQ.1) THEN ; DV=Viral Load
IPRE=INTERC+SLOPE*(TIME-12)
Y=2*LOG(THETA(7))+( (DV-IPRE)/THETA(7) )**2
ENDIF
IF (DVID.EQ.2 .AND. DV.EQ.0) THEN ; NO dropout
Y=-2*(-CMHZ)
ENDIF
IF (DVID.EQ.2 .AND. DV.EQ.1) THEN ; dropout
Y=-2*(CMHZ-HZLA) - 2*LOG(1 - EXP(-HZLA))
ENDIF

Modifications of Hu & Sale Method

- User supplied $SUB CCONTR for joint ELS and LIKE contributions to objective function
- “Standard” code for disease progress prediction

Hu & Sale described a joint model for HIV viral load and drop-out by explicitly computing the -2 log likelihood contribution to the objective function. An alternative method is to use a user defined CCONTR subroutine which uses existing NONMEM routines to compute the likelihood. This alternative method may be simpler to use because it uses ‘standard’ code for the disease progress model prediction.
Modified Code

This illustrates how to joint modelling for all 3 models for missingness. The CRD model is coded by fixing THETA(2) and THETA(3) to zero ($THETA, $OMEGA and $SIGMA records are not shown here) The RD model is coded by fixing THETA(2) to zero and estimating THETA(1). It uses the LOCF value to predict the hazard during the unobserved period after the last known observation until drop-out is known. The ID model is coded by fixing THETA(1) to zero. It uses the predicted value of viral load to predict the drop-out hazard. The $CONTR record identifies the DVID data item to inform the CCONTR subroutine whether an observation is a viral load value or a drop-out value. User supplied CONTR and CCONTR subroutines are defined in the $SUBROUTINE record.

Differential equations are used to integrate the hazard. Compartment 1 is used to integrate the cumulative hazard since the start of the study. Compartment 2 is used to integrate the hazard over the period from the last known observation until the time that it is known that dropout has occurred (at some unknown time since the last known observation). An effect of treatment is assumed to affect the intercept of the disease progress model.

The format of the user supplied CCONTR is shown here. It can be used quite generally. The main user specific feature is to use the value of TYPE (which is determined by a value in the data set) to choose which method should be used to compute the contribution to the likelihood.

### CCONTR

```fortran
SUBROUTINE CCONTR (ICALL,CNT,P1,P2,IER1,IER2
SAVE
C LVR and NO should match values in NSIZES
PARAMETER(LVR=30,NO=50)
COMMON /ROCM4/ Y(NO),DATA(NO,3)
DOUBLE PRECISION CNT,P1,P2,T
DIMENSION P1(*),P2(LVR,*)
TYPE=DATA(1,1)
C Value of TYPE is provided as a user defined data item
IF (TYPE.EQ.1) THEN
  CALL CELS(CNT,P1,P2,IER1,IER2)
ELSE
  CALL CLIK(ICALL,CNT,P1,P2,IER1,IER2)
ENDIF
RETURN
END
```

The format of the user supplied CCONTR is shown here. It can be used quite generally. The main user specific feature is to use the value of TYPE (which is determined by a value in the data set) to choose which method should be used to compute the contribution to the likelihood.

### CCONTR

```fortran
SUBROUTINE CCONTR (ICALL,CNT,P1,P2,IER1,IER2
SAVE
C LVR and NO should match values in NSIZES
PARAMETER(LVR=30,NO=50)
COMMON /ROCM4/ Y(NO),DATA(NO,3)
DOUBLE PRECISION CNT,P1,P2,T
DIMENSION P1(*),P2(LVR,*)
TYPE=DATA(1,1)
C Value of TYPE is provided as a user defined data item
IF (TYPE.EQ.1) THEN
  CALL CELS(CNT,P1,P2,IER1,IER2)
ELSE
  CALL CLIK(ICALL,CNT,P1,P2,IER1,IER2)
ENDIF
RETURN
END
```
### Slide 17

**Dropout Data Format**

<table>
<thead>
<tr>
<th>ID</th>
<th>TRT</th>
<th>TIME</th>
<th>CMT</th>
<th>LOCF</th>
<th>DV</th>
<th>MDV</th>
<th>DVID</th>
<th>EVID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-0.6</td>
<td>26.1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>-0.6</td>
<td>28.1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>28.1</td>
<td>53.2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>75</td>
<td>1</td>
<td>0</td>
<td>53.2</td>
<td>81.6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>108.7</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-0.1</td>
<td>28.8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>28.8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>28.8</td>
<td>53.2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>53.2</td>
<td>81.6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>81.6</td>
<td>108.7</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

ID 2: Dropout is between time 25 and 50. Dropout risk compartment is turned on at 25 (CMT=2).

The NONMEM dataset format is shown here. When DVID is 1 this means the DV observation is of the disease state (e.g. viral load). When DVID is 2 this means the DV observation is drop-out status (0=not dropped out, 1=dropped out)

The LOCF data item is used when testing the random dropout model.

The CMT data item is set to 2 at the time of the last observation prior to dropout. This turns on this compartment so that it accumulates the hazard of dropping out.

In this example the first subject did not drop out during the study and the final record has DV=0 to indicate this. The second subject dropped out between time 25 and 50. A special event record (EVID=2) is used to set the CMT variable to 2 and turn on the second compartment. The final record for this subject indicates that they were known to have dropped out by time 50.

### Slide 18

**LIKE vs -2LL Methods**

**ID Bias and Imprecision**

- **LIKE Success**
  - **TP**: 79%
  - **PPV**: 79%
  - **7 to 4 min**: 79%
  - **-2LL**: 44%
  - **Success**: 68%
  - **73 53 min**: 68%

<table>
<thead>
<tr>
<th>TIME</th>
<th>LIKE Bias</th>
<th>SUCCESS</th>
<th>TP</th>
<th>7 to 4 min</th>
<th>-2LL</th>
<th>SUCCESS</th>
<th>44%</th>
<th>73 53 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>1</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>2</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>3</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>4</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
<td>-0.1%</td>
</tr>
</tbody>
</table>

The influence of using the two different methods (CCONTR and -2LL) for computing the contribution to the likelihood was investigated by Monte Carlo simulation. The probability of being observed to have dropped out is shown using an informative dropout model in the graph. Overall about 50% of subjects are expected to drop-out during the time from 0 to 100.

A disease progress observation was made on up to 4 occasions after entry to the study. During simulation the ID model was used to predict if drop-out occurred between the last visit and the current visit. A linear disease progress model with fixed intercept of 0 was used to describe the disease progress state. Simulated data were then fitted to the ID model using NONMEM.

The CCONTR method was faster and slightly more NONMEM runs minimized successfully. The bias and imprecision of the parameter estimates was similar for both methods. There were major biases in the estimated baseline and informative dropout hazard parameters.
Hu and Sale investigated the three models for missingness using real data sets. They distinguished between the models by assuming the difference in -2 times the log likelihood was distributed according to the chi-square distribution. This assumption was tested by simulating data under the completely random dropout model and then fitting it with the same model and with the random and informative dropout models. The probability of being observed to have dropped out is shown using the CRD model in the graph. Overall about 50% of subjects are expected to drop out during the time from 0 to 100. The addition of one extra parameter with either the RD or ID models required a large change in objective function to reject the null with a 5% Type I error rate. Over 90% of runs minimized successfully with both the null model (CRD) and the alternate model (RD or ID).

Questions

- Is Missingness Informative?
  - Only one bit of information per subject

- Can NONMEM get the right answer?
  - LIKE method with CCONTR is OK
  - Direct coding of -2LL is biased

- Can we distinguish CR, RD and ID?
  - Randomization test shows ΔOBJ is approximately χ² distributed

Missiness models can be used to simulate drop-outs but the data from a drop-out is small and is unlikely to help identify the underlying disease progress model because each subject only provides one bit of information. NONMEM estimates of the drop-out hazard are biased and it seems unlikely that they are reliable for simulation the drop-out process. The very large change in objective function required to reject the null is unexpected and raises questions about the implementation of the model and NONMEMs methods.

More Material

From
John Abowd
Missing Data Mechanisms

- The complete data are defined as the matrix $Y$, a matrix of size $n \times K$.
- The pattern of missing data is summarized by a matrix of indicator variables $M$, a matrix of size $n \times K$.
- The data generating mechanism is summarized by the joint distribution of $Y$ and $M$.

$$m_{ij} = \begin{cases} 
0, & \text{if } y_{ij} \text{ is observed} \\
1, & \text{if } y_{ij} \text{ is missing}
\end{cases}$$

$$p(Y, M | \theta, \phi)$$

From John Abowd

Missing Completely at Random

- In this case the missing data mechanism does not depend upon the data $Y$.
- This case is called MCAR.

$$p(M | Y, \theta, \phi) = p(M | \phi)$$

From John Abowd

Missing at Random

- Partition $Y$ into observed and unobserved parts. $Y = (Y_{\text{obs}}, Y_{\text{mis}})$
- Missing at random means that the distribution of $M$ depends only on the observed parts of $Y$.
- Called MAR.
Not Missing at Random

- If the condition for MAR fails, then we say that the data are not missing at random, NMAR.
- E.g. dropout because of adverse effects or failure of drug to be effective

The Rubin and Little Taxonomy for Dealing with Missing Values

- Analysis of the complete records only
- Weighting procedures
- Imputation-based procedures
- Model-based procedures

Analysis of Complete Records Only

- Assumes that the data are MCAR.
- Only appropriate for small amounts of missing data.
- Used to be common in economics, less so in sociology.
- Now very rare.
Weighting Procedures

• Modify the design weights to correct for missing records.
• Provide an item weight (e.g., earnings and income weights in the CPS) that corrects for missing data on that variable.

Imputation-based Procedures

• Missing values are filled-in and the resulting “Completed” data are analyzed
  – Mean imputation
  – Regression imputation
• Some imputation procedures (e.g., Rubin’s multiple imputation) are really model-based procedures.

Model-based Procedures

• A probability model based on \( p(Y, M) \) forms the basis for the analysis.
• This probability model is used as the basis for estimation of parameters or effects of interest.
• Some general-purpose model-based procedures are designed to be combined with likelihood functions that are not specified in advance.
Little and Rubin's Principles

- Imputations should be
  - Conditioned on observed variables
  - Multivariate
  - Draws from a predictive distribution
- Single imputation methods do not provide a means to correct standard errors for estimation error.

From John Abowd

Hazard Models for Dropout

CRD: $h(t; Y_o, Y_u, \eta, \beta) = \beta_0$

RD: $h(t; Y_o, Y_u, \eta, \beta) = \beta_0 \exp(\beta_1 Y_i - 1)$, for $t_{i-1} \leq t < t_i$

ID: $h(t; Y_o, Y_u, \eta, \beta) = \beta_0 \exp(\beta_2 Y_u)$, for $t_{i-1} \leq t < t_i$

$Y_i$ means LOCF
$Y_u$ means disease progress model prediction

Three models for the hazard are shown. Each one corresponds to the missingness mechanism for dropout. CRD is completely random dropout. The hazard is constant and is not affected by any observed or unobserved value of disease state. RD is random dropout. The hazard depends on the last observed value of the disease state. The last value is carried forward to the current time and therefore it is called a LOCF or Last Observation Carried Forward value. ID is informative dropout. The hazard depends on the predicted value for the disease state.

A Special Case

The probability of dropping out may also be affected by the process of making a visit to see the clinician. The clinician may advise the patient to withdraw at the time of the visit. This might be modelled by having a third category for the drop-out variable. When DV is 2 this indicates the patient dropped out at that time. The probability of this happening is described by an additional probability parameter (THETA(PVISIT)).