ESTIMATING MEDICAL COSTS WHEN FOLLOW-UP DATA ARE INCOMPLETE

Henry Glick
University of Pennsylvania Health System
http://www.uphs.upenn.edu/dgimhsr

Forest Research Institute
Jersey City, New Jersey
September 9, 2005

INTRODUCTION

- Cost estimates are often based on accumulated costs over a certain duration of time
  - Incomplete cost data is common
  - Incorrect analysis of incomplete cost data can lead to biased analyses and incorrect inferences
- Common methods for imputation of outcomes, such as last value carried forward, can be highly unreliable when imputing costs

OUTLINE

- Which cost outcome should be analyzed?
- Review definitions of censoring
- Review common techniques for addressing censored data
- Provide recommendations

WHICH COST OUTCOME SHOULD BE ANALYZED?

- Most common candidates:
  - Total cost (sum of all costs during period of observation, e.g., average cost was $6000 during the 6 months of observation)
  - Average cost per period (e.g. per month) (sum of all costs during period of observation divided by periods of observation, e.g., the average cost was $1000 per month of observation)
  - Repeated measure of cost (direct analysis of monthly data, e.g., average costs were $3000, $1500, $750, $250, $250, $250 for the 6 months of observation and the average for the 6 months was $6000)

WHICH COST OUTCOME? (II)

- Not a lot of discussion in the literature, but....
- The measure that is best depends in part on:
  - presence or absence of censored and missing data
  - Whether the episode of care is time-limited or chronic
- Whether medical service use / cost has been sampled versus fully measured has more to do with how one weights participants’ responses than it does with which of the cost outcomes should be analyzed
  - e.g., if one collects service use / cost data once every 6 months, but asks about service use during the last month:
    * One needs to weight the response data so they reflect service use during the entire 6 months
    * Sampling strategy does not have implications for which of the 3 measures of cost one should use
REPEATED MEASURES ANALYSIS

- Advantages
  - Allows for unbiased analysis of cost data up to the point of censoring
    * Use of repeated measures is a common approach to addressing censoring (e.g., Lin 1997; Lin 2000; Bang and Tsiatis 2000)
    * "...we recommend that costs per patient in each of a number of time periods should always be recorded in trials with censoring..." (O’Hagan and Stevens 2004)
  - When the disease is not time-limited, information about the trajectory of costs over time provides additional information to decision makers

TOTAL COST VS AVERAGE COST PER PERIOD

- Either is reasonable to consider when there isn’t (substantial) censoring and the episode is time-limited
  - Presuming that one derives the average cost per period simply by dividing by a constant, there is little difference between the two
- If there is substantial censoring, one loses information by analysis of either outcome
- If service use / cost is not time-limited, one may provide an impression that use / cost may continue at current levels when a repeated measures analysis might provide information that says it will not

AVERAGE COST PER PERIOD AND CENSORING

- If length of follow-up differs among participants, use of the average cost per period (i.e., dividing total costs by the observed length of follow-up) to address this censoring will be unbiased only if:
  - Data are "missing completely at random" (see below)
  - There is no mortality
    * One will oversample data from those who died, because they are more likely to have complete follow-up, whereas others will have censored follow-up
  - Costs are constant over time
    * If costs vary with time, one will overweight earlier follow-up time and underweight later follow-up time
CENSORED COSTS

- Costs are censored when the cost data for a subject are incomplete
- Censored costs data often occur when follow-up is incomplete
  - Costs are based on an accumulation of resources used over time
  - If a subject is not followed for the entire period of interest, the resources used during the period where no information was gathered are not known
  - For these subjects costs are censored

RIGHT CENSORED DATA

- One can think of censored cost data in the way one thinks of censoring in survival analysis
- An observation on a variable is right censored if all you know about the variable is that it is greater than some value
  - In survival analysis, the censored variable is typically the time of occurrence for some event; cases are right censored because observation is terminated before the event occurs
  - In cost analysis, the censored variable is typically the sum of a flow of costs; cases are right censored because observation is terminated before the flow ends

EXAMPLES OF RIGHT CENSORING

- If the censored variable is the time from randomization until death (in years), you may know only that the variable is greater than 5 years, in which case the person’s death time is right censored at 5 years
- If the censored variable is costs after randomization until death (in $), you may know only that costs are greater than $20,000, in which case the person’s costs are right censored at $20,000

ISSUES TO CONSIDER BEFORE ADDRESSING CENSORED COSTS DATA

- What is the pattern of missing (i.e., which values are missing)?
- What is the reason the data are missing (i.e., what is the mechanism of missingness)

PATTERN OF MISSING

- Non-monotone pattern
  - Item non-response (refuse to answer particular question)
  - A single follow-up interview is skipped
PATTERN OF MISSING (II)

- Monotone pattern
  - Attrition from study or participants lost to follow-up
  - Rolling admission with fixed stopping date

<table>
<thead>
<tr>
<th>Time Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>e</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>r</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>v</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>a</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>t</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>i</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>o</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notation:
1 = alive and in study
0 = Dead
. = missing

REASON FOR MISSING DATA

- Why are the data missing?
  - Is missingness related to the study variables?
  - Example: Data may be censored due to termination of study

- Is attrition predictable by study variables?
  - Is attrition greater for participants who live further from hospital?
  - Are participants with a particular symptom more likely to drop-out?
TYPOLOGY OF CENSORING MECHANISMS

- Rubin has defined 3 types of censoring mechanisms
- Two of the mechanisms are referred to as ignorable, while one is referred to as non-ignorable
- Mechanisms are considered ignorable when the process which generates the missing data does not alter the medical costs between people who are observed and those who are not observed
  - Data may be “missing completely at random” (MCAR), in which case participants who have incomplete follow-up are similar in every way to those with complete follow-up, except for random differences
  - Data may be “missing at random” (MAR), in which case missingness is random within observable categories (i.e. predict missingness)
- The mechanism is non-ignorable if values of the missing observations are conditional on the mechanism that created the missing value
- The appropriate strategy for analyzing censored costs depends on mechanisms that give rise to missing data (Rubin, Biometrika, 1976)

EVALUATING THE MECHANISM OF MISSINGNESS

- Because there often are no data available after someone is censored - an exception might exist, for example, if costs are obtained from an administrative data base while preferences are obtained from patient interview - it is usually difficult to perform direct tests of the mechanism
- Sufficient, but not necessary, indirect tests exist:
  - For ruling out MCAR vs either MAR or nonignorable missingness
  - For ruling out MAR vs nonignorable missingness

JUDGING BETWEEN MCAR VS OTHER MECHANISMS

- If clinical and demographic characteristics of participants with censored data differ significantly from the characteristics of participants without censored data, one can reject an MCAR mechanism
- If costs of participants who eventually have censored data (e.g., during periods prior to their data being censored) differ significantly from the costs of those who never have censored data (during the same periods), one can reject an MAR mechanism
- Lack of evidence to reject MCAR differs from evidence that the mechanism is MCAR

JUDGING BETWEEN MAR VS NONIGNORABLE MISSINGNESS

- If, after controlling for the probability of being missing, costs of participants who eventually have censored data (e.g., during the periods prior to their data being censored) differ significantly from the costs of those who never have censored data, one can reject an MAR mechanism
- Lack of evidence to reject MAR differs from evidence that the mechanism is MAR

TENETS WHEN ANALYZING CENSORED DATA

- Use all available information
- Avoid bias
- Attain the greatest possible precision
- Report uncertainty accurately
GENERAL STRATEGIES FOR ANALYZING MISSING DATA

- All case analysis
- Completed case analysis
- Standard survival analysis (treat cost as the outcome of a survival model and use an indicator variable to represent censoring)
- Imputation
  - hot-deck imputation
  - last value carried forward
  - predicted mean
  - EM Algorithm
- Mixed models / Propensity-score weighted analysis

ALL CASE ANALYSIS
(i.e., Observed Costs)

- Use the sample mean of the observed costs from all study subjects (i.e., ignore missing cost data)
  - Sample mean is biased downward because of failure to account for costs incurred after censoring, but
  - Will not create too large an error, if number of observations with missing data is small and is balanced between the two groups
- Note: Although the direction of the bias is known for the group mean, it is unknown for the difference in means between the randomization groups

COMPLETE CASE ANALYSIS

- Evaluate data from complete cases only (i.e., discard data from all cases where the cost data is not complete)
- Advantages
  - Easy
  - Yields a rectangular file
  - Does not “invent” data
- Disadvantages
  - Sample mean biased toward the cost of the participants with shorter survival times, because longer survival times more likely to be censored
  - Inefficient
  - Discards data (from the incomplete cases): violates the tenet of using all available data
  - Complete cases are often a biased sample
- As with the all case analysis, the direction of bias is unknown for the difference in means between the randomization groups
STANDARD SURVIVAL ANALYSIS

- In this approach, costs are treated as potentially right-censored survival times by attaching the censoring indicator to the observed total cost
  - Usually invalid because - unlike time - participants usually do not accumulate costs with a common rate function
  - High cost participants may generate larger costs at both the survival time and the censoring time

IMPUTATION (I)

- Uses associations among observed variables and outcomes to predict (impute) a plausible value for each missing outcome
  - “Hot-Deck” Imputation
    * Imputes observation based on “donors” in dataset. Donors selected based on similarities in observed variables
  - Last Value Carried Forward
    * If costs in Month t were $500 and costs in month t+1 was missing then impute costs in month t+1 as equal to $500.
  - If MCAR, imputation could made using univariate model (i.e. group means)
  - Predictions often made using multivariable regression.
    * EM Algorithm preferable because it provides a maximization of the likelihood function:
      1. Estimate the missing values (E-step)
      2. Maximize the likelihood (M-step)
      3. Iterate until convergence

IMPUTATION (II)

- Advantages
  - Avoids bias in point estimates if:
    * Missing-data mechanism is “ignorable,” that is, after controlling for observables, the process that gives rise to missing data mirrors a random mechanism
    * The prediction model is valid
  - Yields a rectangular file
  - Retains all the data

- Disadvantages
  - Naive imputations (i.e. means) can create problems
  - Underestimates uncertainty since analysis assumes imputed values are known
    * Standard errors too small, p-values too significant, and confidence intervals that cover less than their nominal coverages
MULTIPLE IMPUTATION

- Used to overcome underestimation of uncertainty from the strategies above

- First proposed by Rubin (1977) and now software available for its application (Solas 2.0)
  - Free software also available from: [www.stat.psu.edu/~jls/misoftwa.html](http://www.stat.psu.edu/~jls/misoftwa.html) (Schafer, J.L. (2000) NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2.02 Software for Windows.)

- Works by imputing several values (M) for each missing value
  - The uncertainty between the estimates of the missing value is combined with the traditional measure of uncertainty to arrive at an overall standard error for the imputed variable

MIXED MODELS

- Often used for analysis of repeated measures in studies subject to attrition

- These models often assume the data that are available for a given subject are representative of that subject's deviation from the average trend lines that are observed for the whole sample

- Random-effects models for longitudinal data with maximum likelihood estimation provide valid inferences in the presence of ignorable nonresponse

- In another approach (Lin et al., 1996), the survival analysis is a Kaplan-Meier of survival. This method is a two-model method that also models costs and then combines the estimate of survival and costs

MIXED MODEL EXAMPLE


- Method proceeds in four steps:
  - Calculate probability of surviving at the beginning of the period (e.g., at the beginning of the month or year)
  - Calculate expected costs per period among those available for follow-up during the period (and alive)
  - Multiply the survival probability per period times the expected costs per period (i.e., the results from 2 steps above). The result is a set of expected costs during the periods which accounts for dropouts and death
  - Sum the per period products to obtain expected total cost

INTUITIONS

- The Lin Interval Method assumes MCAR
  - Assumes participants who are lost to follow-up have the same survival and costs as participants who are available for follow-up at the time when they dropped out of the study

- Lin Interval Method is similar to an imputation based on average costs per period, but it includes added information about the survival of missing subjects
EXTENSION OF LIN INTERVAL METHOD ASSUMING MAR P-MISSING METHOD

- Predict the probability that each subject will be missing (e.g., at the end of follow-up); divide the observations into subsets based on probability of being missing (i.e. propensity score); and repeat four steps for each probability group
  - Estimates of survival and cost can be made using multivariable models (e.g., failure time models and regression analysis)
- This model uses propensity scores to define the pattern of missingness strata
- This extension still assumes a high level of ignorable, MAR missingness
  - Participants who are lost to follow-up are assumed to have the same predicted survival and predicted costs as participants who have similar probabilities of being lost to follow-up, but who were available for follow-up

INVERSE PROBABILITY WEIGHTING

- Lin (2000) proposes weighted regression to address MAR censoring, with weights based on the inverse of the probability of being observed
  - Those who have a high probability of being observed and who actually are observed have little extra weight attached to their costs, because participants like them generally are all observed in the sample
  - Those with a low probability of being observed and who actually are observed have substantial extra weight attached to their costs, because participants like them generally were censored during follow-up

Schulman KA, et al. Results of the economic evaluation of the First study, a multinational prospective economic evaluation. IJTAHC. 1996.
**INVERSE PROBABILITY WEIGHTING**

- Addresses both censoring and missing data
- Procedure is conducted in two stages
  - Predict the probability of being observed for each period (commonly by use of time to event models or repeated measure logit models)
  - Use the inverse of the predicted probability of being observed in a weighted multivariable, repeated measures analysis of costs (e.g., weighted repeated GLM or OLS)
  - Resulting prediction of mean costs per period adjusts for MAR censoring

**STRENGTHS AND WEAKNESSES**

- Is appropriate for either right or left censoring of costs
- Advantage over straight imputation increases as survival probabilities drop
- Can apply multiple imputation techniques
- Development of extension is important because in many cases MAR is more plausible than MCAR

**AIDS EXAMPLE**

- Randomized trial comparing combination therapy vs. monotherapy in a population of AIDS patients
- 597 participants enrolled in two treatment groups (297 in monotherapy, 300 in combination therapy)
- At the end of 1 year, 56% (334 participants) were no longer available for observation
  - 54% (161 participants) in monotherapy group
  - 58% (173 participants) in combination therapy group

**COST DATA**

- Resource use (with the date of service) was recorded prospectively in the trial case report forms
- Costs were estimated by multiplying service use by country-specific unit cost estimates
- Costs were subdivided by month
- Incomplete follow-up resulted in a monotone pattern of unit-level missing data
- Participants who died during the interval do not have censored or missing data; their costs are known to be 0 for all months following death
FOUR METHODS FOR ADDRESSING MISSING COSTS

- We compared 4 methods for addressing missing cost data
  - Observed costs from all study participants (i.e., ignore the problem)
  - Observed costs from completed cases only
  - Lin Interval model (Lin et al. Biometrica [1997]) - MCAR
  - P-missing method (Extended Lin Interval model)
- The last two methods are implemented using both univariate models and Multivariable models

REASONS FOR DISCONTINUATION

<table>
<thead>
<tr>
<th>Reason</th>
<th>Combination</th>
<th>Monotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious Medical Event</td>
<td>31.4%</td>
<td>40.9%</td>
</tr>
<tr>
<td>Non-Serious Event</td>
<td>56.2%</td>
<td>46.5%</td>
</tr>
<tr>
<td>Lost to Follow-up or Other</td>
<td>12.4%</td>
<td>12.6%</td>
</tr>
</tbody>
</table>

See Tables at end of notes

COMPARISON OF RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>Comb</th>
<th>Monotherapy</th>
<th>Diff</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed costs</td>
<td>4855</td>
<td>3648</td>
<td>1207</td>
<td>275 to 2140</td>
</tr>
<tr>
<td>Complete cases</td>
<td>6234</td>
<td>3363</td>
<td>2871</td>
<td>1593 to 4150</td>
</tr>
<tr>
<td>Lin</td>
<td>7057</td>
<td>4752</td>
<td>2305</td>
<td>1067 to 3464</td>
</tr>
<tr>
<td>P-missing</td>
<td>7280</td>
<td>4240</td>
<td>2040</td>
<td>917 to 3433</td>
</tr>
</tbody>
</table>

CONCLUSIONS

- The magnitude of the cost estimates and of the difference in costs between the treatment groups varies depending on the technique chosen to analyze incomplete cost data
- Magnitude of the difference between methods grows
  - As censoring grows
  - As costs differ between subgroups
  - As subgroups are discriminated from each other
- Collect costs per participant in each of a number of time periods when substantial amounts of censoring are expected
  - Rule of thumb: can ignore censoring when 5%-10% of data are censored
- Avoid last value carried forward, complete case analysis, and observed cost methods because they are prone to bias
ADVICE

- Avoid missing values wherever possible, by taking energetic steps to retain subjects in the study
- Collect covariates that are useful for predicting missing values
- Use as much information as possible and apply missing data analysis method that minimizes bias

NEXT STEPS

- Address non-ignorable missingness
  - One should collect as much information as possible about the reasons for drop-out and attempt to incorporate this information into the analysis
  - One should consider the mechanism by which the effect of the study drug dissipates once, when a subject drops out of the study and stops taking the drug
    * Does the effect of the drug continue once the patient no longer complies with the treatment?
  - If the mechanism is unknown, analyses should be conducted to assess sensitivity of results to plausible alternative specifications of the mechanism

REFERENCES

Missing Data:

Rubin DB. Inference and missing data. Biometrika, 63, 581-592 (1976)


Models:


Non-ignorable missing:


Using observed costs within the small intervals:


Using censoring techniques for cost:


***reasons it doesn’t work:


Multiple Imputation:


Mixture Models:


---

### PROBABILITY OF MISSING SUBGROUPS

<table>
<thead>
<tr>
<th>Probab Category</th>
<th>Combination</th>
<th></th>
<th>Monotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% missing</td>
<td>Avg Cst / month</td>
<td>% missing</td>
</tr>
<tr>
<td>High</td>
<td>80</td>
<td>824</td>
<td>85</td>
</tr>
<tr>
<td>Medium</td>
<td>55</td>
<td>630</td>
<td>53</td>
</tr>
<tr>
<td>Low</td>
<td>38</td>
<td>426</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>561</td>
<td>54</td>
</tr>
</tbody>
</table>
## COMPARISON OF CALCULATION, LIN METHOD VS. P-MISSING METHOD

### COMBINATION THERAPY IN MONTH 12

| Probab Categ | Rand N | Obs N | Lin Method | | | | P-Missing | | | |
|-------------|--------|-------|------------|------------|------------|------------|------------|------------|------------|
|              |        |       | Weight     | Surv Cost  | Group Cost | Weight     | Surv Cost  | Group Cost |
| High         | 100    | 20    | 0.157      | 598        | 94         | 0.333      | 560        | 185        |
| Medium       | 100    | 45    | 0.354      | 485        | 172        | 0.333      | 489        | 161        |
| Low          | 100    | 62    | 0.488      | 659        | 321        | 0.333      | 675        | 223        |
| Total        | 297    | 136   | 1.000      | --         | 587        | 1.000      | --         | 569        |

### MONOTHERAPY IN MONTH 12

| Probab Categ | Rand N | Obs N | Lin Method | | | | P-Missing | | | |
|-------------|--------|-------|------------|------------|------------|------------|------------|------------|------------|
|              |        |       | Weight     | Surv Cost  | Group Cost | Weight     | Surv Cost  | Group Cost |
| High         | 99     | 15    | 0.110      | 731        | 81         | 0.333      | 725        | 242        |
| Medium       | 99     | 47    | 0.346      | 284        | 98         | 0.333      | 283        | 93         |
| Low          | 99     | 74    | 0.544      | 88         | 48         | 0.333      | 89         | 29         |
| Total        | 297    | 136   | 1.000      | --         | 227        | 1.000      | --         | 365        |