

SIMULATING ABANDONMENT USING KAPLAN-MEIER SURVIVAL ANALYSIS IN A SHARED BILLING AND CLAIMS CALL CENTER

Quinn D. Conley

Analytics Resource Center
Westfield Insurance
Westfield Center, OH 44251, USA

ABSTRACT

Abandonment is a key indicator of performance and a driver of service level in a call center. Calls that abandon affect the wait times of the remaining calls in the queue and the ability of call center resources to service the remaining calls. This interaction is further complicated when the call center has multiple arrival channels, handled by two groups of resources with a shared pool of resources between them. In this case, a valid call center model necessitates a highly accurate method for modeling abandonment. This paper documents a unique application of Kaplan-Meier survival analysis to model call center abandonment in a discrete event simulation model. The paper also demonstrates the benefits of using Kaplan-Meier verses another approach.

1 INTRODUCTION

This paper is focused on the accurate and straight-forward modeling of abandonment in a discrete event simulation of a call center. No statistical comparison, theoretical model development, or complex distribution fitting is needed because the approach relies on the empirical distribution of the underlying abandonment data. The approach documented in this paper is pragmatic and implementable, with a demonstrated ability to validate on key abandonment and service level metrics.

2 SIMULATION PROJECT OVERVIEW

2.1 Company Background

Westfield Insurance is one of the top 50 property and casualty insurance companies in the USA. Westfield is headquartered in Westfield Center, Ohio (about 45 miles south of Cleveland and 25 miles west of Akron). It is one of the largest non-public companies in Ohio, with over \$1.6 billion in annual revenues. Westfield distributes its products through a network of over 1,000 independent agencies. The company actively writes personal insurance in 10 states, commercial insurance in 21 states, and surety in 31 states.

2.2 Business Problem

Westfield leadership wanted to evaluate options for expanding the use of their call center for claims reporting. The solution for understanding the impact of the proposed change was a discrete event simulation model of the entire claims first notice of loss (FNOL) process.

2.3 Call Center Complexity

A critical piece of the FNOL process is claims call intake in Westfield's call center. The call center handles the intake of insurance claims, claims inquiries, billing inquiries, and billing modifications. The call

center is the single point of contact for customers and the main point of contact for independent agents. In order to ensure a consistent experience for callers and to accommodate the staffing efficiency and flexibility needed for variable, weather-based claims arrivals, the billing and claims call centers have a subset of pooled resources.

The pooled resources are deemed hybrids. Hybrids can answer both claims and billing calls, and are split into two different types. They can either be a hybrid (take next claim or billing call) or a designated hybrid (answer one type of call and are available to answer an overflow call of the opposite type). An example of a designated hybrid would be a customer service representative (CSR) primarily handling billing calls, who will also take an overflow, claims call. Overflow calls are calls that have waited beyond the service level for a primary resource, and have been routed to a designated hybrid resource. For an in-depth description of modeling the resource scenario stated above, along with additional resource complexities, see Conley and Grabau (in press).

Due to the delicate interplay between call abandonment, resource type staffing, and resource to call queue assignment, a highly accurate modeling method for abandonment was essential.

3 LITERATURE REVIEW OF KAPLAN-MEIER ABANDONMENT IN SIMULATION MODELING

3.1 Zohar

Zohar et. al. (2002) uses Kaplan-Meier and simulation for testing the hypothesis that a caller's patience depends on their wait time expectations. The premise is that customers adapt their patience based on prior experience and the current performance of the system. Kaplan-Meier is used to estimate the mean of the current wait time distribution. The mean is then multiplied by a constant and the product is assigned as the caller's abandonment time. Simulation is used to replicate the process.

3.2 Roubos

Roubos et. al (2012) demonstrates that the hyperexponential distribution can be used to model patience in a call center. Kaplan-Meier is then used to develop the empirical CDF for patience. Finally, simulation is used to compare the results of the hyperexponential model to the empirical CDF for patience.

3.3 Major Differences that Make this Approach Unique

The author believes that the following make this method unique compared to the publications cited in the literature review:

- This approach was used for multiple caller types with different abandonment distributions for each type
- This approach accounts for CSR's that have multiple skill sets and answer more than one call type
- The method was implemented in a discrete event simulation model
- The goal of this approach is to replicate the historical abandonment behavior based on actual data, not to develop a model for patience

4 APPROACH

4.1 Data

The call data are separated by channel into claims and billing datasets. The claims channel includes call arrivals from insurance agents, customers, and claimants who are all directed to one queue via a single phone number. The billing channel includes arrivals from agents into an agent queue and arrivals from customers into a customer queue via separate phone numbers.

The number of abandoned calls affects the abandonment rate as well as the service level. The abandonment rate is simply the number of abandoned calls divided by the number of offered calls, in a given time period. Service level is the number of calls answered in less than a certain number of seconds divided by the number of calls answered, in a given time period. When a call waits beyond the service level and then abandons, it positively affects the service level by decreasing the number of answered calls in the denominator of the service level calculation.

Comparing the two caller types for the billing channel in Figure 1 yields a couple of interesting findings. First, the agents that abandon are willing to wait longer than the customers that abandon. Second, there is a drop in the abandoned calls around the service level, showing the impact of the call being queued to a designated hybrid resource once the service level has not been met.

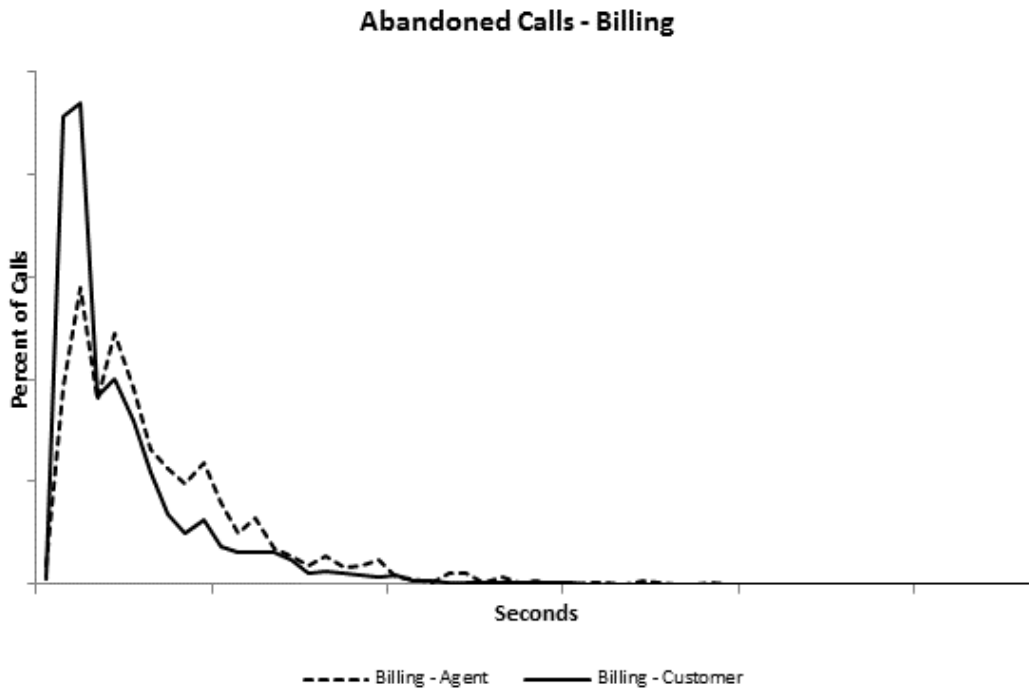


Figure 1: Abandoned Time Distribution for Billing Calls - Agents vs. Customers

By comparing billing and claims abandoned calls in Figure 2, one observes that billing customers are less patient than claims customers.

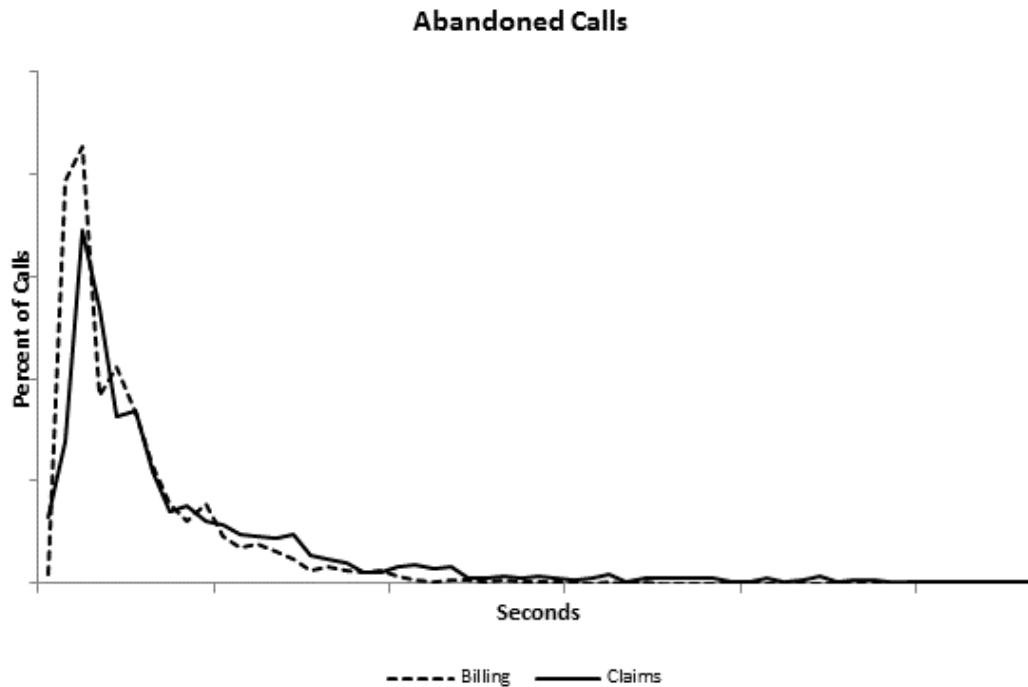


Figure 2: Abandoned call time distribution for Billing vs. Claims

4.2 Kaplan-Meier Method

The problem with only using the abandoned calls to assign abandon times to all entities is that it ignores the observations that were answered, even though they waited. If one were to remove the answered calls from this project, 95% of the wait time observations would be ignored just because they did not abandon! The abandoned call distribution is obviously incomplete in its ability to describe the time that callers are willing to wait before hanging up. The answered calls have something to add, via the censored observations included in Kaplan-Meier survival analysis.

Kaplan-Meier was originally used to analyze the length of time patients survived in medical studies, where death was the event of interest. A unique feature of Kaplan-Meier is that it makes use of censored data. Right-censored data are observations for which the time until the event of interest is not known, yet for which some information is available. For example, a patient that drops out of a medical study is right-censored because the time until death is unknown, yet there is some information on how long they survived.

Kaplan-Meier includes both the abandoned and answered observations in its calculation of survival rate. The event of interest, in this case, is the time that a call abandons. This time is readily available for abandoned calls but, it is not available for answered calls because they are right-censored. In other words, answered calls were serviced before they reached the limit of their patience and abandoned (so the time they would have waited until abandoning is unknown, but there is information about how long they waited). The answered calls contribute to the survival rate calculation up until the point when they are censored (the wait time at which the call was answered). At this point, they are removed from the denominator for the survival rate calculation in the next interval and forward. Including the wait times of both the abandoned and answered calls gives a complete picture of caller patience. For more information on Kaplan-Meier survival analysis, see Kaplan et. al. (1958) and Lee et. al. (2003).

4.2.1 Cumulative Survival Rate vs. Cumulative Hazard Rate

The author recommends using the cumulative survival rate because of its ability to accurately translate to a cumulative probability. The cumulative hazard rate is not a probability; it is a rate of probability. Let T be the survival time. T can be described by the survival function $S(t) = 1 - P(\text{an individual fails before } t) = P(T > t)$. Since $P(\text{an individual fails before } t)$ is defined as the cumulative distribution function $F(t)$ of T , then $F(t) = 1 - S(t)$. It can be shown that the cumulative hazard function $H(t) = -\ln S(t)$. Thus at $t = 0$, $S(t) = 1$, $H(t) = 0$, and at $t = \infty$, $S(t) = 0$, $H(t) = \infty$. Therefore, the cumulative hazard function can be any value between zero and infinity (Lee et. al. 2003). For the purposes of simulating abandonment, the cumulative distribution function needs to be bound by values from zero to one. Furthermore, it can be shown that the cumulative hazard rate has a value of 1 at survival $\approx 36.8\%$. Here is an example to demonstrate:

$$\begin{aligned} -\ln(x) &= 1 \\ \ln(x) &= -1 \\ e^{\ln(x)} &= e^{-1} \\ x &= e^{-1} \approx .368 \end{aligned}$$

4.2.2 Using Survival Rate to Assign Abandon Time

The survival rate from the Kaplan-Meier calculation was used to develop a cumulative distribution function (CDF) for abandonment, which was then used to assign abandon times to individual entities in the simulation.

For calculating the survival rate, calls were separated into billing-agent, billing-customer, and claims segments. IBM SPSS Modeler was used to execute the Kaplan-Meier survival analysis for each call segment. The survival time was set equal to the wait time of each call and the target was set to a flag if the call abandoned. The results for billing in Figure 3 show that agents are more patient than customers.

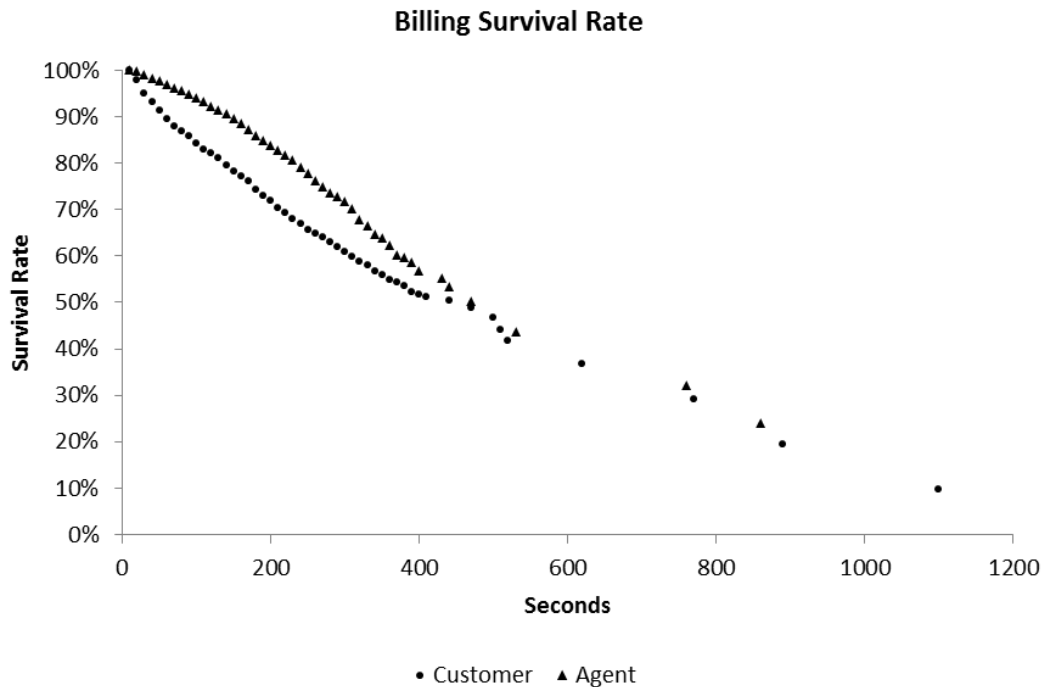


Figure 3: Kaplan-Meier Survival Rate for Billing Calls - Customers and Agents

The results for claims indicate a change in patience for callers that wait longer than a certain amount of time, as shown in Figure 4.

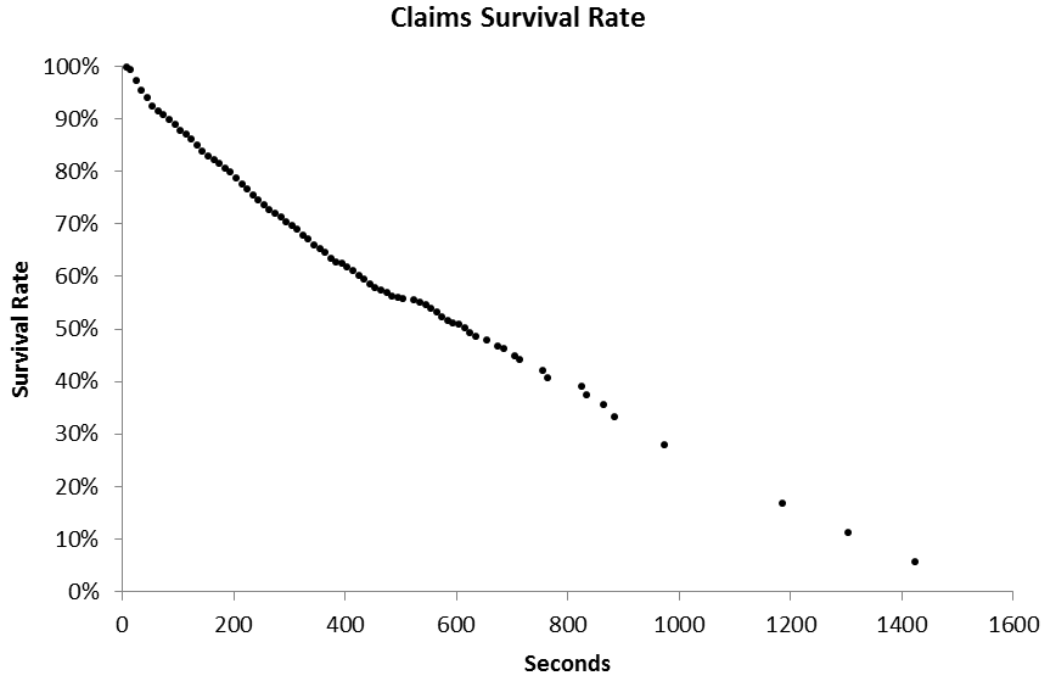


Figure 4: Kaplan-Meier Survival Rate for Claims

Next, the survival rate was used to develop a CDF for abandonment by call segment. As shown in section 4.2.1, the CDF for abandonment is one minus the survival function at time t . Through this calculation, most of the needed CDF was obtained. The last piece of the CDF needed to be estimated because the survival rate did not go to zero for the empirical data. The approach for billing and claims was to fit a curve to abandonment CDF from the last observation to 1.00. The abandonment CDF input for the simulation model was the abandonment CDF including the fitted curve from the last observation to 1.00. As shown in Figure 5, the abandonment CDF for billing customers and agents converged after a certain amount of time. One fitted curve was used for the last piece of the CDF in the model.

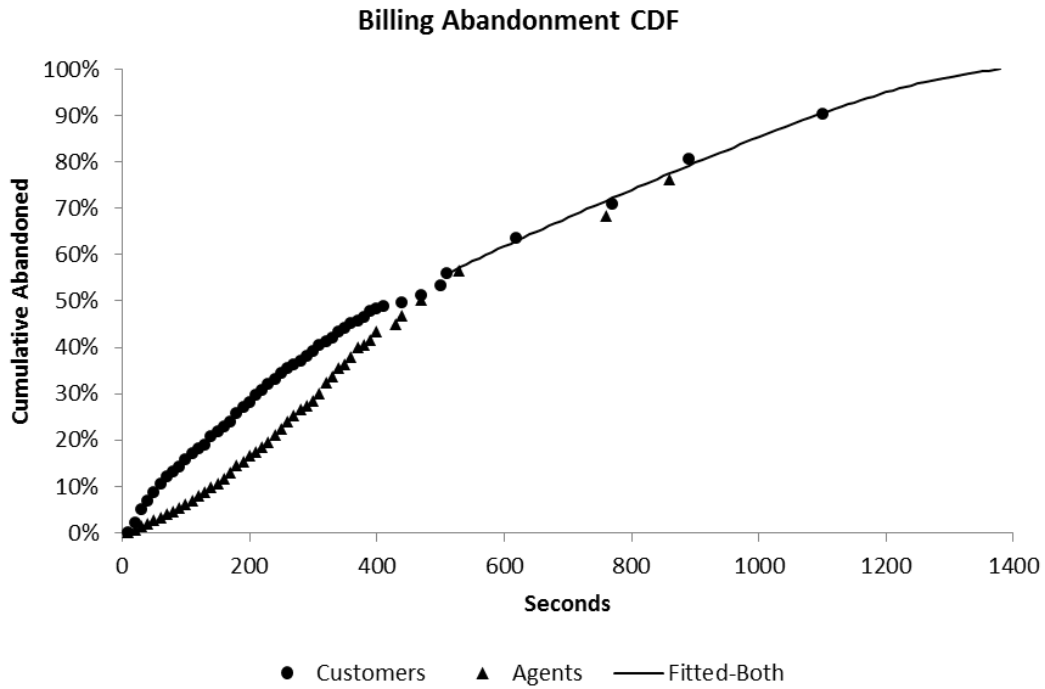


Figure 5: Billing abandonment CDF Estimation

The result for claims is shown in figure 6.

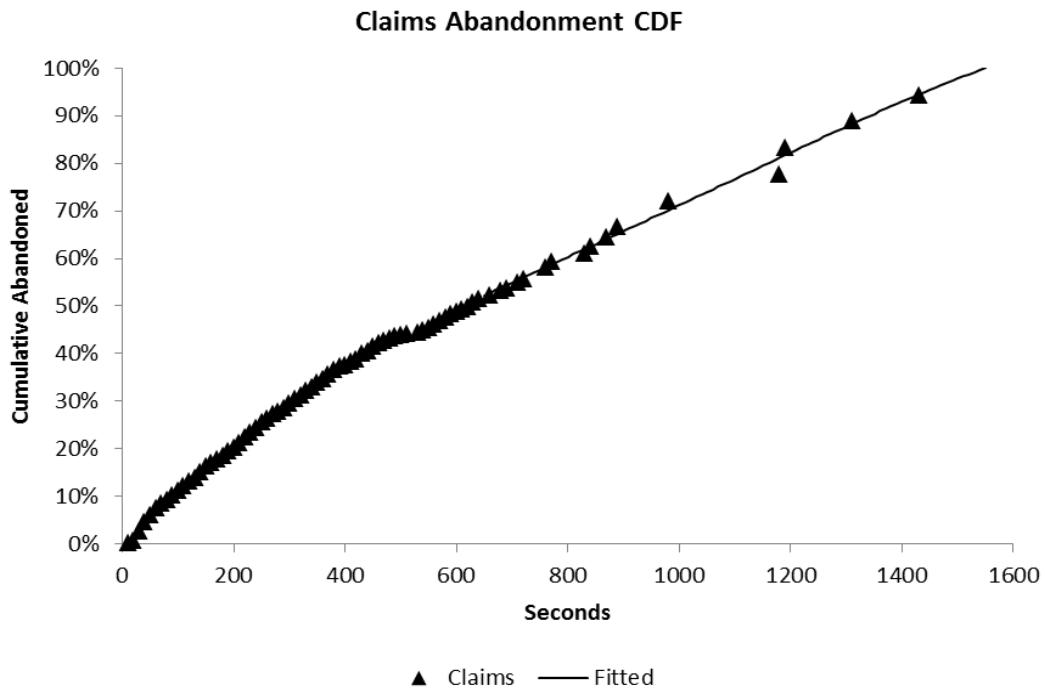


Figure 6: Claims abandonment CDF Estimation

4.2.3 Modeling Approach

Once the CDF for abandonment was established, a method for assigning an abandon time to each entity based on the empirical distribution was needed. Rockwell Automation's Arena simulation software (Rockwell 2012) was used to model the call center due to availability and author familiarity.

In order to assign abandon times to entities, use two variable arrays:

1. The time interval and its associated value of the CDF for abandonment from the Kaplan-Meier analysis. The author used ten second intervals.
2. The time interval and its associated abandon time. The author used the midpoint of the interval for the abandon time.

The method used for assigning the abandon time has three steps.

1. Draw a random number from a uniform (0,1) distribution and assign the number to the entity
2. Search from one to the number of time intervals in variable array 1 for the random number from step 1 that is less than or equal to the CDF for abandonment of the current search time interval. J will be set to the first time interval for which the search is true.
3. Assign an attribute the value of J from step 2, and use the attribute to index variable array 2 and assign an abandon time associated with that index

The method for executing the entity abandonment is a disjointed network with seven steps as seen in figure 7. It has the capability to remove multiple entities from the queue before delaying for the next queue check interval.

1. Generate a control entity
2. Search the phone queue, from one to the number of entities in the queue, for an entity that has a wait time greater than its assigned abandon time
3. Route the control entity based on the current value of J from the search in step 2. If an entity is found in step 2, then J is not zero and the control entity is sent to step 4. If an entity is not found in step 2, then J is set to zero and the control entity is sent to step 7.
4. Remove the entity in the phone queue with rank J and send it to step 5. Route the control entity back to step 2 to re-check the phone queue for additional entities that should be removed based on their wait time and abandon time.
5. Tally the actual abandon time for the entity removed from the phone queue and send to step 6
6. Dispose the entity removed from the phone queue
7. Delay control entity until the next queue inspection

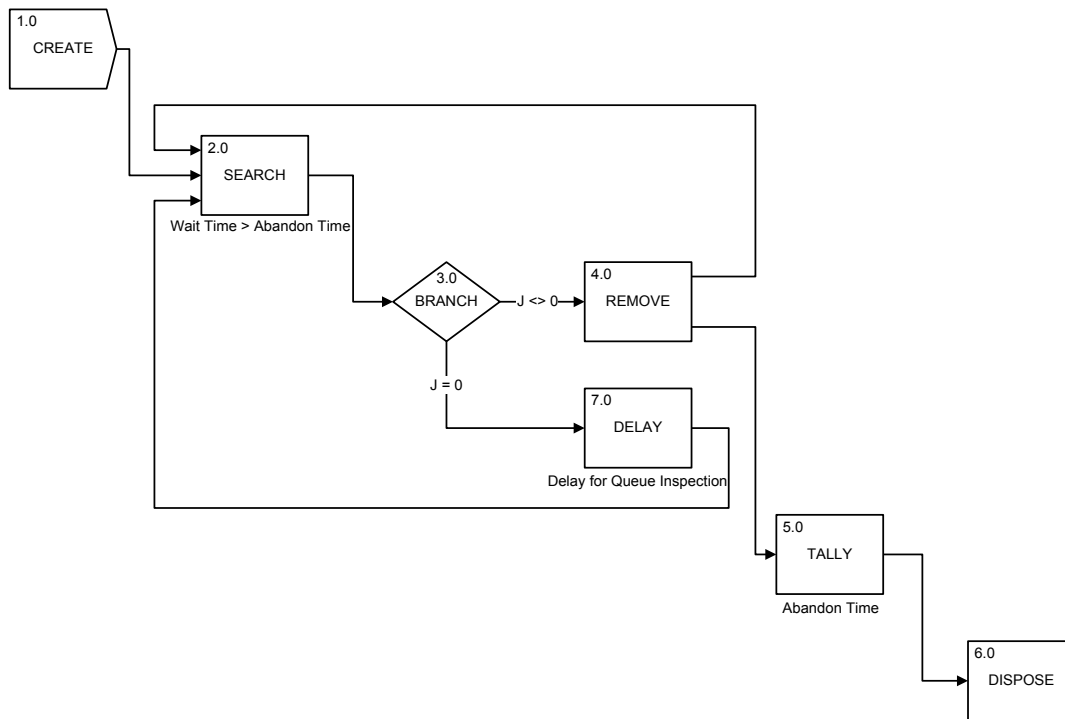


Figure 7: Abandonment Execution Process

4.3 Simulating Using Abandoned Only

For comparison purposes, the abandonment CDF was also derived based on abandoned data only. Therefore, answered calls were not accounted for and Kaplan-Meier survival analysis was not used. The modeling approach used was the same as stated in section 4.2.3. The results are displayed in Figures 10 and 11 in section 5.

5 RESULTS

The validation results for percentage abandoned and service level were statistically equivalent based on a 95% confidence interval. The percent abandoned confidence intervals simultaneously overlap for both billing and claims as shown in Figure 8.

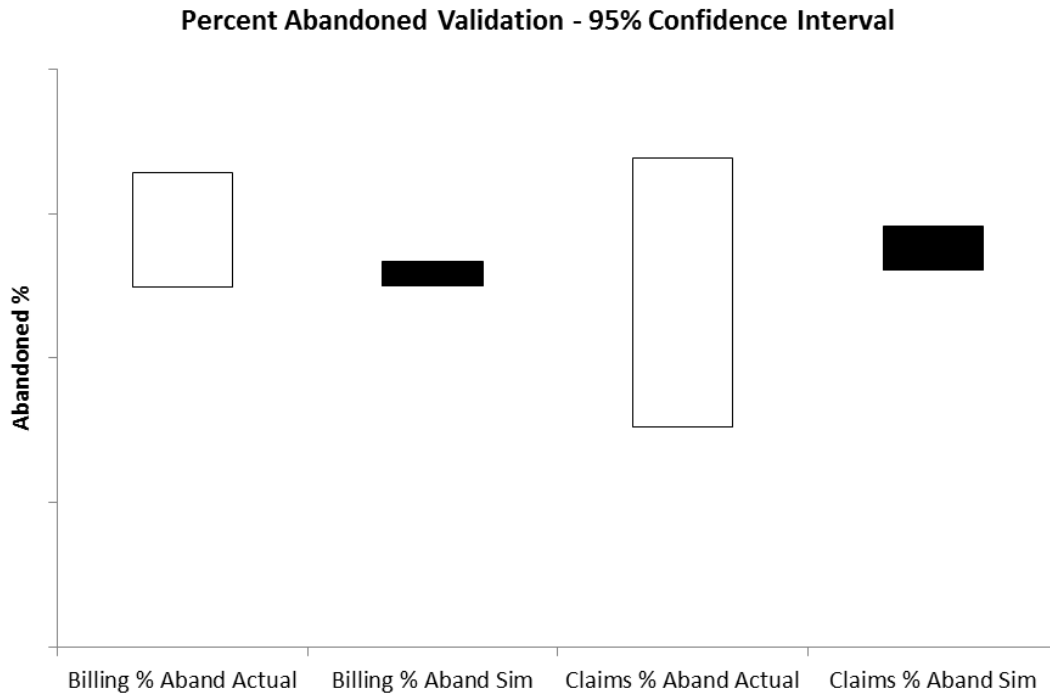


Figure 8: Validation of Abandonment Percentage

The service level confidence intervals simultaneously overlap for billing and claims as shown in Figure 9.

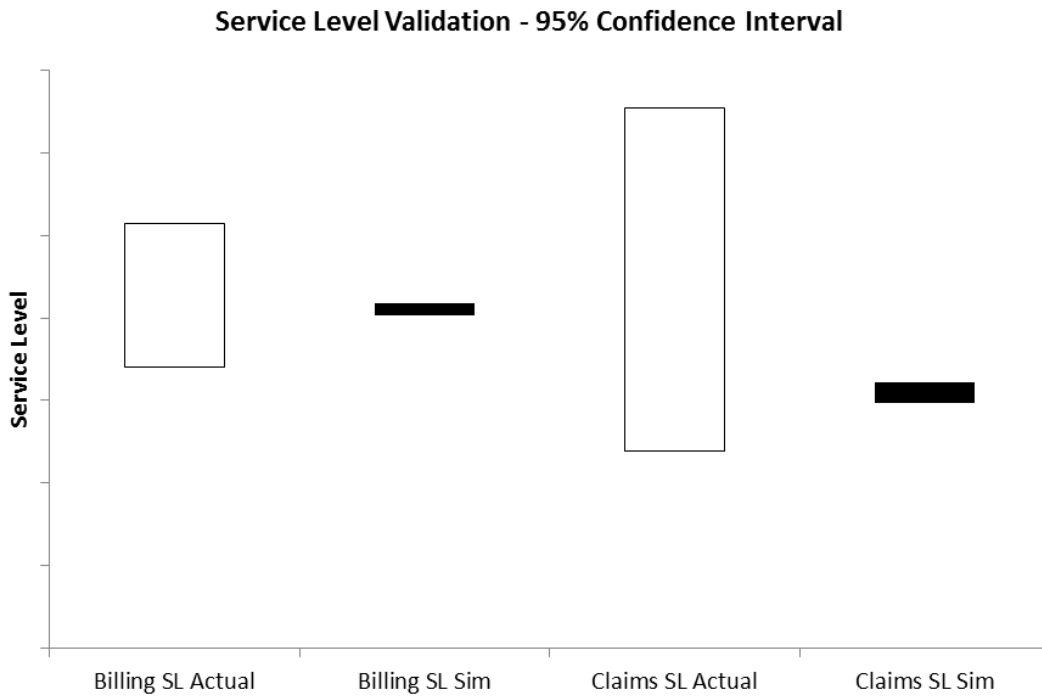


Figure 9: Validation of Service Level

Using the abandoned data only understates the willingness of callers to wait, as demonstrated in Figure 10 by the large spread between the confidence intervals of the actual data and the simulation output.

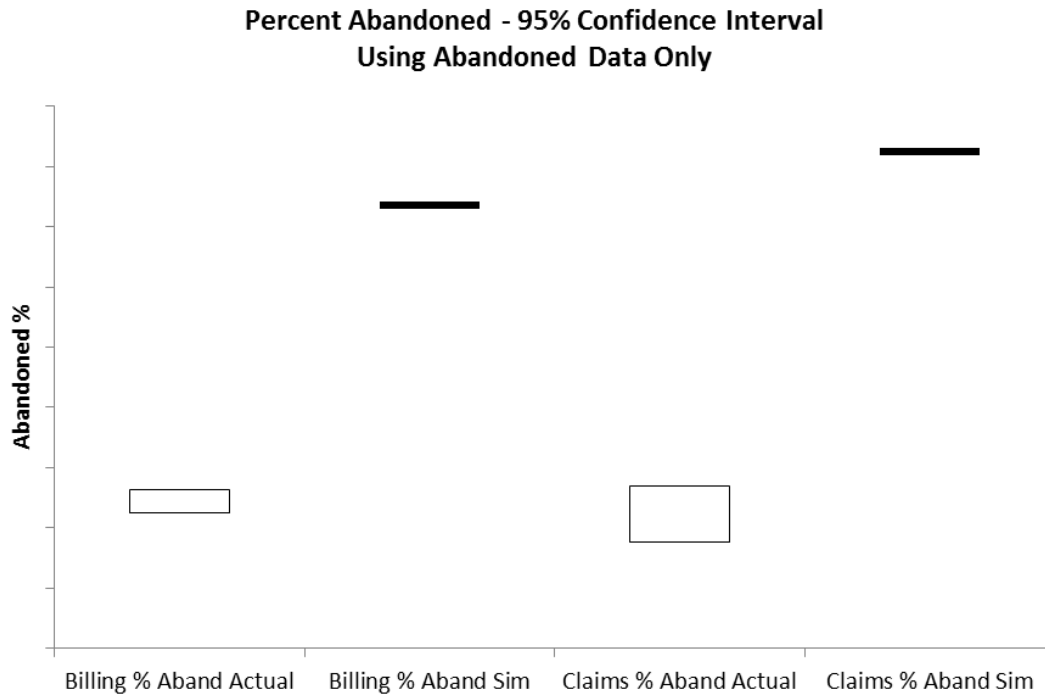


Figure 10: Abandonment Percentage Simulation Results Using Abandoned Data Only

Using the abandoned data only inflates the service level by removing calls that waited from the denominator in the service level calculation. This is demonstrated in Figure 11 through the large spread between confidence intervals and the artificially high service level output from the simulation.



Figure 11: Simulation Results Using Abandoned Data Only

6 CONCLUSIONS

Kaplan-Meier survival analysis is an accurate and efficient method to model abandonment in a discrete event simulation model. By including both the answered and abandoned observations, the analyst obtains a full picture of the actual patience demonstrated by callers. Using this full picture as the CDF for abandonment time enables the simultaneous validation of dependent call center metrics.

REFERENCES

- Conley, Q. D., M. R. Grabau. In press. Simulating a Modified Hybrid Approach to Resource Assignment in a Shared Billing and Claims Call Center. In *Proceedings of the 2013 Winter Simulation Conference*, Edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Kaplan, E. L., P. Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53: 457-481.
- Lee, E.T., J. W. Wang 2003. *Statistical Methods for Survival Data Analysis*. Hoboken: John Wiley & Sons Inc.
- Rockwell Automation, Inc. 2012. Arena (Version 14.0). Milwaukee, WI.
- Roubos A., O. Jouini. 2012. "Call Centers with Hyperexponential Patience Modeling." *International Journal of Production Economics* 141 (1): 307-315.
- Zohar, E., A. Mandelbaum, N. Shimkin. 2002. "Adaptive Behavior of Impatient Customers in Tele-Queues: Theory and Empirical Support." *Management Science* 48 (4): 566-583.

ACKNOWLEDGEMENT

The author would like to thank Mark Grabau for his inspiration to find and implement creative solutions to complex problems.

AUTHOR BIOGRAPHY

QUINN CONLEY is a Business Intelligence Analyst in the Analytics Resource Center at Westfield Insurance. He currently works as an internal, advanced-analytics consultant, focusing on discrete event simulation and predictive modeling. Prior to his current role, Mr. Conley provided analytics support in the rate and profitability management of Westfield's non-standard auto products. Mr. Conley is a graduate of Baldwin-Wallace College in Berea, Ohio; with a B.S. in Mathematics. He will be graduating from Ashland University in Ashland, Ohio in Fall, 2013 with a Master's in Business Administration. His email address is quinnconley@westfieldgrp.com.