# Final Project Report:

## Impact of Diagnosis Time and Demographic Features on Sepsis Mortality

---

CSE-40817 Healthcare Analytics (Fall 2018)

Seth Cattanach and Tommy Clare

12 December 2018

## I. Introduction:

This study seeks to identify exactly how much benefit comes with identifying sepsis sooner. Through several methods of analysis, the goal is for these results to be used by doctors to determine whether an initial sepsis examination would be worth the time and the resources if it meant the difference between a significant number of patients surviving instead of dying from sepsis.

Qualitatively, the questions of, "Do factors like race, age, weight, and sex have an impact on the median survival time of an group of individuals infected with sepsis?" and "By how much does the median survival time of a group of individuals increase if their case of sepsis had been caught and diagnosed from the start of their stay in the ICU?" will be answered by evidence captured by this project's interactive Kaplan-Meier feature.

Quantitatively, the questions of, "Is the difference between the median survival time of an individual statistically different from what it would have been if their case of sepsis had been known from the beginning of their ICU stay?" and "Are gender, age, weight, sex, and ethnicity each significant in a Cox Regression model that predicts the change in hazard ratio for sepsis mortality?"

These questions raise interesting considerations about the tradeoffs between spending time and resources to conduct (possibly unnecessary) sepsis tests on patients, and the advantages that could come from realizing and treating sepsis sooner. The answers may provide justification for running a sepsis test on patients with related diseases, such as metastatic cancer and diabetes, in addition to other prerequisite tests that would normally be conducted.

## *II. Related Work:*

The project was initially inspired by an analysis of sepsis *identification* and how the criteria for sepsis diagnosis affected treatment and mortality risks[1]. This resource (and its linked Git repository) provided the initial datasets upon which this project was completed. A similar

---

[1] *A Comparative Analysis of Sepsis Identification Methods*

study conducted in Germany in 2007 guided this project toward analyzing *demographic* features, as this study controlled for age and sex when analyzing sepsis mortality rates[2].

Similar research has been performed using survival analysis for specific treatment methods, such as a 2017 study on sepsis mortality and "continuous renal replacement theory"[3] - this project, though it uses similar analysis methods (namely, survival analysis and Cox Regression), differs by analyzing diagnosis time as well as health and demographic features instead of specific treatment methods.

Especially in the last two decades, sepsis mortality in cases of severe diagnosis has been rapidly increasing[4] in the United States, leading researchers to question whether these increased rates of mortality and hospitalization are due to factors like delayed diagnosis times or ineffective treatments. This project attempts to answer, in part, whether delayed diagnosis time has any influence on mortality rates (though it does not address hospitalization rates, as does the referenced study[5]. Another study from 2014 concludes that empiric antibiotic treatment is an effective treatment method for severe sepsis *within the first hour of septic shock*[6], further demonstrating the significance of diagnosing sepsis at the earliest possible stage; in contrast, this project demonstrates how significant mortality risks might vary with a late diagnosis time using survival analysis.

One of the more similar studies suggested sepsis mortality risk rises at least 4% for each hour treatment is delayed,[7] although this study didn't account for an important aspect of this project - whether this mortality risk varied significantly on the account of age, gender, weight, etc.

Other studies have alluded to the particularly-devastating effect of sepsis with respect to age, suggesting the elderly population to be especially vulnerable to sepsis[8]. This appears to be

---

[2] *Hospital Incidence and Mortality Rates of Sepsis*
[3] *Biomarkers Predicting Survival of Sepsis Patients Treated with Continuous Renal Replacement Therapy*
[4] *Rapid Increase in Hospitalization and Mortality Rates for Severe Sepsis in the United States*
[5] *Rapid Increase in Hospitalization and Mortality Rates*
[6] *Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock From the First Hour*
[7] *Time to Treatment and Mortality during Mandated Emergency Care for Sepsis*
[8] *Severe sepsis and septic shock in the elderly: An overview*

due to several factors, among which is the effectiveness of the immune system with age[9]. With this in mind, the project aims to account for the significance of age in terms of sepsis mortality.

III. Solution/Method:

This project conducts survival analysis on the patients from the chosen dataset. In an effort to predict the probability of a patient surviving after a given number of days in the hospital, several constants are defined. First, the unambiguous time origin is chosen to be the time at which the first patient was admitted to the ICU. Next, the uniform time scale is set to be days, on the order of the patients' ICU stays. Finally, the event in question is sepsis mortality.

Kaplan Meier curves, which visualize the probability of an event not occuring (in this case, a patient not dying from sepsis) over the time of a patients' stay in the ICU, were used in combination with a user-friendly drop down menu that allows for the control of several factors. An explanation of this feature:

1. The drop down GUI itself allows a user to see the predicted Kaplan Meier probability-of-survival curve for a race, age, weight, and gender of the user's choice. It also outputs the predicted median survival time for a user-chosen combination of demographic features. The primary goal of this GUI was to identify trends that would be more influential in predicting sepsis survival probability than others.
2. A Kaplan Meier curve is generated for both the *actual* time to event data, as well as the *ideal* time to event data that would have been if diagnosis time of sepsis was the admission time. A new dataframe column, made by adding diagnosis-time-into-visit to the actual time before sepsis, accomplished this. Effectively, this new column showed the duration the patient *would have been known* to have sepsis if they were tested for it the as soon as they were admitted.

To determine whether there was a statistically significant difference between the *ideal* time before sepsis mortality corresponding to the median probability of survival and the *actual*

---

[9] *Sepsis and Immunosenescence in the Elderly Patient*

time before sepsis mortality corresponding to the median probability of survival, a two-sample, paired, one-sided t-test was conducted. Forty sample points were taken from the Kaplan Meier curve generator; each sample point consisted of the median times before sepsis mortality (in the actual and ideal cases), paired together. The resulting p-value was used to determine if there was a statistically significant increase in the median time to sepsis mortality for patients with diagnosis time taken into account.

Cox Proportional Hazards (CPH) regression models were used to evaluate the validity of assumptions about each individual demographic feature derived from these intuitive, demographic-featured-adjusted Kaplan Meier curves. The statistically significant factors were taken from the summary of the results, and these significant features were then analyzed using covariate group plots to highlight significant differences in probability of survival for varying values of these features.

## IV. Data and Experiments:

The data was gathered from the GitHub repository of Alistair Johnson, who hosts the code repository associated with *A Comparative Analysis of Sepsis Identification Methods in an Electronic Database*. The repository outlines steps to obtaining .csv files that contain information about sepsis-infected patients, diagnosed on seven different levels of increasing strictness.

Various preprocessing methods were done before statistical analyses were performed:

1. Patients with metastatic cancer and diabetes were removed from the dataset, as these conditions may have had a confounding effect on the dependent variable of interest, time to sepsis mortality.
2. Patients with null values in controlled factors of age, weight, race, and ethnicity were removed from the dataset. For a study focused on the effects of demographic variables on probability of survival, a patient lacking one or more of these traits should be removed.

3. Patients with null or negative values in *suspected_infection_time_poe_days* (that is, the duration after ICU admission (if any) that it took for a patient to be diagnosed with sepsis) were assigned a value of 0. These patients, essentially, were admitted for sepsis, and were given no diagnosis time since this was the basis of their admission.
4. The durations for how long a patient *actually* stayed in the hospital and how long they were *known* to have sepsis (interpreted as their *ideal* time to event data) were each computed as separate columns in the dataframe.
5. Finally, before fitting a Cox Regression model, the *ethnicity* feature values for all patients in the dataset were mapped to one of five categorical values: *White, Asian, Black, Hispanic,* or *Other*.

V. Evaluation and Results:

Running the Kaplan Meier curve GUI allowed us to compare the survival probabilities over *actual* time to mortality and *ideal* time to mortality at once, all the while accounting for four different demographic factors. An example output is shown below in Figures 1-3.

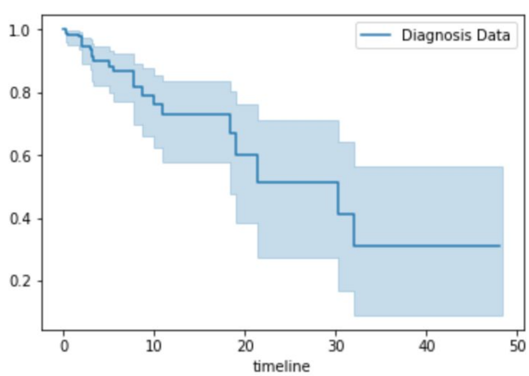


Figure 1. Kaplan Meier Curve For Patients: Race = White, Age ≥ 10, Weight ≥ 145, Gender = Male.

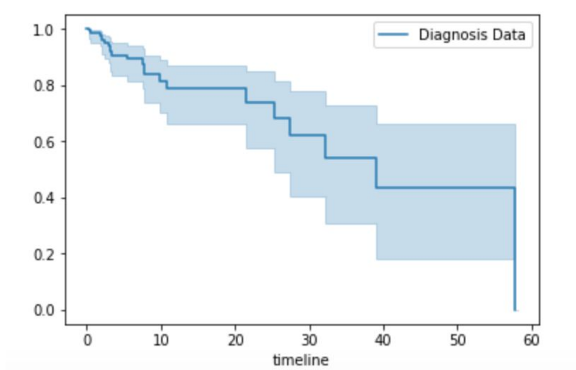


Figure 2. KM Curve With Same Parameters: Event is still *hospital_expire_flag*, but time to event now accounts for sepsis diagnosis time.

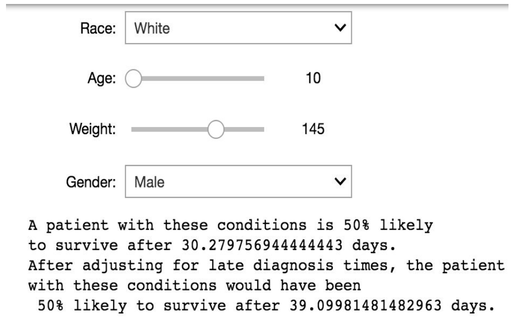


Figure 3. The drop down menu that provided the features shown in Figure 1: Median times of survival corresponding to *actual* and *ideal* time before sepsis mortality.

The example above projects a nearly nine day improvement in the median survival time for patients if sepsis had been diagnosed from the onset of a patient's admission into the ICU.

As mentioned above, forty median survival time sample points like these were collected and used as the data of a two-sample, paired t-test that looked for a significant difference in the median sepsis survival times of patients whose diagnosis times were taken into account. The results of this two-sample test are summarized below:

- *Null hypothesis*: the true difference between the *ideal* time before sepsis mortality corresponding the median probability of survival and the *actual* time before sepsis mortality corresponding the median probability of survival is 0.
- *Alternative hypothesis*: the true difference between the *ideal* time before sepsis mortality corresponding the median probability of survival and the *actual* time before sepsis mortality corresponding the median probability of survival is greater than 0.
- *Test-statistic*: t = 8.94.
- *P-value*: 2.74e-11
- *Conclusion*: Since the p-value, 2.74e-11, is less than alpha=0.05, we reject the null hypothesis. There is evidence in support of the claim that the true difference between the *ideal* median time to sepsis mortality and the *actual* median time to sepsis mortality is greater than 0.

In other words, a typical patient has been statistically significantly proven to survive for a longer amount of time if their sepsis was diagnosed as soon as they were admitted to the hospital. Though Cox Regression analysis later showed this factor had a relatively small influence, this result is still profound: it should give doctors insight into the benefits of spending the time and resources necessary to run a sepsis test as soon as patient with related symptoms enters the ICU.

As for the significance of each individual feature, a basic, qualitative view of the outputs of the Kaplan Meier GUI suggests differences in *both* ideal and actual median time to event data with changes in demographic factors like race and gender. At first, both categories of time to event showed longer durations (in most cases) for white patients than any other race, and longer durations for male patients than female.

*The implications from these points could be significant.* Biological factors aside, this could mean a bias of the attention of medical workers to male patients, or white patients. A Cox proportional hazards model was fit and analyzed to identify any such potential bias.

The results showed that the only statistically significant features in the model were age and weight. Figures 4 and 5 show plots of covariate groups for these two features. On both plots, each line shows a survival curve for all patients *greater than or equal to* a specific age and weight. Note that both curves follow the assumption of proportional hazards, since survival curves of different ages and weights do not intersect and appear proportional.

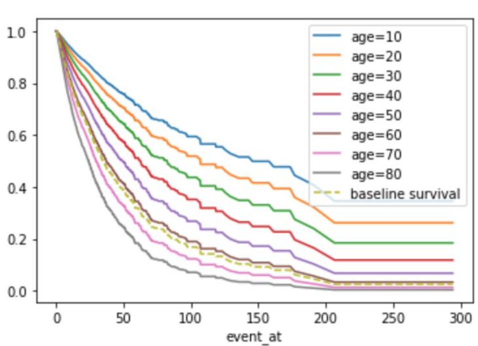


Figure 4. Covariate CPH analysis of age

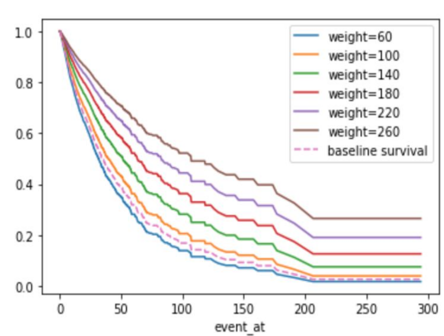


Figure 5. Covariate CPH analysis of weight

## VI. Conclusions

As the median time of survival from admission to a hospital is statistically greater when a patient is known to have sepsis from the onset of admission, this should be evaluated against the costs of staffing, conducting, and reviewing initial sepsis tests for patients exhibiting symptoms of sepsis (or related diseases). Further, because of the statistical significance of age and weight on predicting sepsis mortality, doctors should especially cognisant of a septic infection and related conditions in older patients of lower weight.

**Works Cited**

Dombrovskiy, V. (2007). Rapid increase in hospitalization and mortality rates for severe sepsis in
the United States: A trend analysis from 1993 to 2003. *Society of Critical Care Medicine*. Retrieved from https://journals.lww.com/ccmjournal/Abstract/2007/05000/Rapid_increase_in_hospitalization_and_mortality.4.aspx.

Ferrer, Ricard et al. (2014). Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock From the First Hour. *Critical Care Medicine*. Retrieved from http://www.intensivo.sochipe.cl/subidos/catalogo3/Antibioticos reducen mortalidad en primera hora Shock septico CCM 2014.pdf

Fleischmann, C. (2016). Hospital Incidence and Mortality Rates of Sepsis. *US National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4814768/.

Ho Lee, Jeong & Yeon Kim, Ha & Bae, Eun & Kim, Soo & Kwon Ma, Seong. (2017). Biomarkers Predicting Survival of Sepsis Patients Treated with Continuous Renal Replacement Therapy. Chonnam Medical Journal. 53. 64. 10.4068/cmj.2017.53.1.64.

Johnson, A., & Aboab, J. (2018). A Comparative Analysis of Sepsis Identification Methods in an Electronic Database. *US National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29303796.

Prashant Nasa, Juneja, Deven, & Singh, Omender (2009). Severe sepsis and septic shock in the elderly: An overview. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3956061/.

Seymour CW et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med 2017 May 21; [e-pub]. http://dx.doi.org/10.1056/NEJMoa1703058

Silvia, M., & Perez, A. (2017). Sepsis and Immunosenescence in the Elderly Patient: A Review.

*US National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5329014/.