Handling with missing data in clinical trials for time-to-event variables

Giulia Tonini,PhD Menarini Ricerche, Florence, Italy
Simona Scartoni, Menarini Ricerche, Florence, Italy
Angela Capriati, MD, PhD, Menarini Ricerche, Florence, Italy

ABSTRACT

Missing data is often a major issue in clinical trials, especially when the outcome variables come from repeated assessments. In particular, time-to-event endpoints can be substantially affected by a too conservative treatment of missing data along the observation period. When neglected or not properly treated, missing data may bias the results, reduce power and lead to wrong study conclusions. The advantage of a more sophisticated statistical method versus the traditional clinical method, such as last observation carried forward (LOCF), is still under debate.

We compare the two methods in a clinical study testing the efficacy of an anti-arrhythmic agent versus placebo on the time to Atrial Fibrillation (AF) recurrence, where the maintenance of normal heart rhythm or the occurrence of the AF event was to be daily evaluated by trans-telephonic ECG recorded by the patients. A Cox model is applied for the comparison between treatments.

The dataset presents missing observations due to the fact that recording is missing or ECG is not assessable.

Moreover a simulation is performed to provide an additional example. Both methods for handling missing data are applied. Multiple imputation in SAS uses PROC MI. We examined results and possible problems arising from the fact that PROC MI implements methods which are not suitable for this kind of data.

INTRODUCTION

Missing data in the analysis of clinical trials is a major problem, usually caused by patients drop out before study completion. There are some cases when the problem of missing data arises from the nature of the variables observed. For example, in studies where a continuous monitoring is needed, missing data may arise from lack of compliance or device malfunctioning.

When a time-to-event variable is the primary outcome of a study, it is important to choose the best method to be sure to observe the event, especially when the event does not turn into a permanent condition, as can happen in time-to-death studies.

The International Conference for Harmonization (ICH) guidelines "E9" addresses the issue of missing data in Clinical Trials. In the guideline, it is stated that method for handling missing data has to be declared in the protocol, even if it can be refined in the Statistical Analysis Plan and during the blind review of the final dataset. It is in fact important to reconsider the problem when the amount of missing data is known. The advice coming from regulatory organizations (like EMA and FDA) is to consider different approach to address the problem as a sensitivity analysis.

This paper will focus on a study where the primary endpoint is time to first recurrence of atrial fibrillation. Monitoring of patients is performed using a device for ECG recording that is provided to patients. Since sinus rhythm can restore without medical intervention, it is possible that if monitoring is interrupted, the event is never detected.

In this case, missing data might occur as a result of different reasons, from patients drop out to poor quality of the recorded ECG. Moreover a missing observation at time t does not imply it will be missing at all subsequent times. This means that the pattern of missing data is not monotone.

Every individual is followed up from entry into the study until either event happening or study completion, which can occur before the individual experience a recurrence. Let T_i be the time at which follow up of patient ends, with $C_i =1$ indicating follow up ends at recurrence, while $C_i =0$ indicating that individual i exits the study without experiencing the recurrence (so that T_i is the censoring time). This means that both T_i and C_i are affected from missingness of data. In this particular case where the event is an atrial fibrillation episode, a missing observation means that we have no information whether or not the event happened despite of the fact that the observation is available at any following time.

This is a case of data Missing Completely At Random (MCAR), since the probability of an observation being missing is unrelated to the observed and unobserved data on that individual. We can write it down as $Pr(R_i|Y_i)= Pr(R_i)$ where R_i is the observation for individual i and Y_i is a set of fully observed variables. Since, when data are MCAR, the chance of the data being missing is unrelated to the values, the observed data are therefore representative of the population. However, relative to the data we intend to collect, information has being lost.

THE CASE STUDY

This study is designed as a double-blind, randomised, placebo-controlled, parallel-group study to be conducted in approximately 40 European study sites in Germany, Italy, United Kingdom and Spain.

Approximately 241 male or female patients aged 18 or older suffering from persistent AF suitable for electrical direct current cardioversion (DCC) has been randomised in this study.

For randomised patients, Visit 2 (Day 1) marks the start of the 16-week double-blind treatment period, lasting from Day 1 (the day of Randomisation/start of treatment) until Day 113 or until treatment discontinuation which is also to be considered as 'regular' discontinuation in case of AF recurrence requiring medical intervention.

The primary objective of the study is to assess the efficacy of Ranolazine administered as 3 different doses regimens versus placebo in the maintenance of sinus rhythm (time from randomisation to the first documented episode of AF recurrence) after electrical cardioversion in patients with non-permanent AF (defined as a continuous AF with a minimum duration of 7 days to a maximum of 6 months or requiring termination by cardioversion).

All patients randomised in the study have been provided with a transtelephonic electrocardiographic (TT-ECG) device and has been instructed in its use and to record electrocardiograms every day and whenever the patient experiences symptoms possibly related to a new episode of AF.

TT-ECGs are transmitted to an independent Central Core ECG Laboratory for evaluation. The Central Core ECG Laboratory reviews all electrocardiograms in blinded condition within 48 hours of having received them and informs the local Investigator immediately in case an episode of AF is detected (the relevant ECG are attached to the notification).

The primary efficacy endpoint is time (median; days) from randomisation (Day 1) to first documented AF recurrence. Documented recurrence is defined as AF detected on TT-ECGs by the Core Central Lab or on 12-Lead ECGs performed during a study visit (scheduled or unscheduled).

The dataset is composed by all TT-ECG together with 12-Lead ECG performed at study visits.

Missing data are summarized in Table 1. As, we can see, the most part of the patients have at least one day of missing data. The duration of an interval of missing observation varies from 1 to more than 20 consecutive days. It is more probable to have missing ECGs in the first part of the study period, when the patient has to get used to the TT-ECG device. In particular, 191 out of 240 randomized patients, have at least one missing ECG. Only 29 patients have an interval of missing ECGs longer than 7 days. Not assessable ECGs are considered as missing. Overall, 16% of ECGs to be nominally collected as per protocol are missing.

How to deal with missing data in such a dataset is still an open issue. It is clear that a complete case analysis is not applicable to our case, since only 49 out of 240 patients has complete data. Imputation is then necessary and the choice of plausible estimates is what differentiates the various imputation methods.

We propose here, two different approaches. The first one is a more basic and conservative approach, where a simple strategy inspired to Last Observation Carried Forward (LOCF) is applied. We then examine a Multiple Imputations method. In particular we consider models already implemented in SAS and whether these methods are suitable for this kind of data.

APPROACHES FOR HANDLING MISSING DATA

FIRST APPROACH

The basic approach adopted for handling missing data, is a conservative approach based on the Last Observation Carried Forward (LOCF) philosophy. LOCF is usually implemented when longitudinal measurements are observed for each patient. In its classical application it takes the last available response and substitutes the value into all subsequent missing values.

In our case study, we apply this idea to the ECG dataset. In particular, missing TT-ECG has been imputed according to the following rules:

- One or more consecutive TT-ECG are missing; however, the first successive available ECG (TT-ECG and/or 12-lead ECG) show sinus rhythm: in this case last observation carried forward (LOCF) is applied and sinus rhythm is considered to be maintained for the entire interval of missing TT-ECG recordings. The maximum duration of missing ECG-recordings (TT-ECG or 12-lead ECG) allowing imputation as described above is 7 days; otherwise data is censored at the time of the last ECG before the missing interval (TT-ECG or 12-Lead-ECG available).

- One or more consecutive TT-ECG are missing; however the first two successive available ECG (TT-ECG and/or 12-lead ECG) show atrial fibrillation: in this case it is assumed that the AF episode started on the day following last available ECG (TT-ECG or 12-Lead-ECG) recording.

We then obtain a complete dataset where a Cox proportional hazard model having treatment as major covariate is applied.

SECOND APPROACH

Multiple Imputation method has received a significant amount of attention in recent literature. The idea of multiple imputation first proposed by Rubin [Rubin, 1978], is to impute more than one value for the missing item. The advantage of multiple imputation is that it represents the uncertainty about which value to impute. This is opposed to the first approach which can lead to an underestimation of the variability [Rubin, 1991]. Multiple imputations can be implemented for either longitudinal measurements or a single response. The general strategy is to replace each missing data with a certain number of plausible values from an appropriate distribution. Imputing m values for each missing item, we obtain m different complete datasets. On each of these dataset the planned statistical analysis is applied. Results are then combined so that the final inferential result takes into account the uncertainty caused by the missing data, estimated from the variability of the m independent imputations. If Q_i is the estimate of the unknown parameter and vi is the variance associated to the i-th imputation, then the final estimate for Q is the mean of the m different estimates

$$\bar{Q} = \frac{1}{m} \sum^{m} Q_i$$

The variance of $\bar{Q}$ is the sum of a component within imputation variability and a component between imputations variability. Variance within imputation is $\bar{v} = \frac{1}{m} \sum_{i=1}^{m} v_i$ while between imputations variability is

$$B = \frac{1}{m-1} \sum_{i=1}^{m} (Q_i - \bar{Q})^2$$

Total variance associated to $\bar{Q}$ is then $V = \bar{v} + \left(1 + \frac{1}{m}\right) B$

The process is applied as follows:

•We impute m values for the missing item (obtaining m different data sets)

• We replace each missing value with more than one value from an appropriate distribution.

• We perform m different analyses on the m imputed datasets.

Results are then combined so that the final estimate takes into account uncertainty coming from missing data, which is estimated from the variability of the m independent outputs. The parameter of interest is estimated as well as its variance.

The number of imputations needed depends on the amount of missing information. Little and Schenker [Little et al., 1995] and Rubin [Rubin et al. 1991] indicate that 5 imputations are sufficient when the portion of missing data is around 30%.

In SAS, proc MI , together with proc MIANALYZE, can implement the previously described multiple imputation process. It uses methods that incorporate appropriate variability across the m imputa- tions. The method of choice depends on the patterns of missingness.

The most popular method used by PROC MI is the Monte Carlo Markov Chain (MCMC) algorithm. This is based on the assumption of multivariate normality [Schafer, 1997] which implies that valid imputations may be generated by linear regression equations. The algorithm can be applied for both monotone or non-monotone missing data patterns. Nevertheless, because it assumes normality and linearity, it may not be applicable for imputing categorical variables. One possible solution [Schafer, 1997; Allison et al., 2001], for example when imputing a binary variable, is to round the imputed values so that imputed values greater than 0.5 are set to 1 while anything less is set to 0. However such rounding can produce biased estimates of proportions, especially when the true proportion is near 0 or 1 [Horton et al., 2003]. A paper from Allison [Allison, 2005] illustrates the problem of bias in estimates, comparing MCMC with other methods of imputation based on logistic regression models or linear discriminant model. The problem remains open since, for example, in PROC MI, the logistic regression imputation is available only for monotone missing data patterns.

APPLICATION TO THE CASE STUDY

We considered the dataset from our case study, creating a dummy randomisation list. We added a treatment effect labelling trt=1 for placebo, and trt=2,3,4 for the three different doses of study drug. The highest effect is considered for the highest dose corresponding to trt=4.

We first applied the method following the LOCF approach. In Fig 1 we can see the Kaplan Meier curves on the dataset after correction for missing data, using this first approach. Kaplan-Meier curves for the dataset without any correction are reported in Fig 3. In this case, no correction for missing data means that for the missing days the patient is considered as in sinus rhythm. The correction is indeed conservative. For this reason, we obtain a median time to recurrence which is considerably lower than the median time obtained from the dataset without any correction for missing data.

Cox model shows a statistically significant treatment effect on time to recurrence, with a p-value<0.05.

We then applied the multiple imputations approach. The first difficulty is coming from the fact that the pattern of missing data is not monotone. In fact, considering longitudinal data, it is important to respect the time course of the data. Moreover, the variable we need to impute here is dichotomous (0=no event, 1=event). For data sets with monotone missing patterns, either a parametric regression method [Lavori et al., 1995] that assumes multivariate normality or a nonparametric method that uses propensity scores [Rubin, 1987; Lavori et al., 1995] is appropriate. For data sets with arbitrary missing patterns, a Markov Chain Monte Carlo (MCMC) method [Schafer, 1997] that assumes multivariate normality is used to impute all missing values or just enough missing values to make the imputed data sets have monotone missing patterns.

The procedure we followed to impute missing TT-ECG is the following:

•Proc MI applied to ECG dataset using the option *monotone logistic* obtaining 5 imputed datasets; Imputed values are 1 (event of AF) or 0 (sinus rhythm) for each day of the study follow up. Imputation comes from logistic regression.

•Time-to-event is calculated and Cox model is applied to each imputed dataset; coefficients estimates are obtained (Proc PHREG is applied "by imputation")

Results are combined using PROC MIANALYZE. The mean of the 5 estimates is considered as the result of the final analysis.

The only way to apply PROC MI to our dataset, is to build a monotone pattern of missing data. This implies a destruction of the temporal structure of data, which is instead the information we want to keep from our data.

In Fig 2, Kaplan Meier curves on one of the imputed dataset are shown. Cox model estimates on the imputed datasets shows a statistically significant effect of treatment (p-value<0.05) for four of the five imputed datasets. By the way, when results from the five different imputations are combined, statistical significance of treatment is lost because of the too high between imputation variability. This means that, when we obtain the estimates of the coefficients from the imputed datasets, these might not result to be statistically significant because of a too large variance. To better investigate this result, we applied both approaches to simulated datasets.

SIMULATIONS

A dataset simulation is done to obtain a more complete comparison of the 2 methods (LOCF and MI). This example does not intend to be an exhaustive simulation study but just to demonstrate potential disadvantages in the application of the two approaches. 1000 simulations have been run of a dataset for time-to-event data with missing data as follows. The simulated datasets are built in order to have similar characteristics of the case study dataset. 240 patients are considered with an amount of missing data similar to the one observed in our real dataset. Median time to recurrence has been estimated from real data and a dose-related treatment effect has been included. Missing data pattern is randomly created.

Both approaches are applied to the simulated datasets to impute missing data. After the completed datasets are created, the statistical analysis is performed. Treatment effect on time to recurrence is evaluated by a Cox model in both datasets obtained applied from the 2 methods for missing data (LOCF and MI).

RESULTS

When coefficients of the Cox model are calculated for each simulated dataset, statistically significant treatment effect (p-value<0.05) is shown by LOCF method for the 100% of the simulations. When the MI method is applied, a statistically significant p-value (<0.05) is reached only in the 40% of the simulated datasets (against the 99% with the LOCF approach). This means that the final coefficient estimates does not result to be statistically significant, even if the alternative hypothesis of a positive effect of treatment is true. The MI method generates many cases of false negative (60%) due to a large variability between imputations. The risk is then to not being able to demonstrate the efficacy of the treatment in cases where the treatment would have resulted to be statistically significant.

DISCUSSION

The first approach provides more conservative results. Although it does not take into account uncertainty generated by the imputation procedure, it is a reasonable method for handling missing data. The estimate of the median time to recurrence it is indeed lower than the one obtained from the Kaplan Meier on the incomplete dataset. The risk of this method is to reduce the median time to recurrence at a point where the significance of treatment effect cannot be proved. Since it is a quite conservative approach, this method is commonly accepted by regulatory organizations as a method for handling missing data.

 PROC MI in SAS does not provide a suitable method for our dataset. A possible solution might be to implement a model which imputes the median time to event calculating it from the imputed data at each step. Such a method can be implemented using other software like R. It is fundamental to preserve the time ordering of the variables, using the amount of observed information. Observations coming from the observed data before and after the missing item can be used as predictors. Further work is needed to understand how this approach can be implemented and applied.

CONCLUSIONS

The objective of this paper was to present an example of issues, concerns and available methodology for handling missing data in longitudinal studies where the outcome is time to event, with the particular difficulty that the event does not turn into a permanent condition. It is important to note that more than the statistical analysis, it is the method for monitoring patients which should be carefully examined in order to fine the way to reduce as much as possible the amount of missing data. A major part of the paper was dedicated to the method of multiple imputation, as it has achieved a large popularity in these last years. Our aim was to show that, despite of its popularity, this methodology still needs further development to cover all the possible type of data arising from a clinical study.

We observed that in our case study, MI can affect the statistical significance because of the large variability between imputations, which is not reduced when the number of imputations is increased (from 5 to 25). This might be caused by the fact that PROC MI does not have a method to impute dichotomous variable which does not have monotone pattern of missingness. Moreover, the covariates used to impute the data are not very predictive on the event. An approach considering the probability of having the event based on the hazard function is likely to be more appropriate.

Further work should be performed in order to better understand how well multiple imputation works under different types of missing data. It is very important to understand the limitations of each method and this is the reason why it is always useful to perform a sensitivity analysis, considering different approaches.

Fig 1 Kaplan-Meier curves for time to first AF recurrence in the dataset with missing data imputed using the first apporach
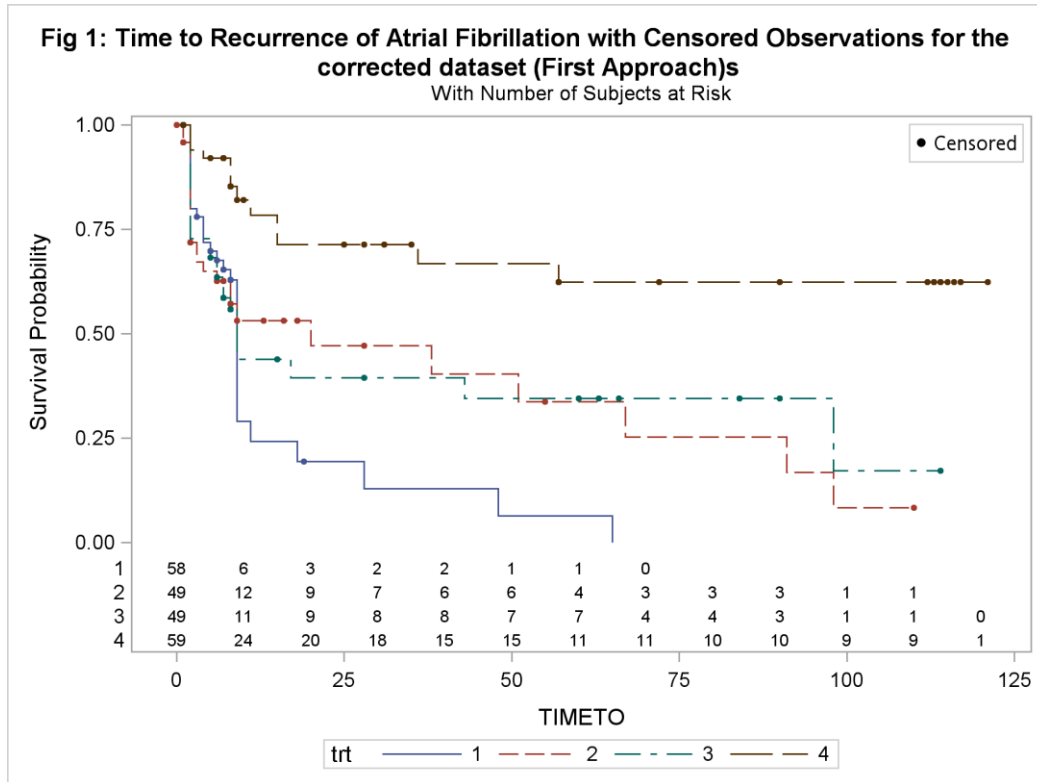

Fig 1: Time to Recurrence of Atrial Fibrillation with Censored Observations for the corrected dataset (First Approach)s

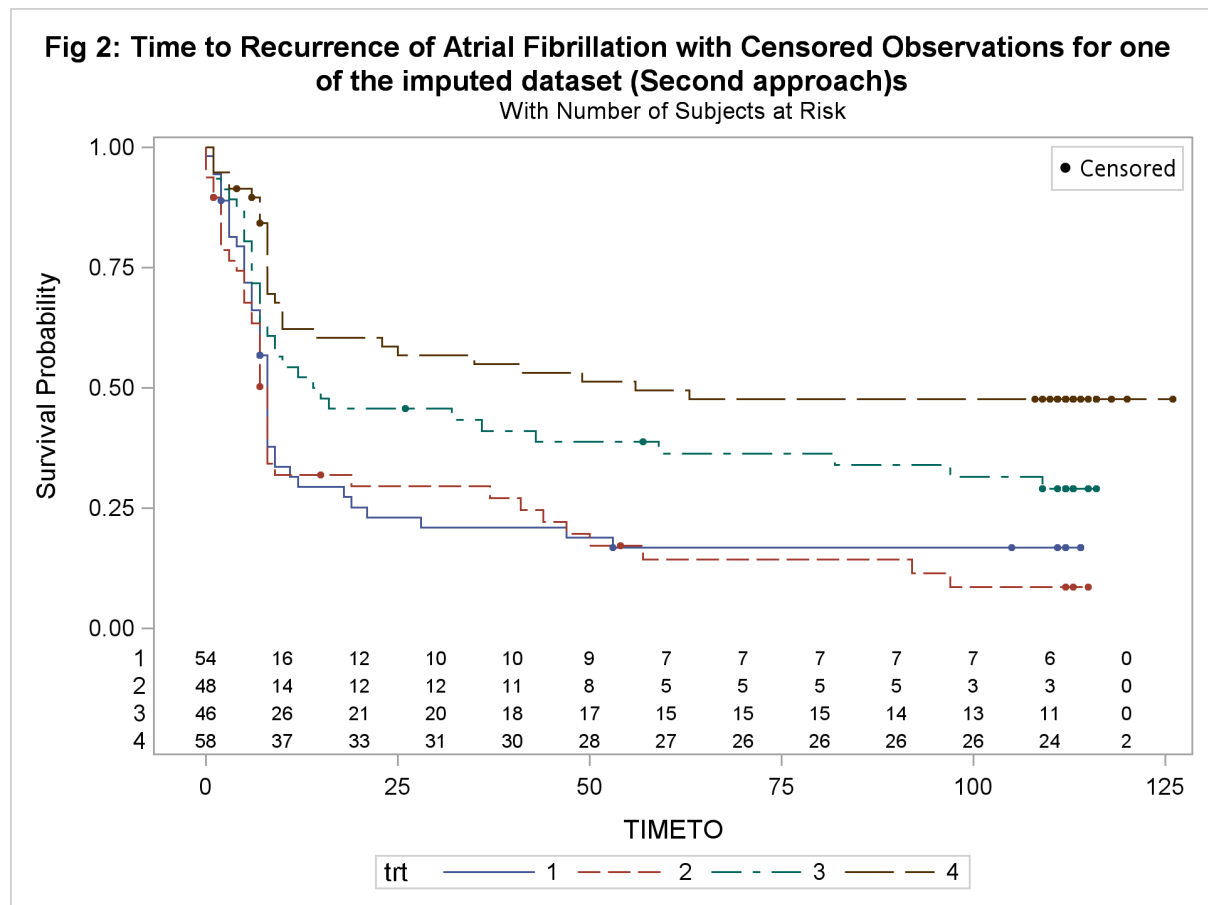**Fig 2 Kaplan-Meier curves for time to first AF recurrence for one of the imputed dataset using MI approach**

Fig 2: Time to Recurrence of Atrial Fibrillation with Censored Observations for one of the imputed dataset (Second approach)s

Fig 3 Kaplan-Meier curves for time to first AF recurrence in the dataset without any correction for missing data
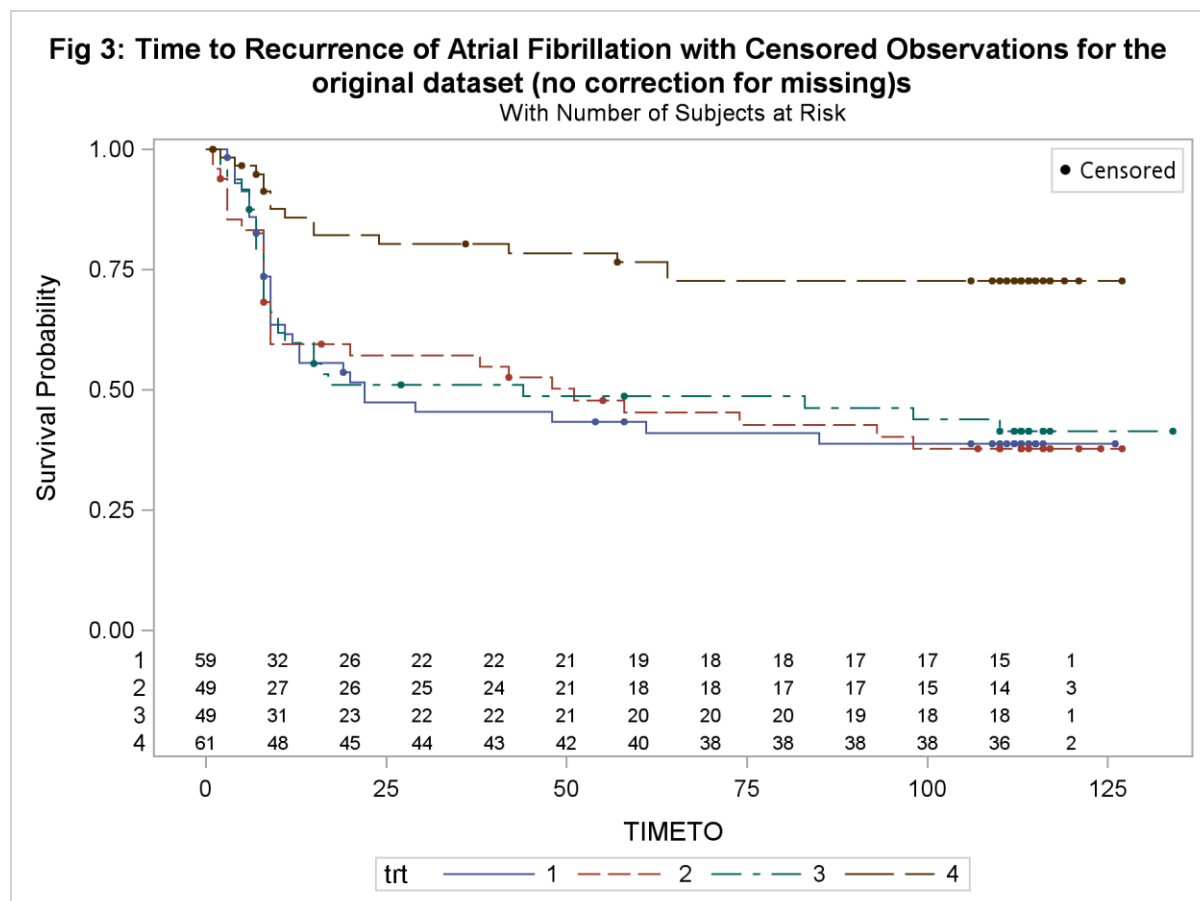

Fig 3: Time to Recurrence of Atrial Fibrillation with Censored Observations for the original dataset (no correction for missing)s
With Number of Subjects at Risk

Table 1

Table 1				
Interval of missing data along the study course				
	Month I	**Month II**	**Month III**	**Month IV**
1-3 days	332	241	266	261
4-7 days	52	28	28	38
>7 days	18	6	5	9
Patient with interval of missing data along the study course				
	Month I	**Month II**	**Month III**	**Month IV**
1-3 days	119	84	89	85

4-7 days	29	8	9	9
>7 days	1	2	2	3

REFERENCES

Allison, PD (2001) Missing Data. *Thousand oaks, CA: Sage Publications*.

Allison PD (2005) Imputation of categorical variables with PROC MI , paper 113-30, *SAS User Group International*, Philadephia, PA

Horton, NJ, Lipsitz SR, Parzen M, (2003) A potential for bias when rounding in multiple imputation. *American Statistician* 57: 229-232.

Lavori, P.W., Dawson, R., and Shera, D. (1995), A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data, *Statistics in Medicine*, 14, 1913–1925.

Little RJA, Schenker N. Handbook of statistical methodology – Missing data. *New York: Plenum Press*; 1995.

Rubin DB, Multiple imputations in sample surveys – A phenomenological Bayesian approach to nonresponse. Imputations and editing of faulty or missing survey data. *U.S. Department of Commerce*. 1978:1-23

Rubin DB, Schenker N. Multiple imputation in health care databases: an overview and some applications. *Statistics in Medicine* 1991;91:473-489.

Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, *New York: John Wiley & Sons, Inc.*

Schafer, J.L. (1997), Analysis of Incomplete Multivariate Data, *New York: Chapman and Hall*.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Giulia Tonini

Menarini Ricerche

Via Rosolino Pilo, 10

50131 Firenze, Italy

Email:gtonini@menarini-ricerche.it