PharmaSUG 2017 - Paper SP05

Combining Survival Analysis Results after Multiple Imputation of Censored Event Times

Jonathan L. Moscovici, QuintilesIMS, Montreal, QC

Bohdana Ratitch, QuintilesIMS, Montreal, QC

ABSTRACT

Multiple Imputation (MI) is an effective and increasingly popular solution in the handling of missing covariate data as well as missing continuous and categorical outcomes in clinical studies. However, in many therapeutic areas, interest has also risen in multiple imputation of censored time-to-event data, because in many cases the Censored at Random (CAR) assumption is not clinically plausible for all subjects and MI allows for Censored Not at Random (CNAR) assumptions. In SAS®, MI is possible through the MI procedure, procedures implementing Bayesian analysis (e.g., MCMC, PHREG) or user-implemented approximate Bayesian bootstrap. In MI, the missing values are filled in and several imputed datasets are created with differing values swapped for the missing ones. Each of those imputed datasets are analyzed separately using the methods that compute the statistics of interest (Kaplan-Meier survival estimates, hazard ratios, etc.). Once these estimates are calculated for each imputed dataset, they are combined, or pooled, using Rubin's rules. These rules assume the estimates to be combined are asymptotically normally distributed. In many cases, such as with survival probabilities, this assumption does not hold and normalizing transformations must be applied beforehand. In this paper, we cover these normalizing methods and present SAS code that implements the necessary data transformations and manipulations for combining various survival analysis estimates such as Kaplan-Meier survival curves (including percentiles), log-rank, Wilcoxon, and other tests for equality of survival curves, and Cox regression estimates of the hazard ratios after multiple imputation of censored time-to-event data. Demonstration is provided using an example imputed data set. Code will be submitted to be made available at http://www.missingdata.org.uk.

INTRODUCTION

Multiple Imputation (MI) (Rubin, 1987) is an effective and increasingly popular solution in handling missing covariate data as well as missing continuous and categorical outcomes in clinical studies. However, interest has also risen in multiple imputation of censored time-to-event data, because in many cases the Censored at Random (CAR) assumption (Heitjan et al., 1991) employed by standard survival analysis methods is not clinically plausible for all censored trial subjects, and MI allows for Censored Not at Random (CNAR) assumptions. Data are considered to be censored at random if the censoring time distribution is independent of the event process, conditional on the observed covariates. Regulatory agencies have noted the need to explore sensitivity to the CAR assumption in the analyses of time-to-event endpoints, as discussed in the Food and Drug Administration (FDA)-commissioned report "The Prevention and Treatment of Missing Data in Clinical Trials" by the National Research Council panel (National Research Council, 2010), and the European Medicines Agency Guideline on Missing Data in Confirmatory Trials (European Medicines Agency, 2010). Analogous to the continuous and categorical outcomes, Lipkovich et al. (2016) proposed multiple imputation-based methods that stress-test the robustness of primary analysis conclusions of time-to-event studies performed under the CAR assumption where subjects withdraw from treatment and follow-up prior to experiencing an event of interest. Multiple imputation provides a flexible tool for such sensitivity analyses, as the imputation model, used to generate imputed values, can be different from the analysis model applied to completed data. This allows the analyst to vary assumptions about missing values (censored subjects) by employing various imputation models and explore the robustness of results under a range of clinically plausible assumptions that could include CNAR assumptions.

Similar to multiple imputation of other types of outcomes, as a result of MI of time-to-event data, multiple datasets are created that share observed values (observed event times), but have different imputed event times for censored subjects. These datasets must be analyzed individually and separately using the time-

to-event analysis methods originally intended. The results from each of these individual analyses are then combined, or pooled, producing an overall result. The MIANALYZE procedure performs this last combining step using Rubin's rules (Rubin, 1987). The workflow for MI and result combination is illustrated in Figure 1.

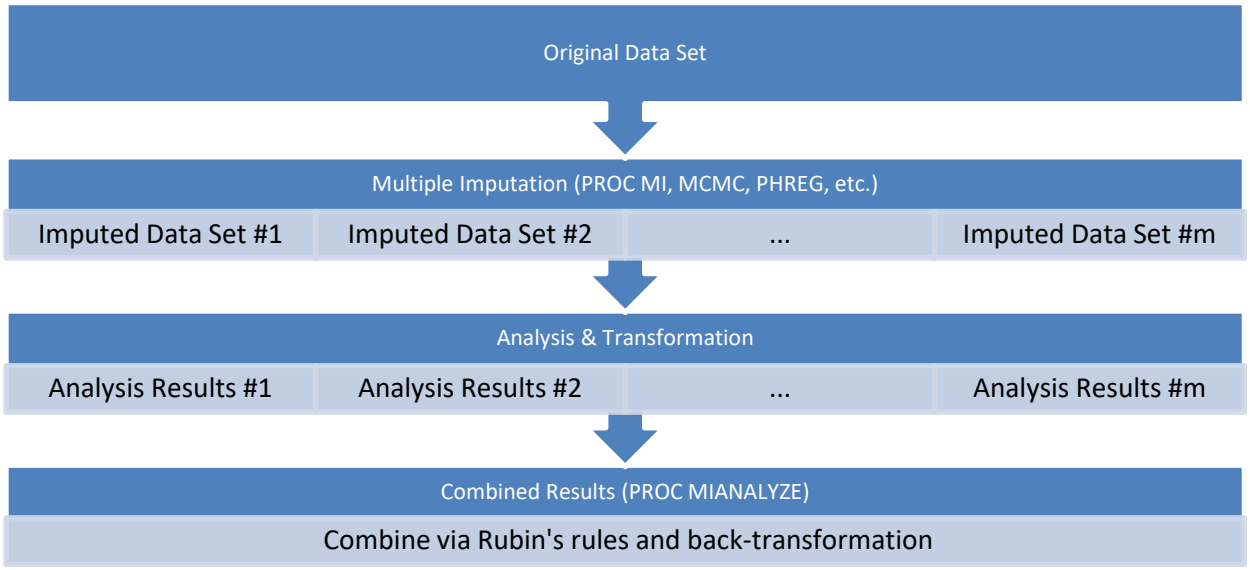

Figure 1. Workflow for MI and result combination

An example of an estimate from survival analysis that can be obtained from the individual imputed datasets and subsequently combined is the estimate of a hazard ratio for the event of interest for the experimental treatment group versus the control. The PHREG procedure, implementing the Cox regression, can be used to produce hazard ratio estimates for each imputed dataset, which would then need to be combined to obtain an overall hazard ratio, as well as its standard error, confidence interval, and an overall test for no treatment effect.

Rubin's rules for combining results from multiple imputed datasets require that the estimated statistics be asymptotically normally distributed. In the case of many typical estimates, such as odds ratios, hazard ratios, relative risks, or survival probabilities, normality does not hold. Therefore, a normalizing transformation is needed before the results can be combined. Van Buuren (2012) suggested transformations for several types of estimated statistics (see Table 1) in order to normalize them.

This paper does not address the question of how to impute event times for censored subjects. For more details on this topic, see, e.g., (Lipkovich et al., 2016; Taylor et al., 2002; Van Buuren, S., 2012; Zhao et al., 2014) and references therein. Rather, this text focuses on solutions for combining estimates that could be routinely produced in time-to-event data analysis reports after multiple imputation, such as Kaplan Meier estimates of the survival curve and survival percentiles, comparison tests of survival distributions (e.g., log-rank, Wilcoxon, and Tarone-Ware test statistics), and Cox regression hazard ratio estimates. The analysis and combination of results are invariant with respect to the assumptions about censored subjects under which multiple imputation was carried out and do not depend on the multiple imputation model used to fill in missing values. So, for any MI model of your choice, the combining process would be the same.

This paper covers the most common survival analysis statistics and includes fragments of SAS code illustrating the key steps. These code fragments belong to a larger set of macros that is planned to be made available at http://www.missingdata.org.uk that can be consulted for a better understanding of the overall process. In the code fragments included in this paper, we preserved the same names of datasets, and (macro) variables as in the full code to facilitate the review of both sources for the readers.

Table 1. Suggested normalizing transformations for select types of statistics. (Excerpts from (Van Buuren, 2012), Table 6.1, p. 156.)

Statistic	Transformation
Odds ratio	Logarithm
Relative risk	Logarithm
Hazard ratio	Logarithm
Survival probabilities	Complementary log-log

The MIANALYZE procedure, implementing Rubin's combination rules, expects as input a dataset where each record contains the results from one imputed dataset. This dataset must contain a variable (column) that represents a point estimate of the analysis statistic and another variable that represents the standard error of the point estimate. The former variable name is specified in the MODELEFFECTS statement and the latter in the STDERR statement. These two quantities are required for Rubin's rules to produce a combined point estimate, standard error, confidence limits, and a hypothesis test for the combined estimate. Therefore, in this paper, when presenting the necessary steps to prepare non-normally distributed statistics for combining, we describe the calculations needed to obtain the normalized point estimates as well as standard errors for the normalized values.

This paper first introduces the example dataset, and then covers the combining of results of Kaplan-Meier survival curves, survival percentiles, some chi-square-distributed statistics for tests of survival curves equality, and Cox regression.

EXAMPLE DATASET

The example dataset used in this paper is from the randomized, double-blind 'Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter (Hammer et al., 1996), otherwise known as the ACTG 175 study[1]. For our purposes, we analyze the secondary endpoint of the study (time to acquired immunodeficiency syndrome (AIDS) or death) and compare two treatment arms (zidovudine monotherapy and zidovudine plus didanosine combination therapy). Table 2 describes the frequency of censoring overall and, in particular, the frequency of subjects lost to follow-up before the planned evaluation period of two years. In this analysis, we consider subjects who had an event or were censored after the planned two-year follow-up period as study completers. We do not impute these. We count these as administratively censored and impute only those subjects who did not experience the AIDS event and were lost to follow-up less than two years after randomization.

As previously mentioned, MI of time-to-event data is most useful for investigation of departures from the CAR assumption, because the standard methods, such as Kaplan-Meier and Cox regression, readily perform survival analyses under the CAR assumption without the need for any imputation of censored data. In this paper, however, we present the results from analyses where multiple imputation of this example dataset was carried out under the CAR assumption. As the focus of this paper is on the implementation procedures, it is helpful to examine the results from the CAR-based multiple imputation and to ensure that they are similar to the results from standard CAR-based analyses not involving imputations, as a high degree of similarity would be expected for approaches that are computationally different but rely on the same assumptions.

[1] With gracious permission from publication author Michael Hughes.

Table 2. Censoring frequencies for ACTG 175 study dataset

	MONOTHERAPY	COMBINATION
# RANDOMIZED SUBJECTS	**619**	**613**
# CENSORED SUBJECTS (% OF RANDOMIZED)	**523 (84.5%)**	**548 (89.4%)**
# LOST TO FOLLOW-UP BEFORE 2 YEARS (% OF RANDOMIZED)	**104 (16.8%)**	**95 (15.5%)**

COMBINING KAPLAN-MEIER SURVIVAL CURVES AFTER MULTIPLE IMPUTATION

A typical analysis of time-to-event data often includes estimation of survival curves using the Kaplan-Meier method. In other words, we are interested in computing survival probabilities at specific times across the follow-up period for these data. Kaplan-Meier estimates of the survival probability function are not normally distributed and thus would require a normalizing transformation before they can be combined with the Rubin's rules. As in reference (Morisot et al., 2015), a complementary log-log normalizing transformation is recommended in order to satisfy the assumption of Rubin's rules.

For each of m imputed datasets that have been created and for a set of time points $t_j, j = 1, \dots, J$, we can compute survival estimates $\hat{S}_i(t_j), i = 1, \dots, m,$ and their variances $Var\left(\hat{S}_i(t_j)\right), i = 1, \dots, m$ using the LIFETEST procedure. Let

$$\widehat{\mathrm{K}}_{\mathrm{i}}(\mathrm{t_j}) = \log[-\log(\hat{S}_i(t_j))]$$

describe the normalizing complementary log-log transformation for the survival estimates.

Along with the transformed point estimates $\widehat{\mathrm{K}}_{\mathrm{i}}(\mathrm{t_j})$, PROC MIANALYZE requires specification of the corresponding survival estimate standard errors on the transformed scale. To compute them, let

$$\mathrm{U_i(t_j)} = Var(\widehat{\mathrm{K}}_{\mathrm{i}}(\mathrm{t_j}))$$

We use the delta method (Oehlert, 1992) with $g\left(\hat{S}_i(t_j)\right) = \log[-\log(\hat{S}_i(t_j))]$ to obtain

$$\mathrm{U_i(t_j)} = Var\left(\widehat{\mathrm{K}}_{\mathrm{i}}(\mathrm{t_j})\right) \approx \left(\frac{-1}{\log\left(\hat{S}_i(t_j)\right) * \hat{S}_i(t_j)}\right)^2 * Var\left(\hat{S}_i(t_j)\right) = \frac{Var\left(\hat{S}_i(t_j)\right)}{\left[\log\left(\hat{S}_i(t_j)\right) * \hat{S}_i(t_j)\right]^2} \quad (1)$$

Since the numerator, $Var(\hat{S}_i(t_j))$, is a known quantity computed by PROC LIFETEST using Greenwood's formula, this equation is a tractable expression for the variance of the transformed estimates. The square root of this estimate gives the standard errors that can be passed to PROC MIANALYZE along with the transformed point estimates, in order to combine the results:

$$S.E.\left(\widehat{\mathrm{K}}_{\mathrm{i}}(\mathrm{t_j})\right) = \sqrt{\mathrm{U_i(t_j)}}$$

In summary, to combine survival probabilities at pre-specified time points, we obtain these survival estimates and their variances for each imputed dataset using PROC LIFETEST, transform them using the equations above, and combine them using PROC MIANALYZE which implements the Rubin's rules. The combined transformed estimates, $\overline{K}(t_j)$ and $SE[\overline{K}(t_j)]$, can then be back-transformed to the original scale using the following expressions for survival probabilities and standard errors:

$$\bar{S}(t_j) = \exp[-\exp[\overline{K}(t_j)]]$$

$$SE[\bar{S}(t_j)] = SE[\overline{K}(t_j)] * \bar{S}(t_j) * \log[\bar{S}(t_j)]$$

SAS Code Fragment 1 demonstrates this process by first performing the normalizing complementary log-log transformation on the survival probabilities contained in the dataset *_spc_pl*, and obtaining the associated standard errors as described in Equation 1, excluding cases where the survival probabilities are 0 or 1. The transformed results are passed along for combining to PROC MIANALYZE: name of the variable containing the transformed point estimates is specified in the MODELEFFECTS statement and the name of the variable containing the transformed standard errors is specified in the STDERR statement. The combined results are thereafter back-transformed in the last DATA step.

```
/* Complementary log-log transformation of survival probabilities and
standard errors*/
/*_spc_pl contains Survival and StdErr, which are the original estimates
obtained from PROC LIFETEST that will be transformed*/
DATA _spc_pl1 _spc_pl2; set _spc_pl;
   if 0 < Survival < 1 then do;
      Survival_cll = log(-log(Survival));
      /*S.E. transformation*/
      Survival_stderr_cll =
      sqrt((1/(log(Survival))**2)*((StdErr**2)/(Survival**2)));
     output _spc_pl1;
   end;
   else output _spc_pl2; /*for Survival=0 or Survival=1 estimates*/
run;
proc sort data=_spc_pl1; by &by &strata &trtvar &timelistvar _Imputation_;
run;

/*Step to remove instances of a timepoint only appearing in one dataset,
since if only one estimate exists there is nothing to combine*/
DATA _spc_pl1; set _spc_pl1;
    by &by &strata &trtvar &timelistvar _Imputation_;
    if first.&timelistvar and last.&timelistvar then delete;
run;

/*Combine transformed estimates*/
proc mianalyze data=_spc_pl1;
    by &by &strata &trtvar &timelistvar;
    modeleffects Survival_cll;
    stderr Survival_stderr_cll;
    ods output parameterestimates=_tmc_survprob_comb;
run;

/*Back-transform combined survival probability estimates and compute CI*/

DATA _tmc_survprob_comb1; set _tmc_survprob_comb;
    Survival_comb = exp(-exp(Estimate));
    Survival_StdErr_comb = abs(Survival_comb * StdErr *
log(Survival_comb));

    zRight = quantile("Normal", 1-&alpha / 2);
       Survival_LCL_comb = Survival_comb**(exp(zRight*StdErr));
       if Survival_LCL_comb<0 then Survival_LCL_comb=0;
       Survival_UCL_comb = Survival_comb**(exp(-zRight*StdErr));
run;
```

SAS Code Fragment 1. Complementary log-log transform of survival probability estimates after multiple imputation

The discussion above assumed a set of time points $\{t_j\}_{j=1}^{J}$ for which survival probability estimates are obtained from the Kaplan-Meier analysis of each of the m imputed datasets. We are often interested in

obtaining combined estimates of the entire Kaplan-Meier survival curve which is a step function with constant step boundaries located at the observed event times. Across multiple imputed datasets, the originally observed event times are the same, but the imputed event times can, of course, be different. Therefore, by default, PROC LIFETEST produces survival estimates at different time points for each imputed dataset. Because of this, to obtain an overall combined survival curve, we need to choose a common set of time points and estimate survival probabilities for this common set from Kaplan-Meier analysis of each imputed dataset. For this purpose, the TIMELIST option of the PROC LIFETEST statement can be used to request survival estimates at specific time points, which may be different from originally observed event times.

More specifically, to obtain a combined estimate of the Kaplan-Meier survival curve, we use a set of time points $\{t_j\}_{j=1}^{J}$ which is the union of the originally observed event times (not imputed) with supplementary time points chosen to span the time period between the minimum and maximum imputed event times across all imputations. Note that we are not adding any additional imputed events, but simply obtaining survival estimates at additional time points across the follow-up period. These extra time points are added to better represent the time periods containing imputed event times across multiple imputed datasets. There may be different ways in which these supplementary time points can be selected. Here, we choose as many of these time points as the number of unique imputed event times across all imputations and spread them uniformly. In other words, if across all imputations there are *r* unique imputed event times, the smallest being s_1 and the largest being s_r, then survival probabilities are here estimated for the following set of time points in addition to the set of observed event times:

$$\{s_1, s_1 + \frac{s_r - s_1}{r}, s_1 + 2\left[\frac{s_r - s_1}{r}\right], \ldots, s_r\}$$

Creation of the list of supplementary time points is illustrated in SAS Code Fragment 2.

```
/*create a list of equally spaced supplementary time points between the
smallest imputed event time to the largest based on all unique imputed
event times provided in the dataset &ietdata */
   proc sql noprint;
      select min(&timevar) into :_scc_mintime from &ietdata;
      select max(&timevar) into :_scc_maxtime from &ietdata;
      select count(unique(&timevar)) into :_scc_nimptime from &ietdata;
      select (max(&timevar) - min(&timevar))/count(unique(&timevar)) into
:_scc_bytime from &ietdata;
   quit;

   %let &etlist=%str(&_scc_mintime to &_scc_maxtime by &_scc_bytime);
```

SAS Code Fragment 2. Generating a list of supplementary time points for survival probability estimation when combining estimates of the Kaplan-Meier survival curves

Using the resulting combined survival probability estimates across the observed event times and supplementary time points, we can produce a combined survival curve. Figure 2 presents the survival curves, by treatment group, for the original (not imputed) data and the combined survival curves for the multiply imputed data. Because imputations were done with a multiple imputation model assuming CAR, which is the same as assumptions behind the Kaplan-Meier analysis of original data, the survival curves obtained with and without imputation are very similar, as expected.

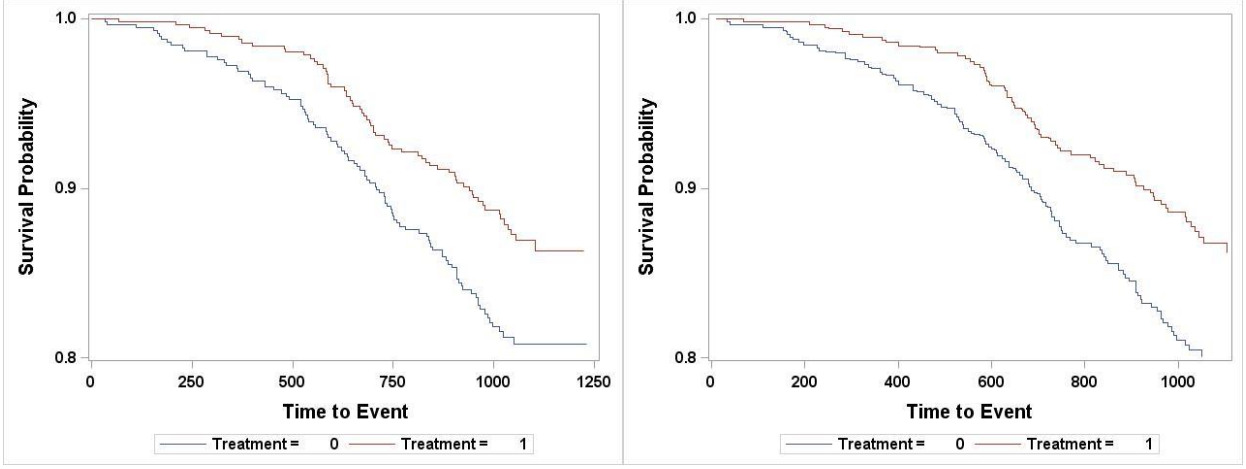

Figure 2. Survival curves based on original data (left) and combined results after MI (right)

OBTAINING SURVIVAL PERCENTILES FROM THE COMBINED PROBABILITIES

Survival percentiles can be computed from the combined survival probabilities $\bar{S}(t_j)$ and their standard errors using standard methods. Alternative approximations exist (Marshall et al., 2009; Collett, 2003) but are not the focus of this paper. The general expression to estimate the $100p$ survival percentile is (SAS Institute Inc., 2008):

$$q_p = \min\{t_j | S(t_j) < 1 - p\}$$

The standard error of the percentile $\mathrm{q_p}$ can be approximated by the Taylor series approximation to the variance of a function of a random variable:

$$SE[\mathrm{q_p}] = \frac{d\bar{S}(q_p)^{-1}}{dq_p} SE[\bar{S}(\mathrm{q_p})]$$

The derivative $\frac{d\bar{S}(q_p)}{dq_p}$, which is actually an estimate of the survival p.d.f. (probability density function) at the time q_p, can be estimated by

$$\frac{\bar{S}(\mathrm{t_+}) - \bar{S}(\mathrm{t_-})}{\mathrm{t_+} - \mathrm{t_-}}$$

where

$$\mathrm{t_+} = min\{\mathrm{t_j}: \bar{S}(\mathrm{t_j}) \leq (1-p) - \varepsilon\}$$
$$\mathrm{t_-} = max\{\mathrm{t_j}: \bar{S}(\mathrm{t_j}) \geq (1-p) + \varepsilon\}$$

and ε is a small number such as $\varepsilon = 0.05$. In some cases, such as when t_+ and t_- end up being equal, a larger ε may be necessary. Then, the 95% confidence limits around q_p can be computed using the normal approximation:

$$\mathrm{q_p} \pm 1.96 \times SE[\mathrm{q_p}]$$

The SAS Code Fragment 3 demonstrates the standard error computation. The identification of the terms $\mathrm{t_+}$ and $\mathrm{t_-}$ and their corresponding survival estimates for the derivative equation is lengthier and can be found in the full code, but the general idea is to look for records satisfying the conditions defining $\mathrm{t_+}$ and $\mathrm{t_-}$ in a dataset containing time points $\{t_j\}_{j=1}^{J}$ and their combined survival estimates as described in the previous section.

For example, Figure 3 illustrates records found in the temporary dataset *_q2* created in order to locate q_p, t_- and t_+ for the 10th percentile using the *surv_quar_comb* macro. In this case, the yellow region contains time points 680, 684 and 692: the records that satisfy the associated conditions from above. These, and related values, are used in SAS Code Fragment 3 to compute the standard errors. Variable *t_minus* contains a lagged value of time point (*timelist* variable) and the indicator variable *done1*=1 marks the record on which the condition for q_p is first met (and from which we use q_p=*timelist* and t_-=*t_minus*). The next record satisfies the condition for t_+ as marked by *done2*=1. After a pass through the dataset to verify these conditions, the required records and values can be extracted based on variables *done1* and *done2* – using the records where they turn from 0 to 1.

timelist	Survival_comb	Survival_StdErr_comb	t_minus	t_plus	s_t_p	done1	done2
667.000	0.9057100247	0.0124389123	660	.	0.9057100247	0	0
671.000	0.905392244	0.012523712	667	.	0.905392244	0	0
677.879	0.9044103985	0.0124275903	671	.	0.9044103985	0	0
679.000	0.9024616642	0.0124158976	677.879042	.	0.9024616642	0	0
680.000	0.9008454888	0.0125000406	679	.	0.9008454888	0	0
684.000	0.8988962533	0.0124726815	680	.	0.8988962533	1	0
692.000	0.8972802908	0.0125557412	680	692	0.8988962533	1	1
697.230	0.8966311991	0.0125539232	680	692	0.8988962533	1	1
703.681	0.8953402197	0.0126419225	680	692	0.8988962533	1	1
706.000	0.893724302	0.0127231453	680	692	0.8988962533	1	1

Figure 3. Computing the 10th percentile using the surv_quar_comb macro

```
   /* Compute standard error of percentile (ster_t_p) and confidence
interval (t_p_LCL, t_p_UCL) */
   ster_t_p = (-(t_plus - t_minus)/(s_t_plus - s_t_minus))*ster_s_t_p;
   zRight = quantile("Normal", 1-&alpha / 2);
   t_p_LCL = t_p - zRight*ster_t_p;
   t_p_UCL = t_p + zRight*ster_t_p;
```

SAS Code Fragment 3. Calculating combined survival percentiles from combined survival probabilities and standard errors

COMBINING CHI-SQUARE–DISTRIBUTED STATISTICS AFTER MULTIPLE IMPUTATION

After multiple imputation, it is possible to combine the test results from all available tests in PROC LIFETEST: log-rank, Modified Peto-Peto, Peto-Peto, Wilcoxon, and Tarone-Ware (SAS Institute Inc., 2008). Each of these tests is based on a chi-square distributed statistic that can be normalized using a Wilson-Hilferty transformation (Wilson et al., 1931). If X^2 is chi-square distributed with d degrees of freedom, then

$$W = \sqrt[3]{X^2/d} \approx N(1 - \frac{2}{9d}, \frac{2}{9d}) \text{ , or}$$

$$\frac{\sqrt[3]{X^2/d} - (1 - \frac{2}{9d})}{\sqrt{\frac{2}{9d}}} \approx N(0,1)$$

Test statistics and their corresponding degrees of freedom can be found in the ODS output dataset HomTests produced by PROC LIFETEST when using the STRATA statement with TEST option. These test statistics obtained from each imputed dataset can be normalized using the above formula and the normalized values can let be input for PROC MIANALYZE in order to obtain combined results for these chi-square based tests, as in SAS Code Fragment 4.

```
DATA _tmc_tests1; set _tmc_tests;
*** ChiSq is a chi-square distributed statistic (e.g., logrank) and DF is
its corresponding degrees of freedom as produced by PROC LIFETEST ***;
    ChiSq_wh=((ChiSq/DF)**(1/3) - (1-2/(9*DF)))/SQRT(2/(9*DF));
    ChiSq_stderr_wh = 1.0;
run;
proc sort data=_tmc_tests1; BY &by test _Imputation_; run;
proc mianalyze data=_tmc_tests1;
   by &by test;
   modeleffects ChiSq_wh;
   stderr ChiSq_stderr_wh;
   ods output parameterestimates=&resprefix._tests;
run;
```

SAS Code Fragment 4. Combining of chi-square-distributed statistics after multiple imputation using the Wilson-Hilferty transformation

Table 3 presents the p-values associated with several test results, both in the combined case after multiple imputation and the original observed data set. The MI results are slightly more conservative, but substantially similar in this example, as would be expected given the CAR-based multiple imputation model.

Table 3. P-values for treatment effect based on the original observed data and combined results after MI

	P-Value				
	Log-Rank	**ModPeto**	**Peto**	**Tarone**	**Wilcoxon**
Multiple Imputation	0.0077	0.0066	0.0066	0.0060	0.0055
Observed Cases	0.0050	0.0042	0.0043	0.0033	0.0025

COMBINING RESULTS OF COX REGRESSION

Cox regression, implemented in PROC PHREG, is often used to estimate hazard ratios, e.g., to compare event hazards between treatment groups. Cox regression model parameter estimates from PROC PHREG represent log hazard ratios. They are normally distributed and do not need any further transformation to be combined with Rubin's rules using PROC MIANALYZE. However, the end results need to be exponentiated to obtain the numerical values of the hazard ratios and associated confidence intervals. SAS Code Fragment 5 illustrates Cox regression analysis of multiply imputed data using PROC PHREG, combining of results, and exponentiation of the combined log-hazard ratios in the last DATA step.

```
proc phreg data=_tmc_1;
  by &by _Imputation_;
  class &trtvar &classvars;
  %if %length(&strata) >0 %then strata &strata ;
  model &timevar * &censvar(&censval) = &trtvar &covars / risklimits
  ties=efron rl;
  ods output ParameterEstimates=_es;
run;

/* Combine model coefficients. */
proc sort data=_es; by &by Parameter CLassVal0 _Imputation_; run;
proc mianalyze data=_es;
  by &by Parameter ClassVal0;
  modeleffects Estimate;
```

```
   stderr stderr;
   ods output ParameterEstimates = _es_mianal;
run;

/*Exponentiate in order to obtain hazard ratio estimates and confidence
intervals */
data &resprefix._hr;
   set _es_mianal;
   Log_HR_comb=Estimate;
   HR_comb=exp(Estimate);
   HR_LCL_comb=exp(LCLMean);
   HR_UCL_comb=exp(UCLMean);

   keep &by Parameter ClassVal0 Log_HR_comb HR_comb HR_LCL_comb HR_UCL_comb
   Probt;
   rename Probt=HR_pval_comb;
run;
```

SAS Code Fragment 5. Combining Cox regression results after multiple imputation

Table 4 contains Cox regression hazard ratio estimates and associated confidence intervals obtained from the observed and multiply imputed data. As expected with the CAR-based multiple imputation model used in this example, the hazard ratios and confidence interval widths are very similar.

Table 4. Cox regression estimates from the original observed data and combined results after MI

	Estimate (Combination vs. Monotherapy)	95% Confidence Interval
Multiple Imputation	0.618	0.450, 0.849
Observed Data	0.624	0.455, 0.855

CONCLUSION

This paper describes the methods for combining results of survival analysis when multiple imputation is used to impute event times for censored subjects. In order to apply Rubin's rules to combine the results from multiple imputed datasets into an overall result, many statistics estimated in the course of survival analysis need to be transformed to satisfy asymptotic normality assumptions required by the Rubin's rules. This paper describes the necessary transformation steps and thus expands a toolbox of methods that facilitate the use of multiple imputation not only with continuous and categorical data (Ratitch et al., 2013), but also with censored time-to-event data. The focus of this text is on combining common survival statistics, such as survival probabilities, survival percentiles, Cox regression hazard ratios, and the chi-square distributed statistics available in PROC LIFETEST for tests of equality of survival distributions (log-rank, modified Peto-Peto, Peto-Peto, Wilcoxon, and Tarone-Ware). Methods to carry out multiple imputation of event times for censored subjects are discussed elsewhere, e.g., (Lipkovich et al. 2016; Taylor et al., 2002; Van Buuren, 2012; Zhao et al., 2014). Although multiple imputation can be performed using various imputation models and under a variety of assumptions, the combination methods presented here are the same regardless of the particular MI method applied beforehand. The motivation for this work is to empower statisticians to evaluate the robustness of clinical trial survival analysis results to the CAR assumption underlying the standard Kaplan-Meier and Cox regression methods, as multiple imputation provides a powerful and flexible tool to perform analyses under a range of deviations from the CAR assumption.

REFERENCES

Collett, D. 2003. *Modelling survival data in medical research*. Second. London: Chapman & Hall/CRC.

European Medicines Agency. 2010. Guideline on Missing Data in Confirmatory Clinical Trials. 2 July 2010. EMA/CPMP/EWP/1776/99 Rev. 1". Accessed March 15th 2017 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf

Hammer, S.M., Katzenstein, D.A., Hughes, M.D., Gundacker, H., Schooley, R.T., Haubric,h R.H., Henry, W.K., Lederman, M.M., Phair, J.P., Niu, M., Hirsch, M.S., Merigan, T.C. 1996. "A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter." *The New England Journal of Medicine*, 335:1081–1090.

Heitjan, D.F, Rubin, D.B. 1991. "Ignorability and coarse data." *Annals of Statistics*, 19:2244-2253.

Lipkovich, I., Ratitch, B., O'kelly, M. 2016. "Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints." *Pharmaceutical Statistics*,15(3):216-29.

Marshall, A., Altman, D.G., Holder, R.L., Royston, P. 2009. "Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines." *BMC Medical Research Methodology*, 9:57.

Morisot, A., Bessaoud, F., Landais, P., Rébillard, X., Trétarre, B., Daurès, J.P. 2015. "Prostate cancer: net survival and cause-specific survival rates after multiple imputation." *BMC Medical Research Methodology*, 15:54.

National Research Council. 2010. *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academic Press.

Oehlert, G. W. 1992. "A note on the delta method." *American Statistician*, 46: 27–29.

Ratitch, B., Lipkovich, I., O'Kelly, M. 2013. "Combining Analysis Results from Multiply Imputed Categorical Data." *Proceedings of the Pharmaceutical SAS Users Group Conference (PharmaSUG 2013)*, paper SP03. Cary, NC: SAS Institute Inc

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

SAS Institute Inc. 2008. SAS/STAT ® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Taylor J., Murray S., Hsu, C.-H. 2002. "Survival estimation and testing via multiple imputation." *Statistics and Probability Letters*, 58:221-232

Van Buuren, S. 2012. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC Press.

Wilson, E.B., Hilferty, M.M. 1931. "The distribution of chi-squared." *Proceedings of the National Academy of Sciences*, Washington, 17:684–688.

Zhao, Y., Herring, A.H., Zhou, H., Ali, M.W., Koch, G.G. 2014 "A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring." *Journal Biopharmaceutical Statistics*, 24(2):229-53.

ACKNOWLEDGMENTS

We would like to thank QuintilesIMS Inc. for supporting this work and conference participation.

CONTACT INFORMATION

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

Jonathan L. Moscovici
QuintilesIMS Inc.
E-mail: jonathan.moscovici@quintiles.com