Systemization of Time-to-event Analyses for Pharmacometric Applications

Julie A Passarell, Caroline Passarell, Jill Fiedler-Kelly

Cognigen Corporation, a *SimulationsPlus* company; Buffalo, NY
http://www.cognigencorp.com

T-003

ABSTRACT

Objectives. In addition to the values of efficacy and safety endpoints, the timing of such endpoints relative to the start of treatment is often of interest. A commonly used statistical methodology for analyzing such time-to-event data is survival analysis (SA). Building upon the implementation of CDISC standards in data collection facilitating the standard format of clinical trial data and previous systemization efforts,[1] the objective of this work was to develop a system to standardize analysis dataset creation, exploratory data review, and SA procedures for exposure-response (E-R) analyses of time-to-event endpoint data.
Methods. SAS® software was used to develop a code library to transform source clinical trial data into an analysis-ready dataset for use in E-R SA. A library of SAS code for the creation of standard exploratory graphs and tables was also developed. A systematic approach to statistical analysis using SAS PROC PHREG and R was developed, based on standard methods for model building and discrimination[2] including the calculation of standard statistics and production of typical diagnostic plots.
Results. The standardized process for dataset creation, exploratory data analysis, and SA was tested on 8 compounds and refined as new variations and additional data checks were identified. This refined process and systematic approach resulted in a greater than 50% decrease in analyst time required for evaluation of E-R relationships for time-to-event endpoints. Other positive benefits of system implementation include a reduction in training time for new pharmacometricians and improved quality and consistency of reporting for E-R SA.
Conclusions. Standardization of analysis-ready dataset creation, exploratory graphical evaluation, and the survival analysis process for time-to-event endpoints has proven instrumental in generating timely understanding of E-R relationships to facilitate model-based decision making under tight timelines and allows for the evaluation of additional endpoints and synthesis of findings across endpoints.

METHODS

Data Considerations

- How is the endpoint of interest defined? Is the definition the same across studies or is standardization necessary?
- Define the start and stop dates of interest usually based on dosing of the drug.
- Determine the epoch of time to be considered for analysis (that is, if dosing is daily, consider daily records; if dosing is weekly, consider weekly records, etc.).

Analysis Dataset Construction

- A dataset requirements form was developed specifically for the evaluation of exposure-response (E-R) relationships for time-to-event endpoints using survival analysis.
  - Used to communicate the structure and content of the dataset required for the analysis.
  - Requires the input of key pieces of information for the analysis from source data (for example, endpoints, time, covariates, etc.) and facilitates the selection of programming templates from the code library used for assembly of the analysis dataset.
  - Provides the list and order of data deletions.
- A standard dataset build process was developed (**Figure 1**):
  - Step 1 - Covariate dataset is built based on the requested stationary and time-varying covariates specified in the requirements, for those patients in the population of interest.
  - Step 2 - Endpoint data is processed from the derived (analysis data model) or source datasets. Records are added over time for a specified epoch of time (that is, daily records) until the event occurs or the patient is censored ("no event").
  - Step 3 - Required covariates are appended onto the endpoint records.
  - Step 4 - Individual exposure measures are appended onto the merged endpoint and covariate data.
  - Step 5 - Graphing-control variables are appended onto the merged endpoint, covariate, and exposure data.
- Standard rules for data checking and handling of data anomalies are built into the code templates (for example, management of missing dates and imputation of missing covariates).
- Existing program templates can be used to incorporate concomitant medications and/or generation of drug exposures, as needed.
- Template programs have been developed and QCd to accommodate these and other variations of the data assembly process.
  - Template programs provide a solid starting point, and can be used as-is or customized if the data, study design, or requirements do not align with the templates.
  - These new or updated templates are then added to the code library for use by the entire data programming department to facilitate and reinforce the use of standardized, QCd code on future projects and programs.

Exploratory Data Analysis

- A library of SAS® code for the creation of standard exploratory graphs and tables specifically for time-to-event endpoint data was developed to facilitate rapid evaluation and assessment of E-R relationships. Exploratory plots of raw data provide enhanced understanding of the informational content of the data relative to the models to be tested and evaluations to be performed.[2,3,4]
- Standard exploratory data analysis library includes:
  - Scatterplot matrices of individual exposure measures and calculation of correlation measures for each pair of exposures;
  - Scatterplot matrices of continuous covariates of interest and calculation of correlation measures for each pair of continuous covariates;
  - Frequency distributions of the individual exposure measures, overall and stratified by dose;
  - Frequency distributions of the time-to-event endpoint, stratified by censor;
  - Summary statistics for each of the patient covariates and exposure measures, overall and stratified by dose and by censor;
  - Kaplan-Meier plot of survival versus time, stratified by categories of covariates;
  - Kaplan-Meier plot of survival versus time, stratified by quantiles of drug exposure (for time-varying exposure, survival probabilities are calculated using the survfit function in R[5]); and
  - Kaplan-Meier plot of the hazard function versus time, stratified by categories of covariates or quantiles of drug exposure.

Survival Analysis

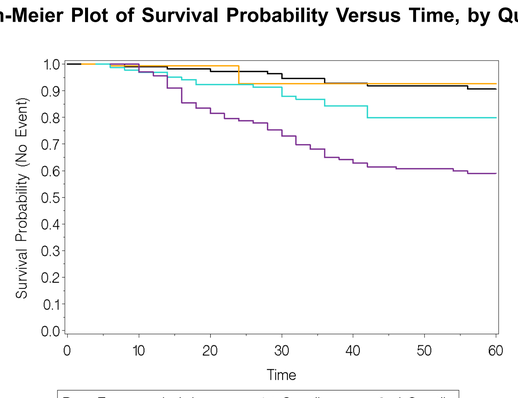
- A systematized approach to statistical analysis of time-to-event data using SAS PROC PHREG was developed.[3,6,7]
- Standard SA methods for model building and discrimination are used to facilitate the calculation of standard statistics and production of typical diagnostic plots for model building and evaluation.[2,7]
- **Figure 2** describes the standard process of the analysis of time-to-event data.

RESULTS

Exploratory Data Analysis

- Example frequency distribution of time-to-event or censor, stratified by censor (**Figure 3**).
- Example Kaplan-Meier plot of survival probability versus time, stratified by treatment shows whether the probability of survival (no event) is similar across dose groups (**Figure 4**).
- Example Kaplan-Meier plot of survival probability versus time, stratified by drug exposure shows whether there is a trend for increasing or decreasing probability of survival (no event) across quartiles of drug exposure (**Figure 5**).
  - When drug exposure is time-varying, the extended Kaplan-Meier estimator is used to calculate the probability of survival using the survfit function in R.[5,8,9]

Figure 3. Frequency Distribution of Time-to-event or Censor



The number on the x−axis represents the median of the range of values for that bar.
The number above each bar represents the number of patients in each bar.

Figure 4. Kaplan-Meier Plot of Survival Probability Versus Time, by Treatment

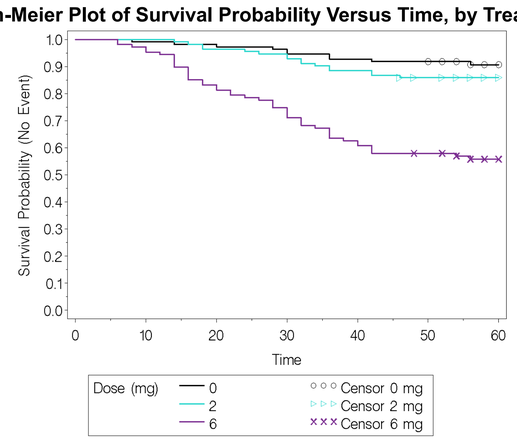


Figure 5. Kaplan-Meier Plot of Survival Probability Versus Time, by Quartiles of Drug Exposure

Survival Analysis

- Check the proportional hazards assumption.
  - The appropriateness of the proportional hazards regression method and the validity of the results depends on the correctness of the proportional hazards assumption. Graphical and analytical methods can be used to verify this assumption for each drug exposure and covariate to be tested. This assumption implies that the graphs of log[-log S(t)] versus log(time) are parallel. Plots are inspected to determine if curves are vertically separated by an approximately constant amount as shown in **Figure 6**.
- Evaluation of time-varying drug exposure in the base model.
  - Example table of univariate fits of each drug exposure measure (**Table 1**) can be produced automatically using the standard program to facilitate the selection of the most appropriate exposure measure to correlate with the endpoint.
  - Covariate evaluation - example table illustrating forward selection results is provided in **Table 2**. This table is automatically generated and can be exported as a formatted Word® table for direct import into technical reports or presentations.

Figure 6. Check the Proportional Hazards Assumption by Plotting Log[-Log S(t)] Versus Log(time)

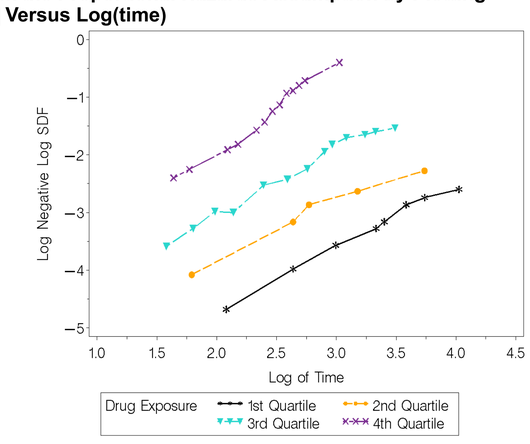


Figure 1. Standard Dataset Build Process

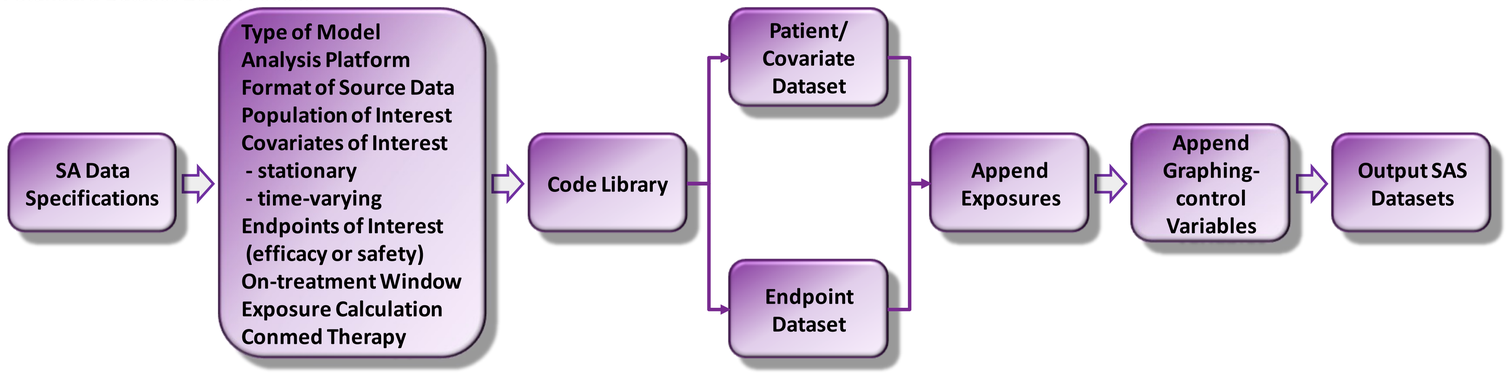


Figure 2. Survival Analysis Process With Standard SAS Program Naming Conventions

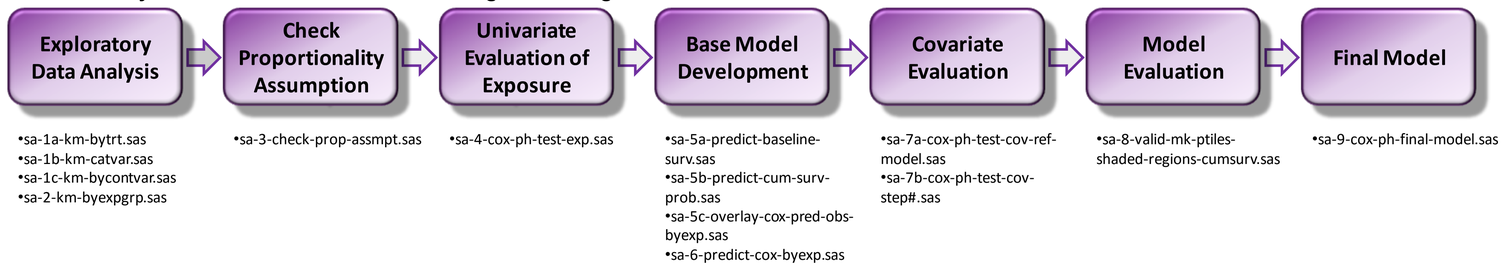


Table 1. Summary of Drug Exposure Evaluation for Exposure-Response Analysis of Time-to-event Data

Model	Exposure Measure	N	Estimate	Standard Error	%SEM	Hazard Ratio	95% CI for Hazard Ratio		*P value*	AIC Intercept and Covariates
							Lower Bound	Upper Bound		
Reference -2 Log L and AIC = 873.398										
1	C_{av} (ng/mL)	8228	0.2112	0.0484	22.9226	1.235	1.123	1.358	<.0001	772.854
	Interaction between C_{av} and time	8228	-0.0063	0.0027	42.7126	0.994	0.988	0.999	0.0192	
2	C_{max} (ng/mL)	8228	0.2314	0.0534	23.0774	1.260	1.135	1.399	<.0001	789.250
	Interaction between C_{max} and time	8228	-0.0071	0.0029	41.5566	0.993	0.987	0.999	0.0161	
3	C_{min} (ng/mL)	8228	0.2036	0.0466	22.8779	1.226	1.119	1.343	<.0001	787.262
	Interaction between C_{min} and time	8228	-0.0060	0.0026	43.1227	0.994	0.989	0.999	0.0204	
4	$AUC_{0\text{-}24}$ (ng × h/mL)	8228	0.0088	0.0020	22.9107	1.009	1.005	1.013	<.0001	787.934
	Interaction between $AUC_{0\text{-}24}$ and time	8228	-0.0003	0.0001	42.7193	1.000	1.000	1.000	0.0192	

Abbreviations: AIC, akaike information criterion; $AUC_{0\text{-}24}$, area under the concentration-time curve from time 0 to 24 hours; C_{av}, average steady-state concentration; CI, confidence interval; C_{max}, maximum drug concentration; C_{min}, minimum drug concentration; N, number of observations; %SEM, standard error of the mean expressed as a percentage.

Table 2. Summary of Forward Selection of Covariates for the Exposure-Response Analysis of Time-to-event Data

Step	Covariate		Estimate	Standard Error	%SEM	Hazard Ratio	95% CI for Hazard Ratio		df	*P value*	Change in -2LL[b]
							Lower Bound	Upper Bound			
Reference -2 Log L and AIC = 772.854											
1	Age (y)		-0.0178	0.0110	61.47	0.982	0.961	1.004	1	0.1038	-2.71171
	Baseline weight (kg)		-0.0176	0.0075	42.60	0.983	0.968	0.997	1	0.0189	-5.77451
	Baseline body mass index (kg/m^2)		-0.0529	0.0279	52.70	0.949	0.898	1.002	1	0.0578	-3.67355
	Sex		0.1609	0.2348	145.94	1.175	0.741	1.861	1	0.4932	-0.46733
	Ethnicity		-0.1684	0.3749	222.67	0.845	0.405	1.762	1	0.6534	-0.21041
	Race	Black	0.5649	0.3376	59.76	1.759	0.908	3.410	2	0.0943	-8.92744
		Other races grouped	1.1060	0.3630	32.82	3.022	1.484	6.156		0.0023[b]	-8.92744
No covariates met the criteria for statistical significance of α=0.01.											

Abbreviations: AIC, akaike information criterion; CI, confidence interval; df, degrees of freedom; %SEM, standard error of the mean expressed as a percentage.
[a] Covariates contributing a change in minus twice the log likelihood of at least 6.64 (α = 0.01, 1 df) will be considered significant.
[b] This covariate is not considered statistically significant since the model contains effects for Black and Other races grouped (2 degrees of freedom).

- Goodness-of-fit plot of Schoenfeld residuals versus time ideally will show an even scatter of points across time (**Figure 7**).
- Model evaluation can be performed using a visual predictive check. This standard pharmacometric method can be applied to time-to-event data by, first, simulating exposure measures and then calculating the time-to-event predictions to assess the predictive ability of the model (**Figure 8**).
- Final model
  - In order to obtain a plot of the model-predicted probability of survival over time with the observed Kaplan-Meier estimates overlaid, the following steps are taken:
    - Using the final E-R model estimates, predict the baseline probability of survival where exposure = 0 to be used to predict the survival at the first time point (t = 0).
    - At each time (t), use the model to calculate S(t) and S(t+1) using the median value of the covariate at time t. Then calculate P(t) = [S(t) - S(t + 1)]/S(t), which is the estimated conditional probability of the event in the interval of time t to t + 1. Then piece together the predicted survivor function by S*(t) = [1 - P(1)][1 - P(2)]...[1 - P(t)].[3]
  - Standard code reads in the final model output to produce a plot of the final E-R model predictions versus time, stratified by dose and significant covariate(s) with Kaplan-Meier estimates of the observed time-to-event data for each dose level (**Figure 9**).
    - This figure provides a better understanding of the concordance between the model-predicted probabilities of survival and the observed Kaplan-Meier estimates of the occurrence of the event over time.

Figure 7. Goodness-of-fit Plot of Schoenfeld Residuals Versus Time

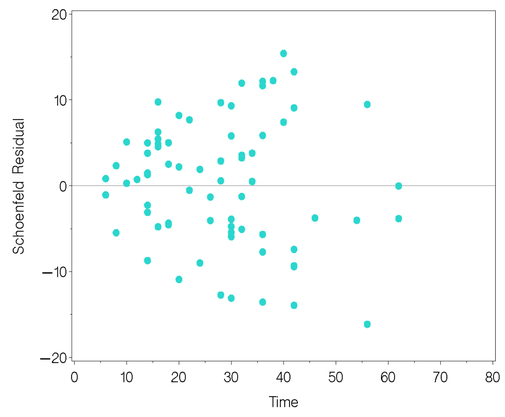


Figure 8. Visual Predictive Check Plot of Simulated Percent of Patients With Event Versus Time With Kaplan-Meier Estimates of the Observed Data by Dose

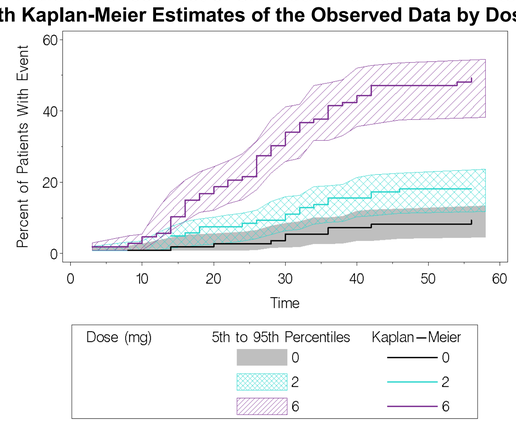
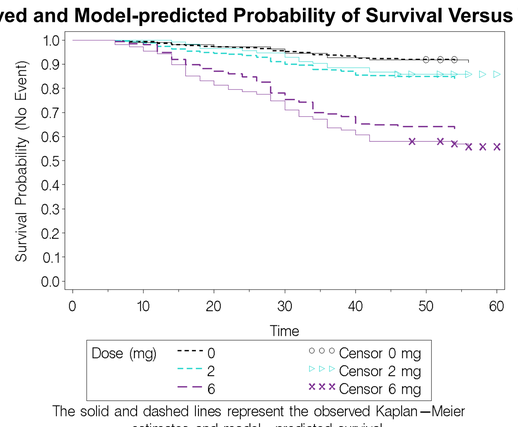


Figure 9. Observed and Model-predicted Probability of Survival Versus Time, by Dose



The solid and dashed lines represent the observed Kaplan−Meier estimates and model−predicted survival.

CONCLUSIONS

Standardization of the processes of analysis-ready dataset creation, exploratory graphical evaluation, and survival analysis for time-to-event endpoints has proven instrumental in generating timely understanding of E-R relationships to facilitate model-based decision making under tight timelines and allows for the evaluation of additional endpoints and synthesis of findings across endpoints.

REFERENCES

1. Passarell J, Passarell C, Hitchcock D, Fiedler-Kelly J, Grasela T. Systemization of logistic regression analysis for pharmacometric applications. Poster presented at: American Conference on Pharmacometrics (ACoP): October 4-7, 2015; Crystal City, VA.
2. Kleinbaum DG and Klein M. Survival analysis: a self-learning text. 3rd ed. New York: Springer; 2012. 700 p.
3. Allison PD. Survival analysis using SAS®: a practical guide. 2nd ed. Cary (NC): SAS Institute Inc.; 2010. 336 p.
4. Cantor AB. SAS® survival analysis techniques for medical research. 2nd ed. Cary (NC): SAS Institute Inc.; 2003. 248 p.
5. R: A language and environment for statistical computing [computer program]. Version 2.10.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-9000051-07-0, URL http://www.R-project.org. 2009.
6. SAS [computer program]. Version 9.4. Cary, NC: SAS Institute. 2013.
7. Collett D. Modelling survival data in medical research. 2nd ed. Boca Raton: CRC Press LLC; 2003. 410 p.
8. Snapinn SM, Jiang Q, Iglewicz B. Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. Am Stat. 2005;59(4):301-307.
9. Therneau TM and Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer; 2000. 350 p.

Passarell J, Passarell C, Fiedler-Kelly J. Systemization of Time-to-event Analyses for Pharmacometric Applications. Poster presented at: American Conference on Pharmacometrics (ACoP8); October 15-18, 2017; Fort Lauderdale, FL.

For additional information, please contact Julie Passarell, MA
Cognigen Corporation, a SimulationsPlus company
1780 Wehrle Drive, Suite 110, Buffalo, NY 14221
(716) 633-3463, ext. 266 or julie.passarell@cognigencorp.com