Variability of Decision-Making for Treating Critically Ill Patients In Septic Shock - Comparisons Of Observational Data From Different Intensive Care Units (ICU)

Naftali Arnold, David Orenstein, Eytan Slotnik
Guided by Prof. Michael Beil and Mr. Gal Hyams

October 6, 2023

1 Abstract

A common setting in an intensive care unit (ICU) is the treatment of septic shock, indicated by low mean arterial pressure (MAP), with norepinephrine (NOR), a drug that increases blood pressure (BP). Presently, the endorsed guidelines strongly support this approach, with the treatment strategy predominantly based on NOR infusions. However the specific treatment strategy, i.e. timing and dosage, remain controversial.

Guidelines propose that the NOR infusions will try to attain some target MAP of 65 mmHg. In practice, the target and overall strategy are frequently altered by the staff based on parameters such as the patient's MAP history, other vitals and many other considerations. For instance, medical staff may choose to "overshoot" when treating a patient under specific conditions, such as concomitant brain oedema.

This common scenario gives us with a unique opportunity to analyze the decision-making process of practitioners within confined state and action spaces. Previous work have examined the existence of variability in treatment strategies. In this study, we will use the higher time resolution of the MAP measurement time in the SICdb and eICU databases to better identify variability and determine if MAP is the predominant or only a factor in the treatment strategy, i.e. to define the time and dosing of NOR.

Our results show substantial variability of treating septic shock. Further investigations are warranted to examine whether the same policy was used along temporal trajectories of indivudual patients.

2 Introduction

2.1 Background

2.1.1 Variability in Medical Practice

Variability of patients as well as of therapies is an inherent aspect of medical practice. Medical practitioners often have to diverge from standardized (protocolized) treatment suggested in guidelines due to a variety of factors, including patient-specific considerations, e.g. the specific combination in an individual patient or his/her personal preferences, as well as logistical or public health constraints [4]. On the other hand, the variability of the medical treatment could also signify that a certain percentage of the patient population is receiving sub-optimal treatment, wasting resources on ineffective or even harmful approaches. The consequences of such 'negative' variability are extensively discussed in other work. . [3]

The ability to detect and analyze variability in practice could provide us with valuable insights into the quality of medical care in distinct units or hospitals.

2.1.2 Guidelines for treating sepsis

This notion of variability as part of medical practice gets a unique test case in the treatment of sepsis and septic shock in intensive care units (ICUs). Sepsis is a life-threatening condition caused by an infection that triggers an exagerrated immune response leading to the failure of vital organs. Part of the clinical presentation can be low blood pressure (measured as mean arterial blood pressure - MAP) that is insufficient to ensure the perfusion of organs with blood and, hence, oxygen supply. If left untreated, this situation can lead to imminent death. Thus, managing low blood pressure in sepsis, aka 'septic shock' is an essential part of the guidelines for treating sepsis [5]. It is based on administering drugs, known as vasopressors' such as noraderenaline (NOR), which increase blood pressure within a minute. The other pillar of treating septic shock is 'fluid resuscitation', i.e. the intravenous administration of large volumes of crystalline solutions. However, the effect of this intervention is measurable at a different time scale (>3-5 minutes). Thus, in situations with dangerously low blood pressure (MAP $<$ 65 mmHg), the fastest intervention to restore organ perfusion in sepsis is injecting vasopressors. The specific NOR rate is primarily determined by the patient's MAP and is usually linearly normalized with the patient's weight.

The MAP plays a crucial role in the decision-making process for treating septic shock, to the extent that the treatment strategy for life-threatening low MAP could be conceptualized within a state-action framework, with the infusion rate of NOR serving as the response to MAP measurements. This configuration offers an optimal setting for testing the variability of treatment, given the relatively compact and one-dimensional nature of both state and action spaces.

2.1.3 The Reality of Treating Septic Shock

In practice, the official guidelines regarding the treatment of sepsis and septic shock, especially the use of vasopressors to achieve a certain MAP lack robust evidence and is mostly based on opinions which are not shared by all experts.[5][6] Thus, the actual treatment of septic shock

is expected to vary in clinical practice, according to the preferences of individual doctors, with a still unknown relationship with outcome. Based on studies in intensive care, it is fair to expect an impact of this variability on mortality.[7]

While this statement highlights the necessity to establish a way to assess adherence to guidelines, a more fundamental question arises: Are there patterns of adherence / non-adherence to guidelines with or without biological justification? A positive answer to this question, coupled with the observation that the official guidelines aren't being followed, implies the possibility of distinct treatment strategies across different ICU types treating different types of patient populations with different combinations of diseases. Furthermore, it's crucial to recognize that while attributing all variability to a treatment strategy might be convenient, it's quite likely that a portion of the variability stems from the variability among doctors within the same ICU.[8]

Given that managing low blood pressure is regarded as a crucial part of the treatment of septic shock, the MAP-NOR relationship holds significant information concerning variability of fundamental treatment strategies. As previously mentioned, this data is easily processable, enabling attainable variability analysis.

2.1.4 Variability Analysis Data Requirements

The objective of this work is to demonstrate that the treatment policies depend on the provided MAP measurements. Naturally, consistent and accurate measurements of MAP are essential. Furthermore, any analysis of the MAP-NOR relationship will require knowledge of the patient's weight, since the NOR rate is normalized by this value. We will go over the challenges associated with meeting each of these requirements.

2.1.4.1 MAP Measurements Time Resolution Using the latest gold standard medical database, MIMIC-IV, would provide us with a time resolution of approximately 60 minutes between each measurement. This is insufficient for our purposes since the half-life of norepinephrine (NOR) is only a minute.

This work approaches the challenge of variability analysis with renewed confidence, thanks to the use of data with higher measurement resolution found in the eICU and SICdb databases. In the eICU, the typical interval between measurements is 5 minutes, while in SICdb it is even more frequent, with a 1-minute interval between measurements.

In this work, resolution will be an essential aspect, although it is important to continually consider the complexity trade off concerning computing resources such as data size and run time.

2.1.4.2 Patient Weight When a medical practitioner decides to administer NOR to a patient, the first step in determining the rate is to decide on a value in units of mcg/min, micograms of NOR per minute. This value can be interpreted as the practitioner's response to the patient's condition. The final infusion rate is adjusted based on the patient's weight, resulting in the actual NOR rate units being in mcg/kg/min,which represents micrograms per minute per kilogram or, in a more accurate mathematical notation $\text{kg} \cdot \frac{\text{mcg}}{\text{min}}$. However, in practice, the weight may not be accurately documented, and as we will explore, a significant number of patient stays lack this value.

Many of the techniques applicable for analysis would necessitate the use of weight normalization, which could result in a significant reduction of the population. To address this challenge, we can explore methods that do not depend on the specific NOR values or rely on the reasonable assumption that the normalization is linear.

2.2 Related Work

The notion of of different types of variability and the attempt to harness that fact to propose better policies was introduced at 2018 in "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care"[1]. A Reinforcement Learning based model was developed to suggest optimal treatment patterns: mainly doses of Vasopressors and IV fluids. The target of the model was to reduce the patients mortality as much as possible, be it in the ICU or 90-day mortatlity. The model included a set of 48 patient variables including demographics, vital signs, laboratory values, fluids and vasopressors received and fluid balance demographics, to vitals signs and laboratory results.

While the paper claimed to achieve lower mortality rates, it also faced criticism and raised concerns about the practical application of the model in clinical settings. One of the critics, in the paper "Does the Artificial Intelligence Clinician learn optimal treatment strategies for sepsis in intensive care?" [11], argued that the focus of the work on long-term rewards does not align with real-world decision-making practices. The review suggested the use of short-term rewards, specifically emphasizing the importance of stabilizing a patient's vital signs and maintaining their MAP within recommended values.

2.3 Data

2.3.1 eICU Database

Critical care patients are monitored closely through the course of their illness. As a result of this monitoring, large amounts of data are routinely collected for these patients. Philips Healthcare has developed a telehealth system, the eICU Program, which leverages these data to support management of critically ill patients.[9] This database is intended to be freely available to support various applications, including the development of machine learning algorithms, decision support tools, and clinical research.

Recorded data include vital signs such as heart rate, heart rhythm, blood pressure, respiratory rate, and more. Additionally, the database contains information related to a patient's relevant medical history and information about drug infusions that have been manually entered by care staff.[9]

An important aspect of the eICU dataset is its thorough documentation of the patient's treatment unit type. This enables us to concentrate on unit types with similar guidelines, such as Medical Intensive Care Units (MICU), Surgical Intensive Care Units (SICU), and Medical-Surgical Intensive Care Units (Med-Surg ICU), while excluding unit types like neurological ICUs, which handle septic shock with different considerations, e.g. on a background of brain oedema

2.3.2 SIC database

SICdb, also refereed to as the Salzburg data set, provides insight to over 27 thousand intensive care admissions, including therapy and data of their preceding surgery. Data was collected between 2013 and 2021 from 4 of the intensive care units in the University Hospital Salzburg, Austria, having more than 3 thousand intensive care admissions per year in 37 beds. The data set is deidentified and contains, amongst others, case information, laboratory, medication, monitoring and ventilation data. SICdb provides highly granular once-per-minute-data for blood pressure.

This work is one of the first analyses with this new data base.

2.3.3 Representing a Patient Stay Trajectory

In the context of identifying trends within large groups, the fundamental unit of the data is the individual patient stay. This typically involves demographics of the patient such as sex, residence, and age, as well as stay specifics from admission reasons to discharge status, and the patient's heart rate, blood pressure, respiratory rate, peripheral oxygen saturation and many other vitals throughout the treatment.

For the purpose of this research, we will restrict the inference, and hence the data on a specific patient stay, to include age, unit treatment type, timed MAP measurements, and timed NOR infusion rates.

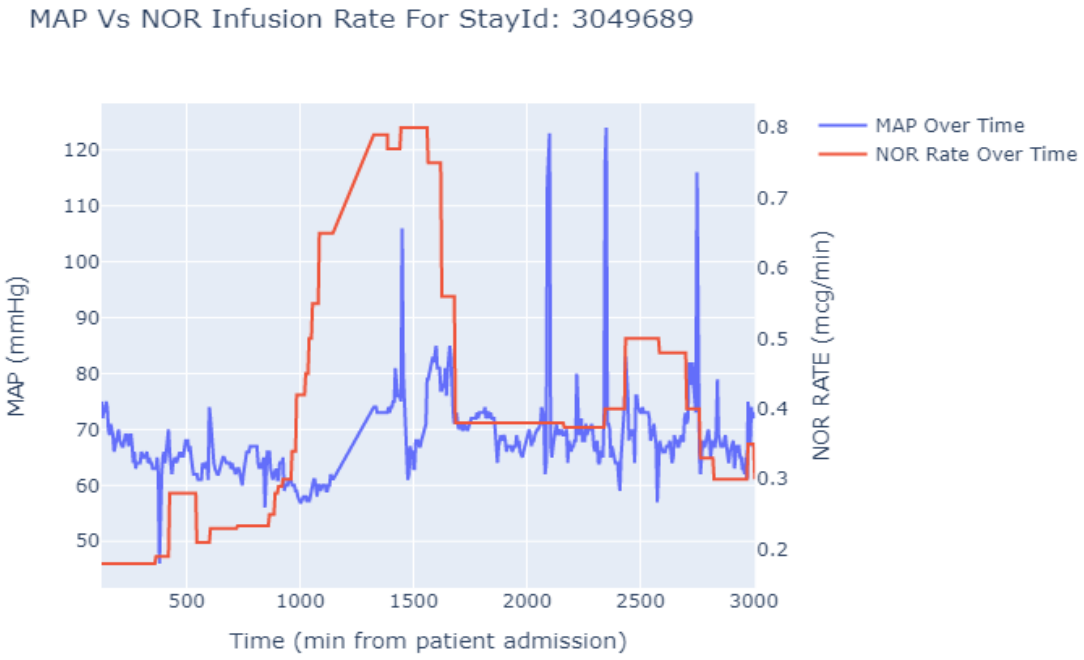

Figure 1: Example of a patient trajectory

3 Results

In our study, we conducted various analyses to investigate the relationship between NOR infusion rates and MAP. The results were computed using two approaches. First, by taking the MAP-NOR data from each group as a whole and computing statistics. Second, the patient stay-wise approach, which involves initially computing some statistic for each patient, usually a scalar value, and then analyzing the distributions of these different statistics among the units.

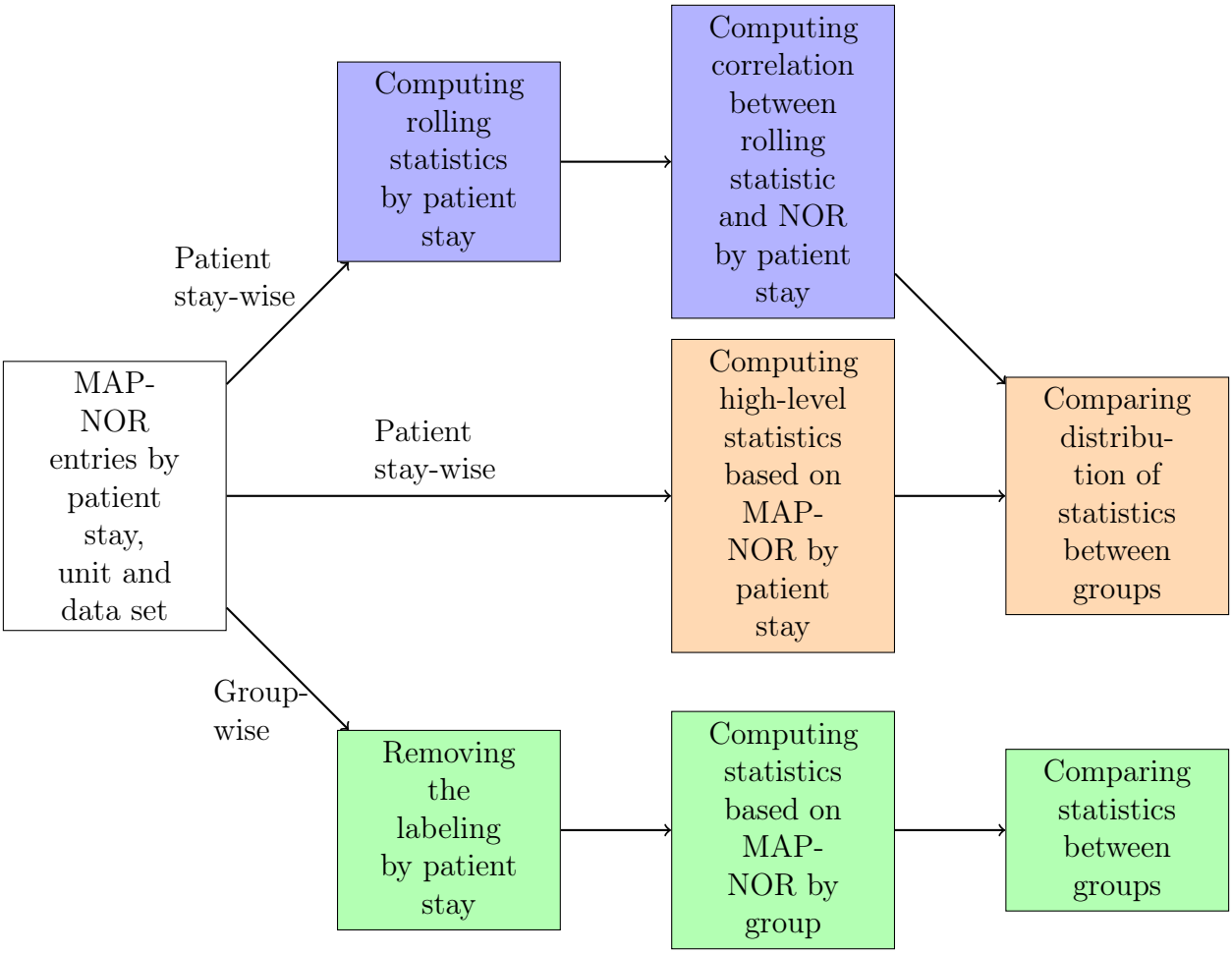

Figure 2: Pipe line for analyses based on groups and patient stays

3.1 MAP-NOR Per Group

This section will focus on analyses that were conducted per group. Focusing on values in the range of 30-80 for MAP and 0.05-50 for NOR rates, these are the typical ranges for patients in septic shock. Among units that share similar guidelines, we primarily focused our analysis on the Med-Surg-ICU and MICU units due to the significant patient populations in these units. Within these units, we conducted a more in-depth investigation of the actual decisions to change the NOR infusion rate, either by increasing or decreasing.

The motivation for binning the entries regardless of the patient stay is the biological

model assuming that the NOR is primarily influenced by the MAP and should trigger a similar decision-making process across all patients. This process determines the mgc/min value of NOR, which is then normalized by weight, which is patient-specific. Therefore, to recover the mcg/min value, which is distributed similarly within each group, we need to use NOR/weight, as the NOR rate is determined by $\text{kg} \cdot \frac{\text{mcg}}{\text{min}}$.

3.1.1 MAP-NOR Rate Pearson Correlation

We initiated our analysis by computing the Pearson correlation coefficient to examine the relationship between MAP and NOR Rates. The resulting Pearson correlation coefficient was observed to be close to zero. This near-zero correlation suggests that there is little to no linear association between MAP and decisions for a specific NOR rate in our data-sets.

3.1.2 MAP-NOR Distribution

In light of the initial findings of the Pearson's correlation analysis, our next objective was to assess the density distribution of these combined values of MAP and NOR and visualize them using a heat map.

Furthermore, we explored the association between NOR infusion rates and MAP values by identifying the MAP value with the highest frequency for each NOR rate. To present these findings effectively, we created plots where the weights were used to represent the percentage of patients who received a specific NOR rate relative to the total number of patients in the study.

This analysis used 2 subsets of the entries -

i. Positive Changes - Entries where the decision to increase the NOR rate, i.e. a higher value of NOR in mcg/min than the previous entry. Usually triggered by lower than expected MAP.

ii. Negative change - Similarly, entries where the decision to decrease the NOR rate, i.e. a lower value of NOR in mcg/min than the previous entry. Also known as "weaning", slow reduction of NOR when MAP is higher than expected.

These considerations concern decisions to increase the NOR rate when the patient is in septic shock, i.e. MAP is below 65 mmHg, which is life-threatening and requires optimized interventions, i.e. quick but not overshooting actions. Decreasing the NOR rate if the MAP is perceived as too high, usually above 65 mmHg, is based on different thoughts: (i) a too drastic reduction of the NOR rate may put the patient back into shock and (ii) it remains unclear if the criteria for the target MAP and for fast action can be relaxed if the patient is now more stable. The latter phase, thus, requires, a different decision-making, that is expected to vary much more than those for the first scenario.

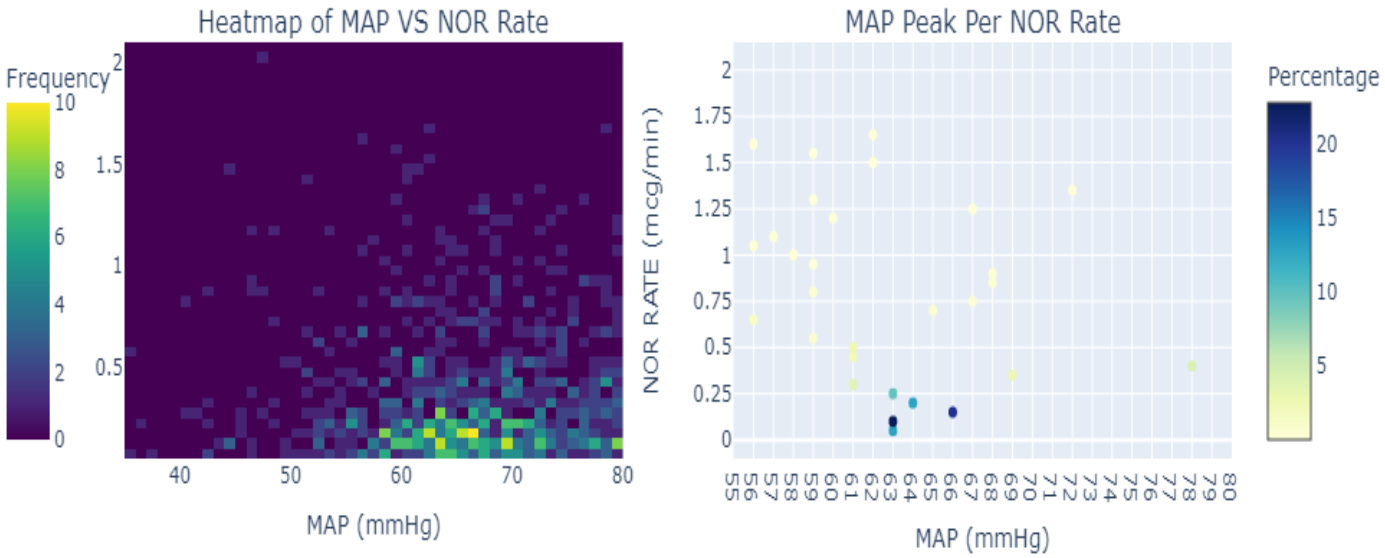

Figure 3: Heatmap of MAP Vs NOR rate. Med-Surg-ICU & positive changes in NOR rate

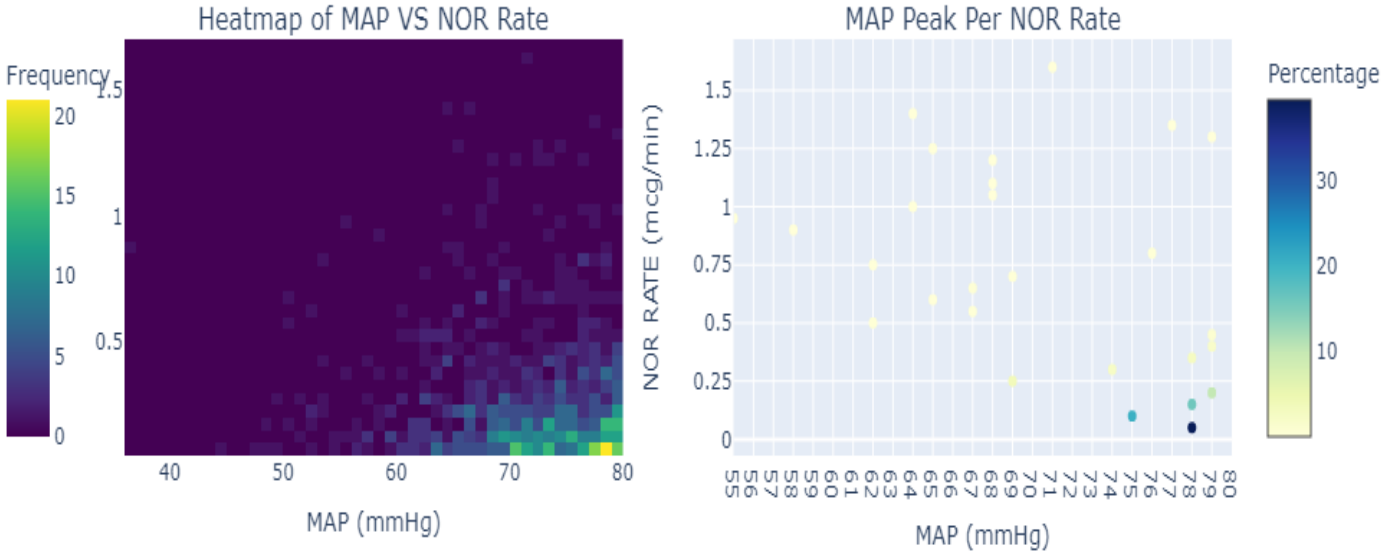

Figure 4: Heatmap of MAP Vs NOR rate. Med-Surg-ICU & negative changes in NOR rate

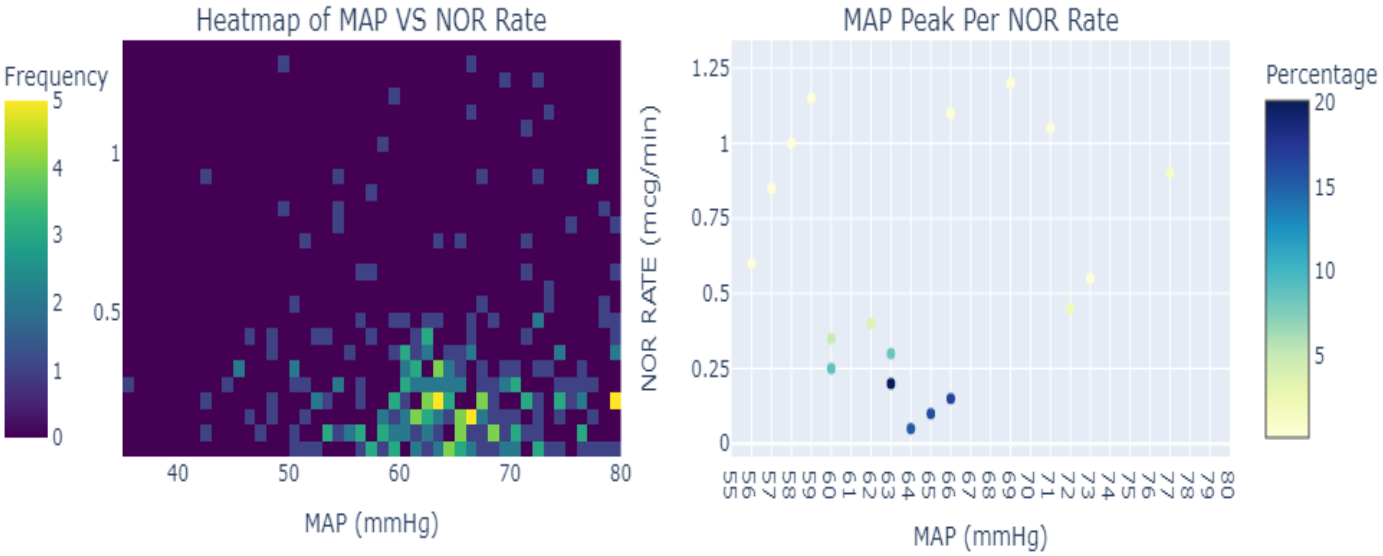

Figure 5: Heatmap of MAP Vs NOR rate. MICU & positive changes in NOR rate

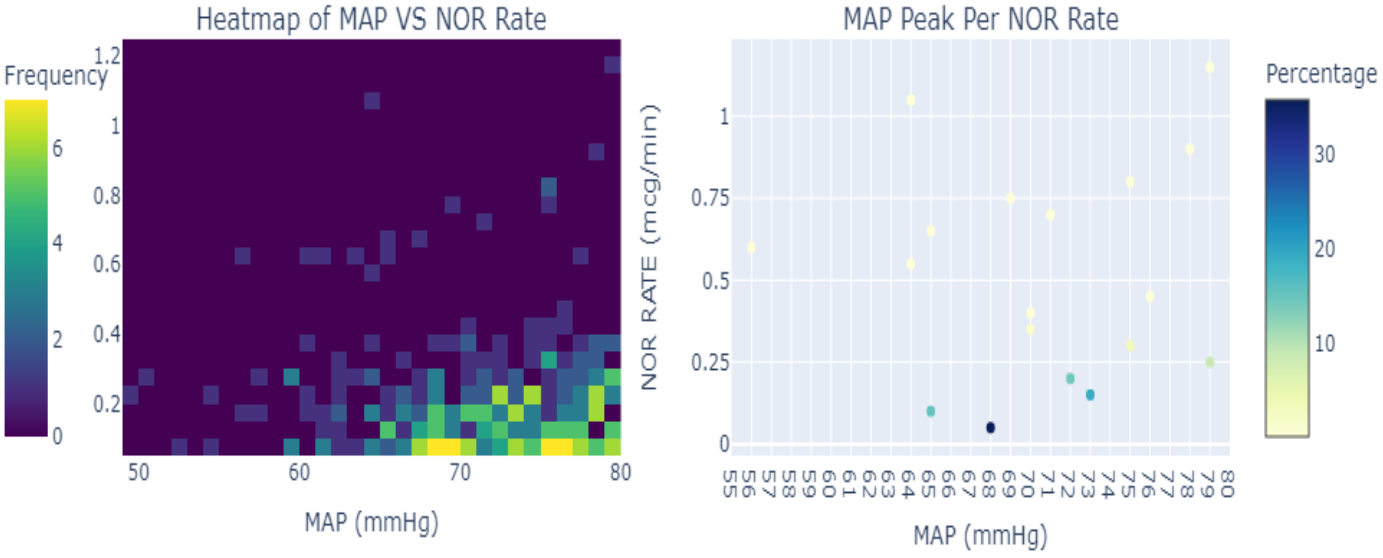

Figure 6: Heatmap of MAP Vs NOR rate. MICU & negative changes in NOR rate

In the previously presented plots, we observe a notable adherence to clinical guidelines within both the Med-Surg-ICU and MICU settings. Specifically, practitioners tend to adjust NOR infusion rates in pursuit of a target MAP around 65-66 mmHg. Conversely, NOR rates are decreased when MAP values surpass this threshold. However, it's worth noting that the level of adherence, though notable, falls short of the expected extent, i.e. close to 100% in this setup. We observe instances, consisting of about 1% of the values, where NOR infusion rates are increased even when MAP values are above 70 mmHg.

3.1.3 MAP-NOR Model Fitting

In the subsequent phase of our results analysis, following the processing done in the previous section (including binning the entries by positive and negative change, and the identification of the most frequent decisions), we undertook a deeper examination of these specific decisions. In order to visualize the relationships between NOR Rate and MAP, we employed a two-step process. First, we took the scatter plots presented in the previous section (Figures 3-6) and transposed them, effectively interchanging the axes. This transformation resulted in positioning the NOR Rate along the x-axis and MAP along the y-axis. Subsequently, we applied polynomial models, specifically considering degrees 1 to 3.

Once we had this new representation of our data, we proceeded to fit polynomial models

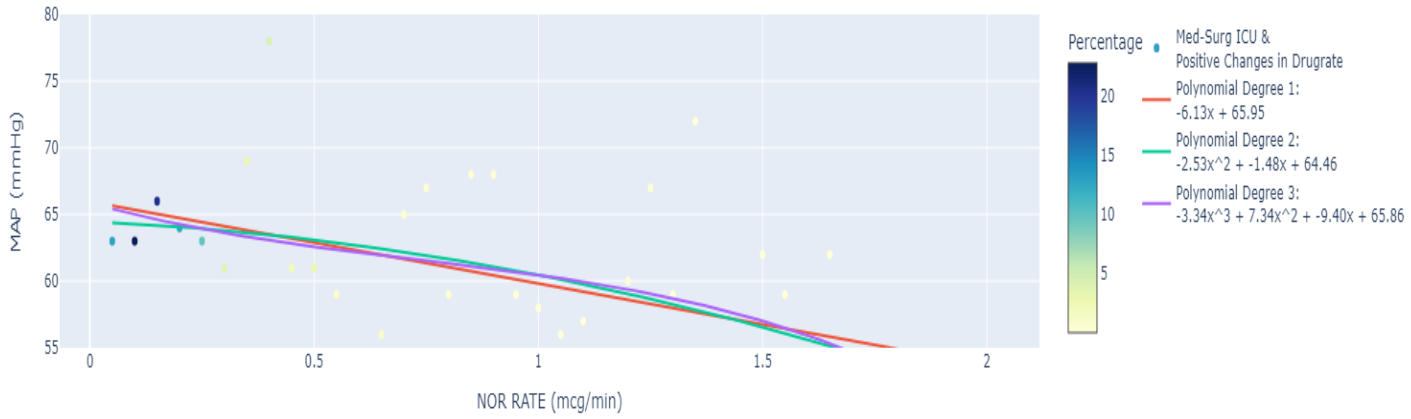

Figure 7: Polynomial fitting. Med-Surg-ICU & positive changes in NOR rate

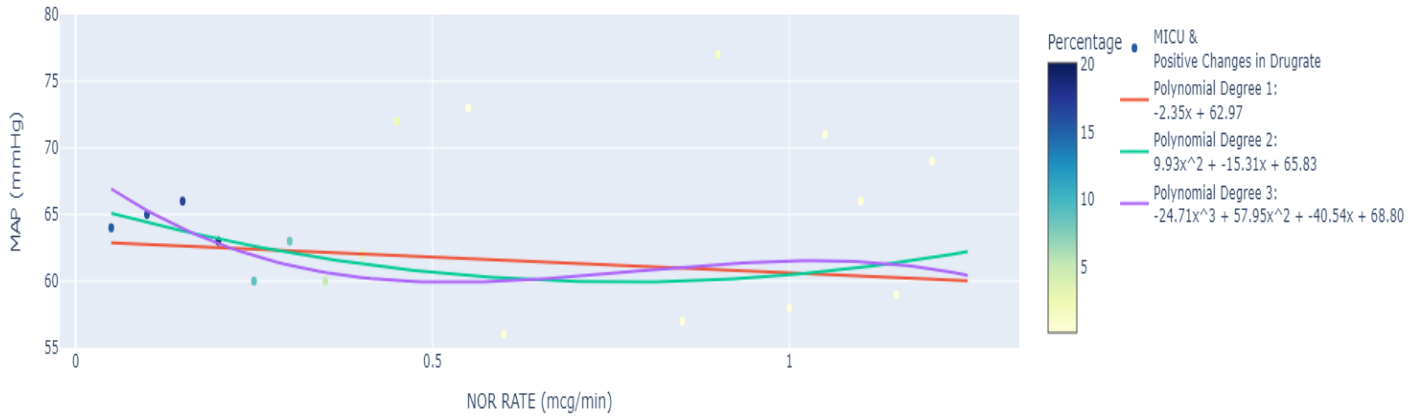

Figure 8: Polynomial fitting. Med-Surg-ICU & negative changes in NOR rate

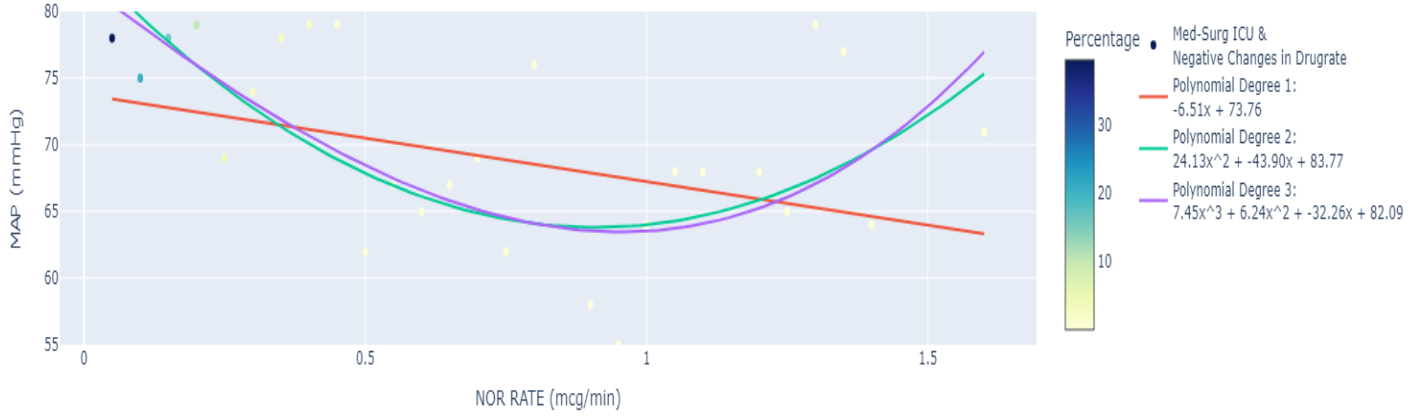

Figure 9: Polynomial fitting. MICU & positive changes in NOR rate

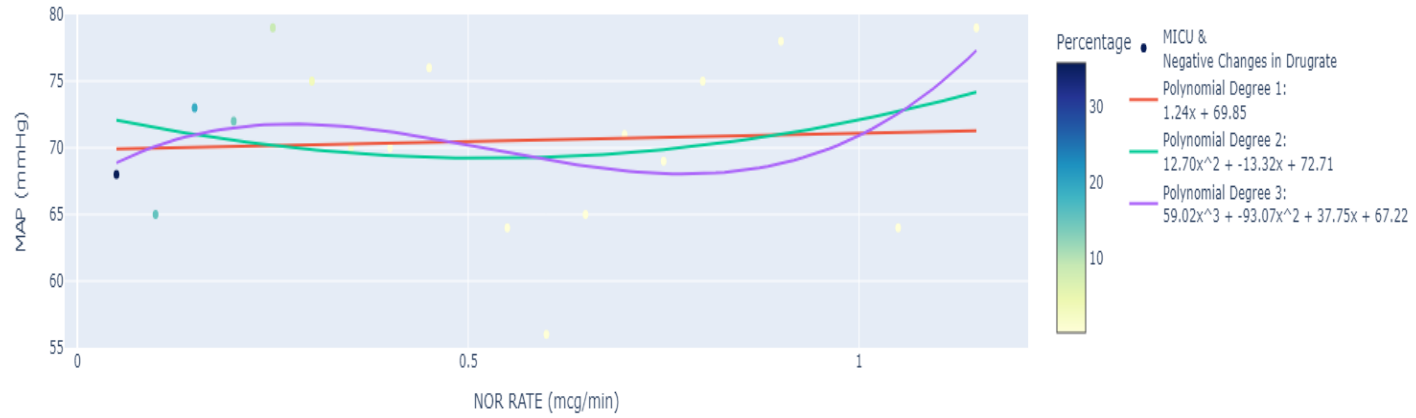

Figure 10: Polynomial fitting. MICU & negative changes in NOR rate

A comparison between the MICU and Med-Surg ICU reveals differences in the coefficients, especially in the intercept, which is significantly lower in Med-Surg ICU. Both units do not take advantage of the expressibility given in higher degrees of polynomials when restricted to negative changes, additionally, the trends seem fixed. In the positive change, the linear slope appears similar, though the behavior in higher degrees suggests that MICU has a different distribution with high NOR rates.

3.2 MAP-NOR Per Patient Stay

This section will concentrate on statistics extracted for each patient stay. Given that weight information is unavailable for most patients in eICU, the primary focus of this section was on statistics that are agnostic to the weight normalization process.

The Comparisons will be between differnt units:

1. MICU - Medical ICU. The unit where in the hospital where critically ill patients receive care for a variety of conditions.

2. SICU - Surgical ICU. Treats for critically ill patients recovering from surgeries. Most patients recuperating immediately after elective or emergency surgery receive care in this unit.

3. Med-Surg ICU - Medical-Surgical ICU. Mixed unit that admits both kinds of critticlly ill patients.

Additionally eICU and SICdb will be compared, that could be interpreted as comparison between the American and Austrian policies on vasopressor therapy in septic shock in clinical practice.

3.2.1 Patient Stay Wise High Level Statistics

The statistics that will be compared between the different units and different data sets are:

i. "First Time To Reach Target" - Time in minutes since admission with septic shock to the unit until reaching a MAP value between 60 and 70 mmHg.

ii. "Time At First NOR" - Time in minutes since admission with septic shock to the intensive care unit until getting the first NOR dosage.

iii. "MAP At First NOR" - MAP value while getting the first NOR dosage.

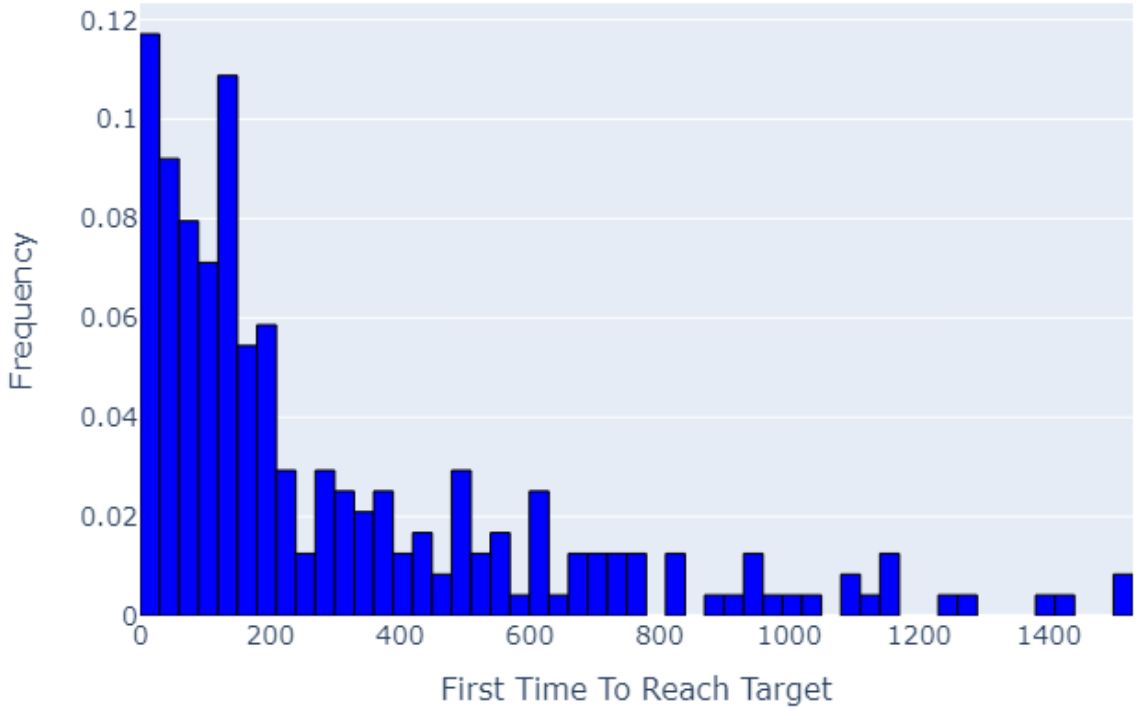

Figure 11: The distribution of the "First Time To Reach Target" statistic in each unit per patient stay In MICU

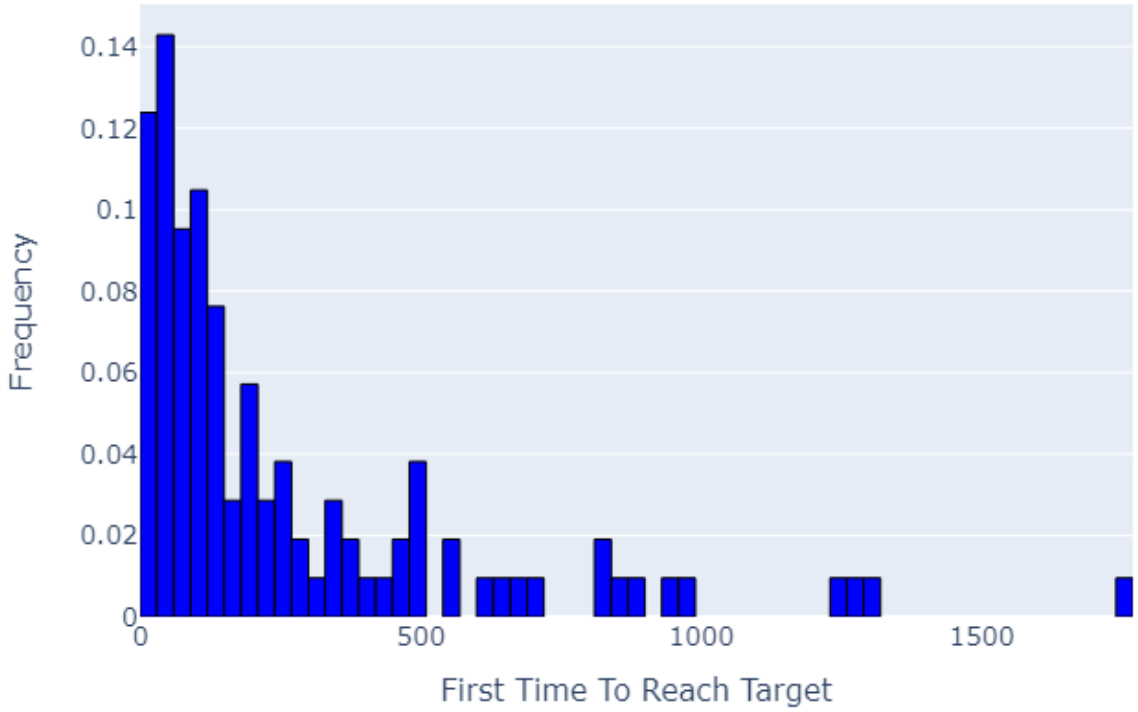

Figure 12: The distribution of the "First Time To Reach Target" statistic in each unit per patient stay In SICU

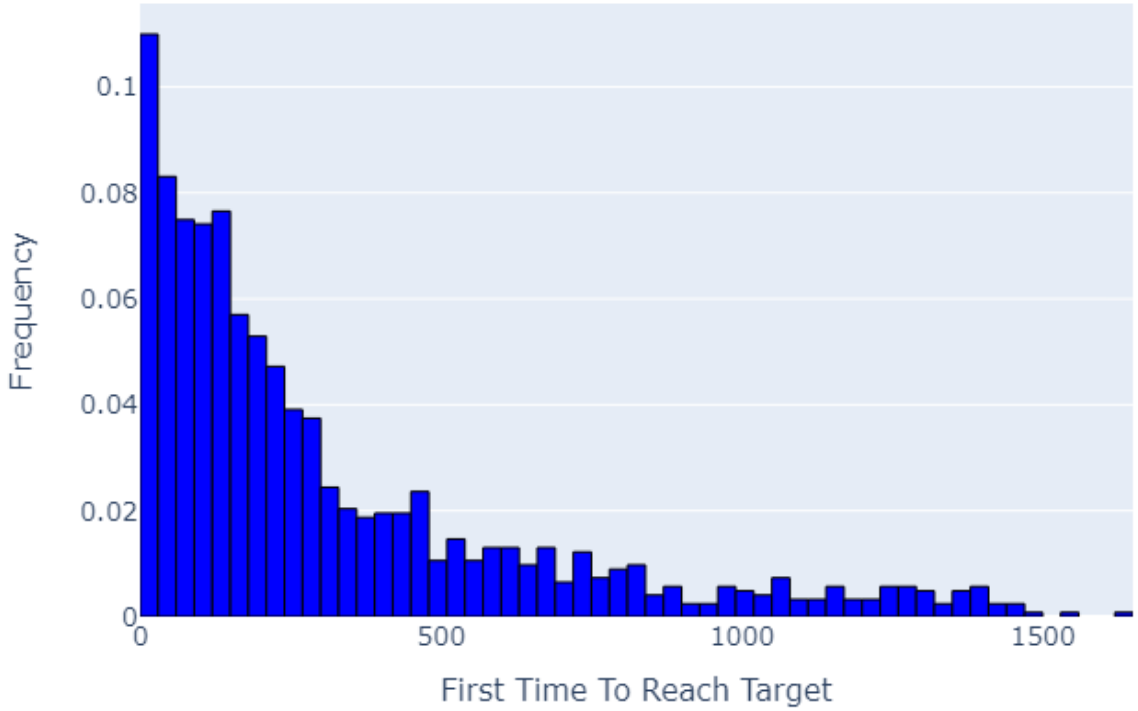

Figure 13: The distribution of the "First Time To Reach Target" statistic in each unit per patient stay In Med-Surg ICU

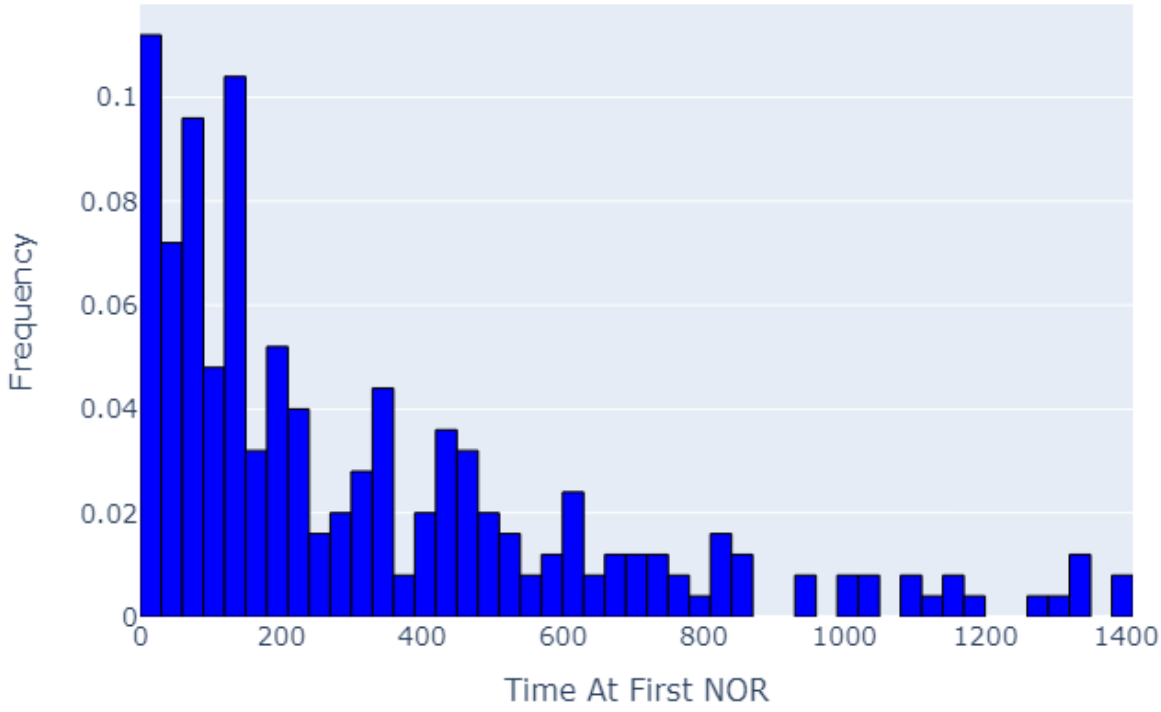

Figure 14: The distribution of the "Time At First NOR" statistic in each unit per patient stay In MICU

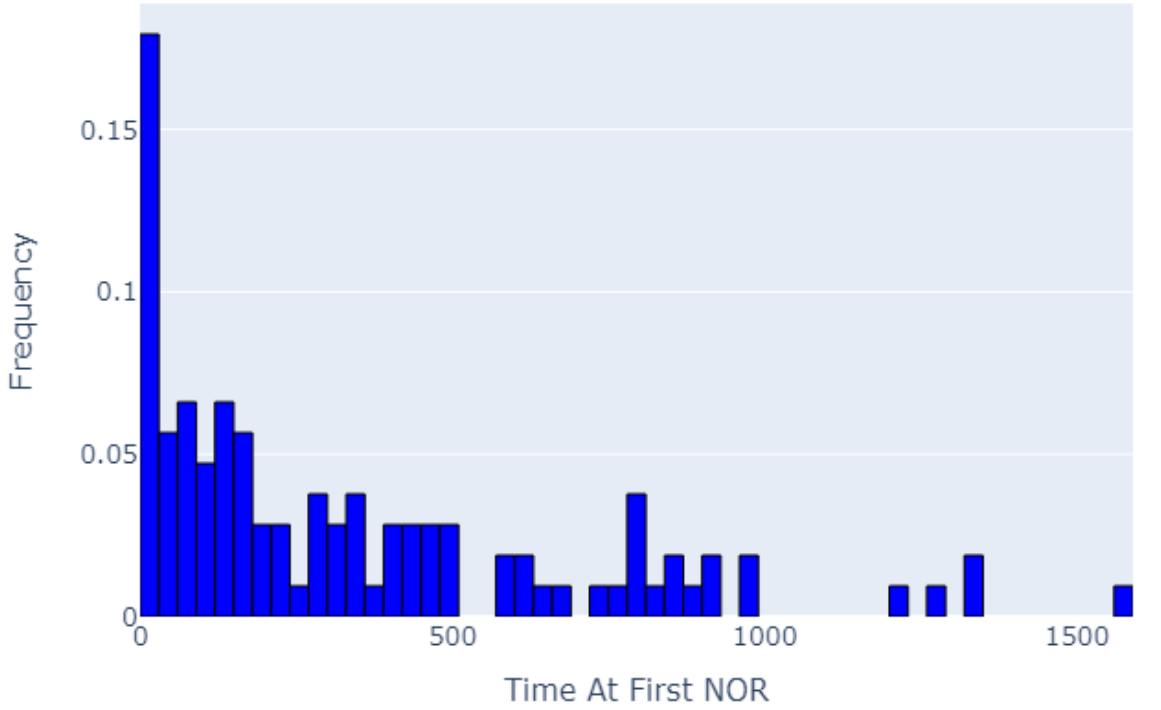

Figure 15: The distribution of the "Time At First NOR" statistic in each unit per patient stay In SICU

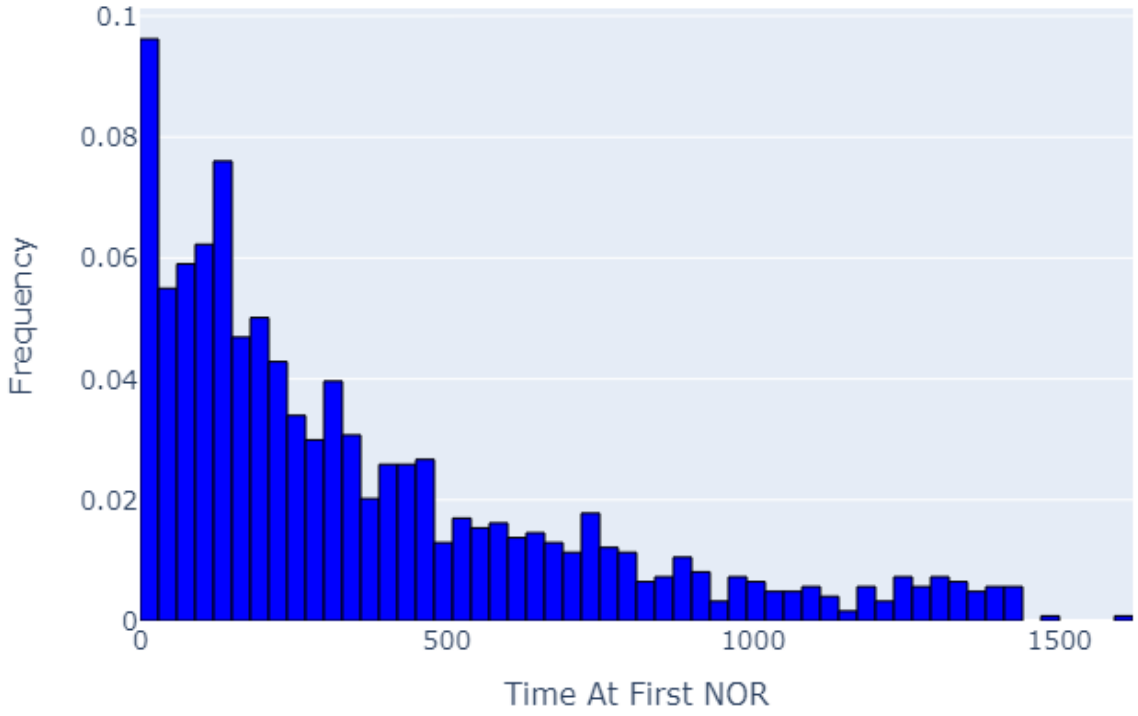

Figure 16: The distribution of the "Time At First NOR" statistic in each unit per patient stay In Med-Surg ICU

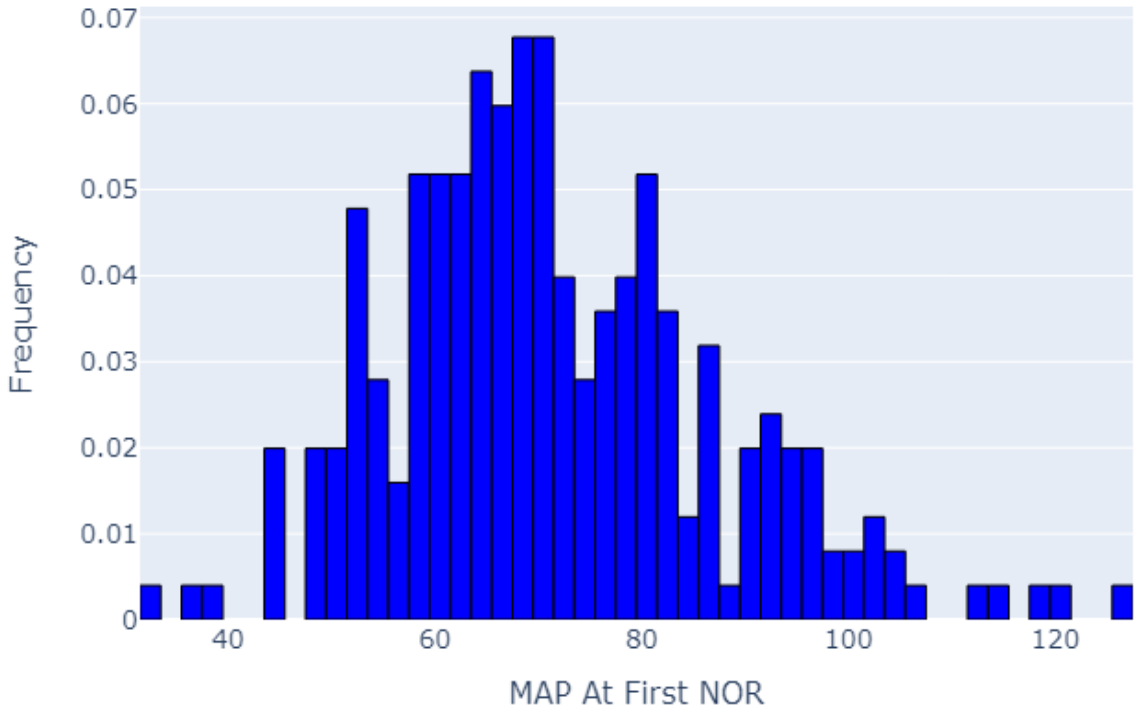

Figure 17: The distribution of the "MAP At First NOR" statistic in each unit per patient stay In MICU

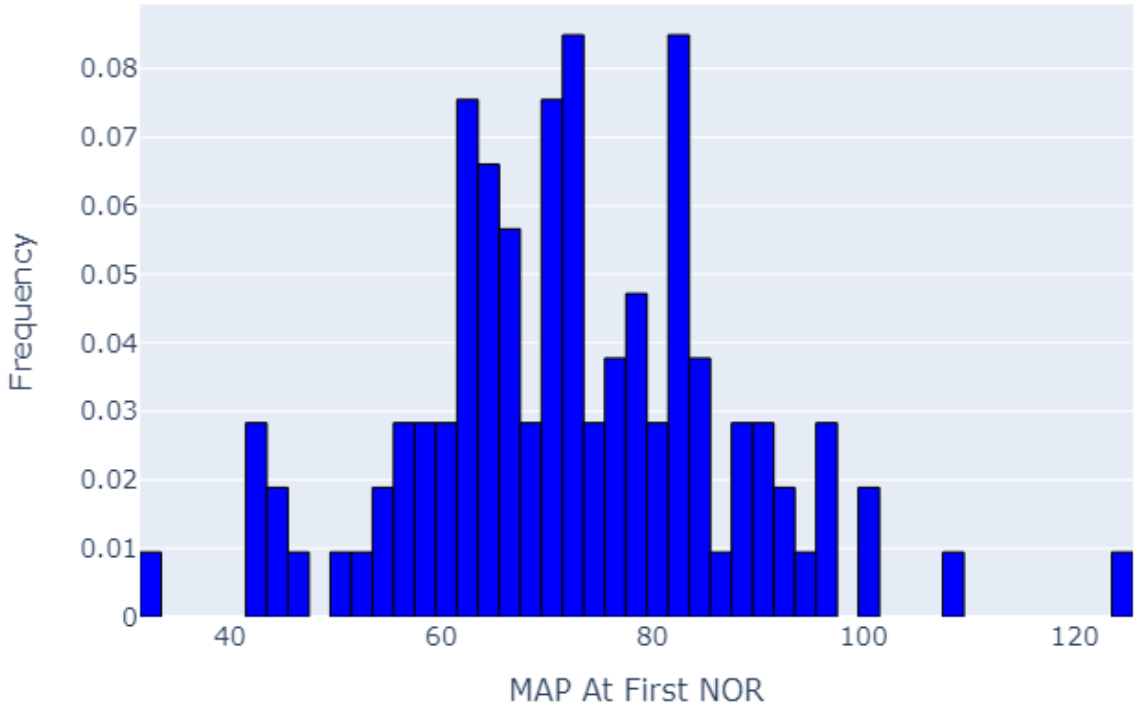

Figure 18: The distribution of the "MAP At First NOR" statistic in each unit per patient stay In SICU

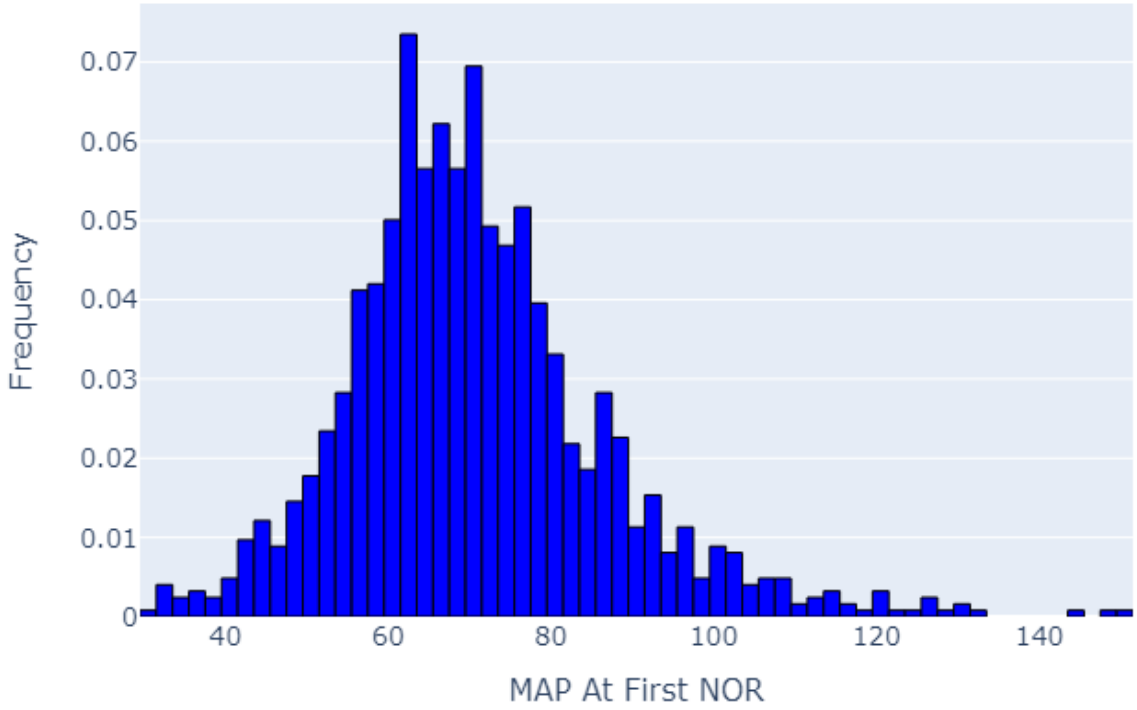

Figure 19: The distribution of the "MAP At First NOR" statistic in each unit per patient stay In Med-Surg ICU

Using non-parametric statistical tests Two-sample Kolmogorov-Smirnov and Mann Whitney U it can be clarified how different the distribution are from each other.

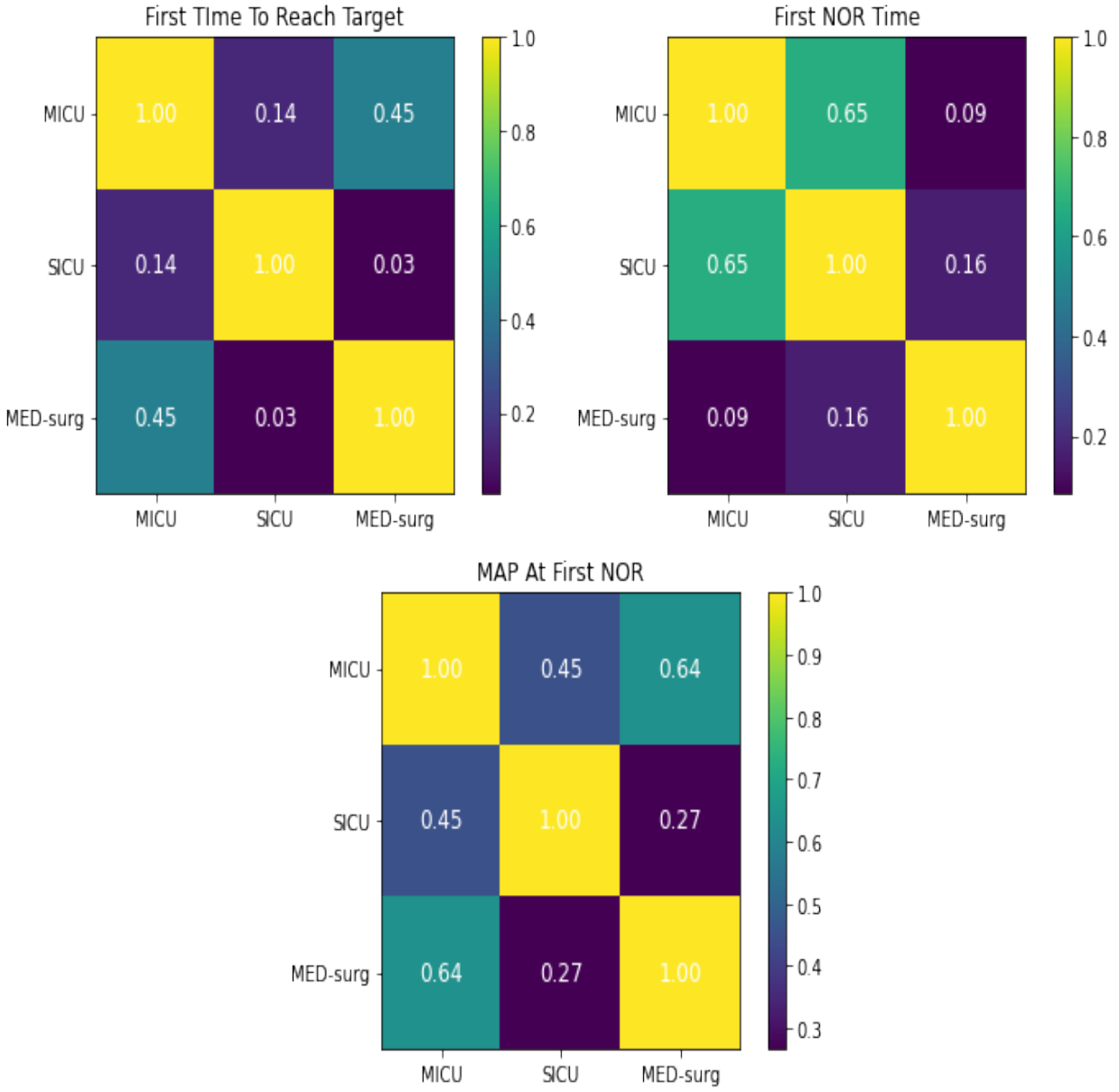

Figure 20: p-values of Two-sample Kolmogorov-Smirnov test

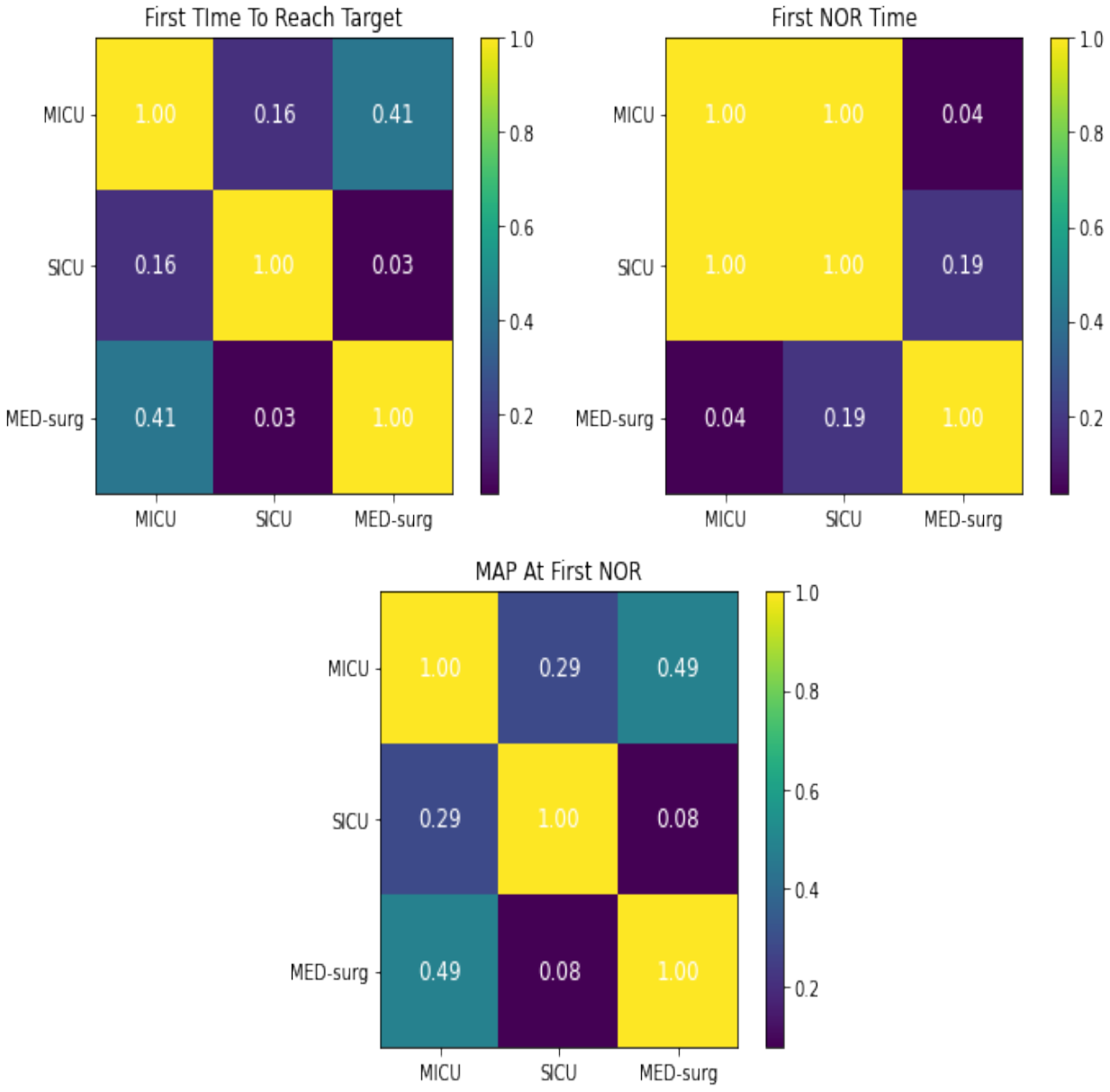

Figure 21: p-values of Mann Whitney U test

Applying these tests, for the most part, indicates that there is no significant difference between the distributions.

Using permutation test it can be assessed whether there is a significant difference in means between two units. In the following graphs the red line indicates the real mean difference between the units.

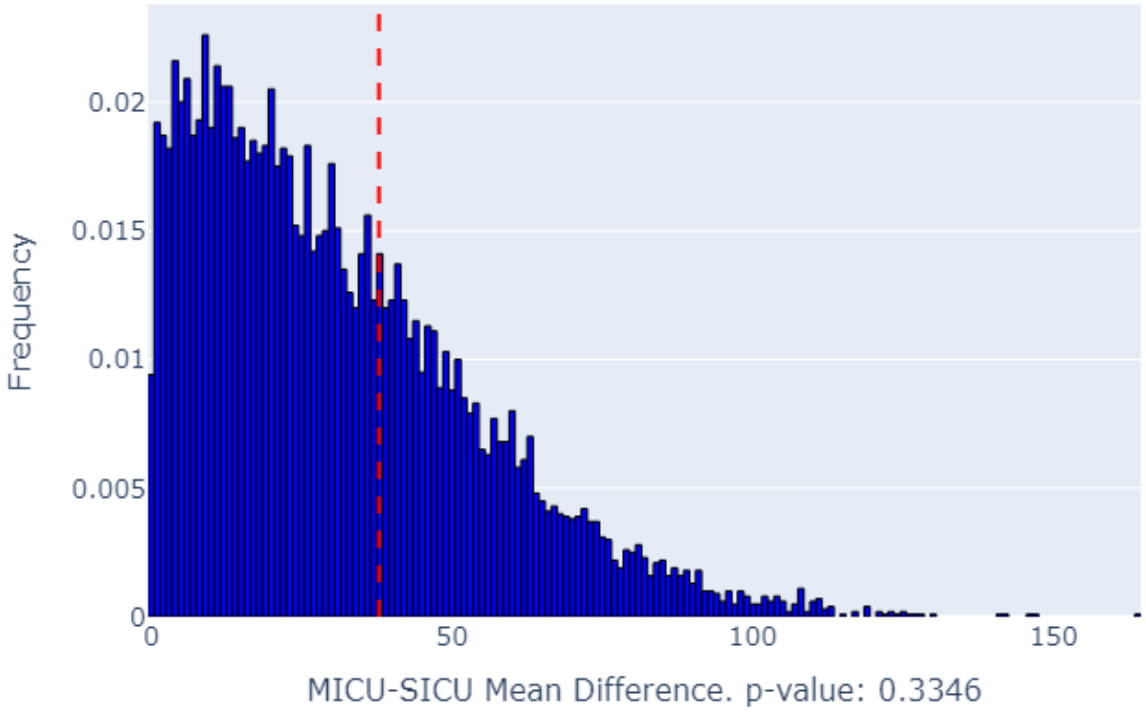

Figure 22: "First Time To Reach Target" - permutation test of Mean difference MICU-SICU

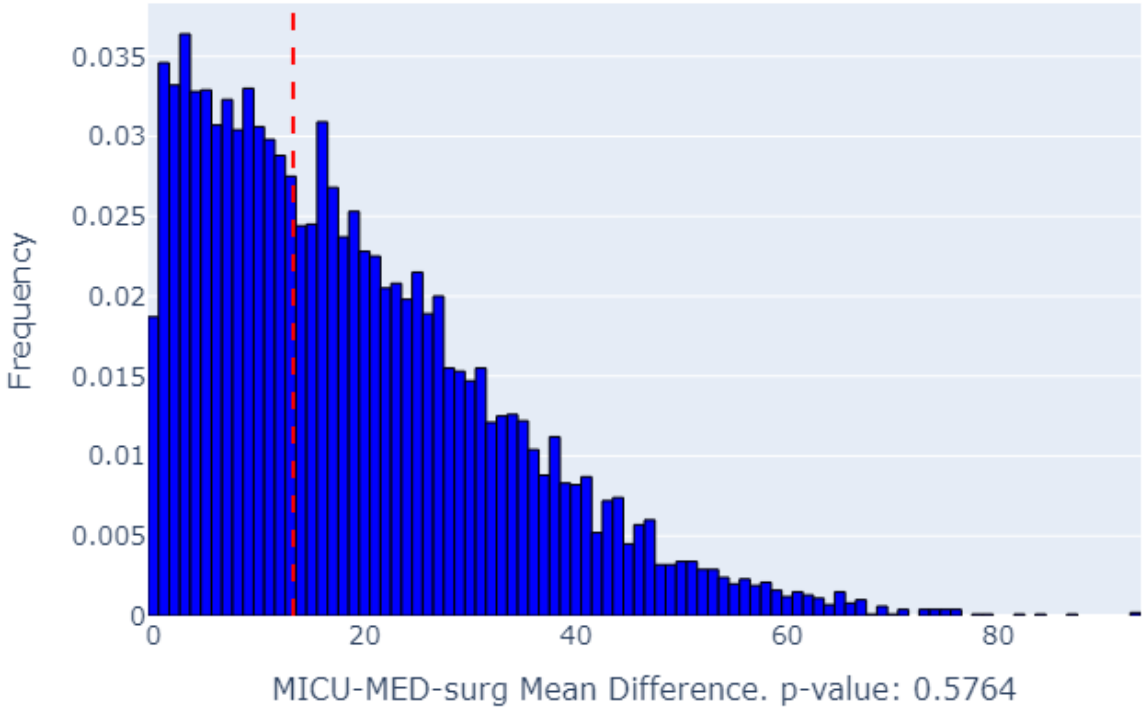

Figure 23: "First Time To Reach Target" - permutation test of Mean difference MICU-MED Surg ICU

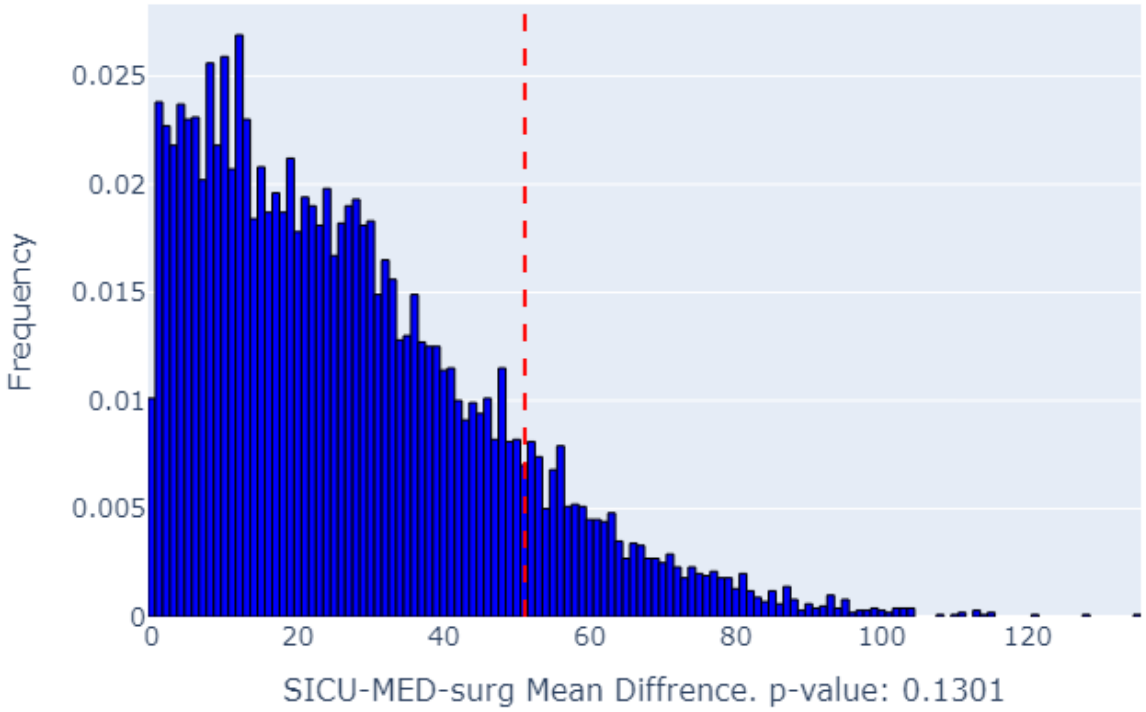

Figure 24: "First Time To Reach Target" - permutation test of Mean difference SICU-MED Surg ICU

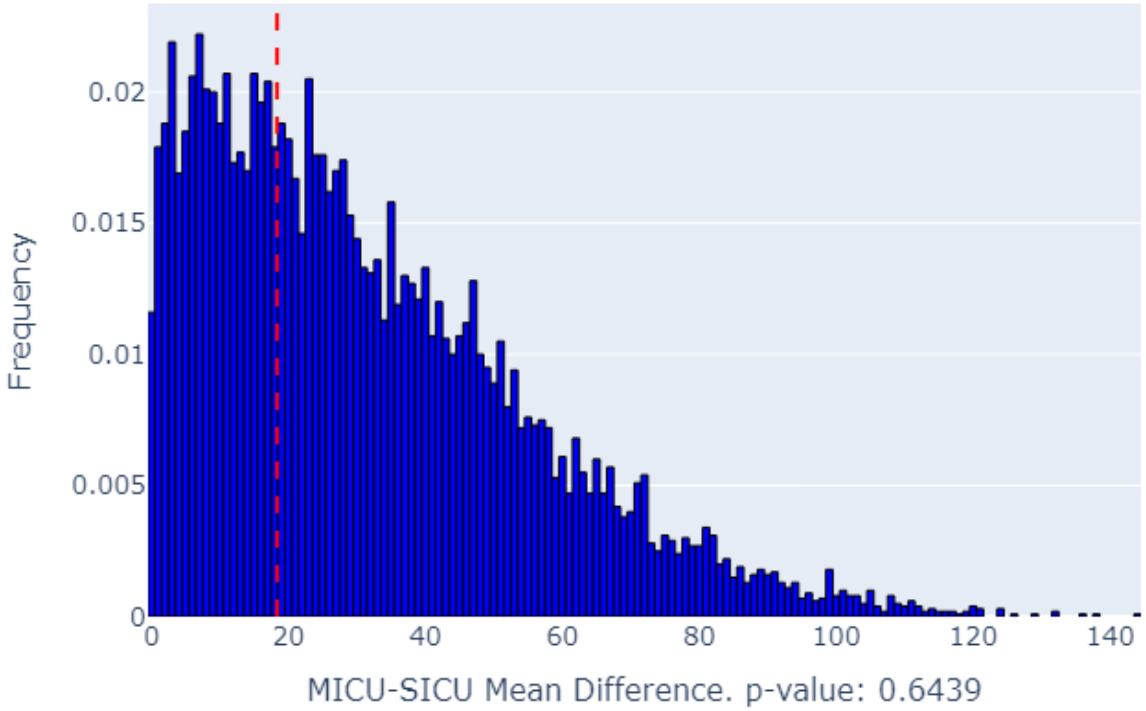

Figure 25: "Time At First NOR" - permutation test of Mean difference MICU-SICU

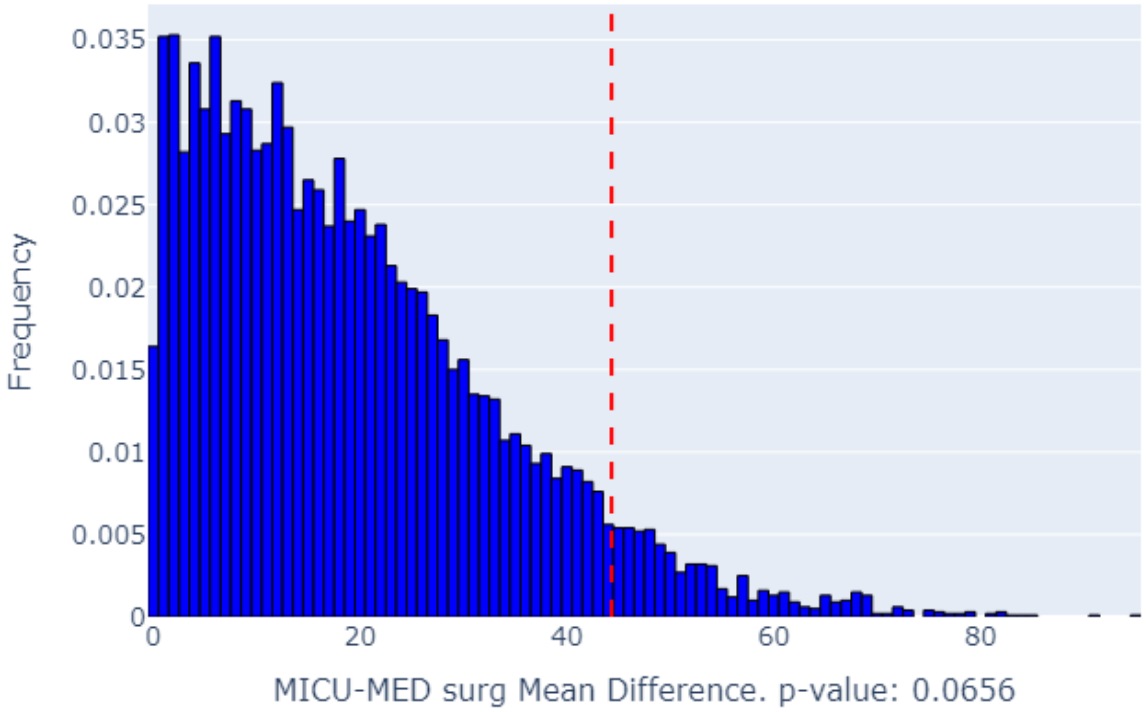

Figure 26: "Time At First NOR" - permutation test of Mean difference MICU-MED Surg ICU

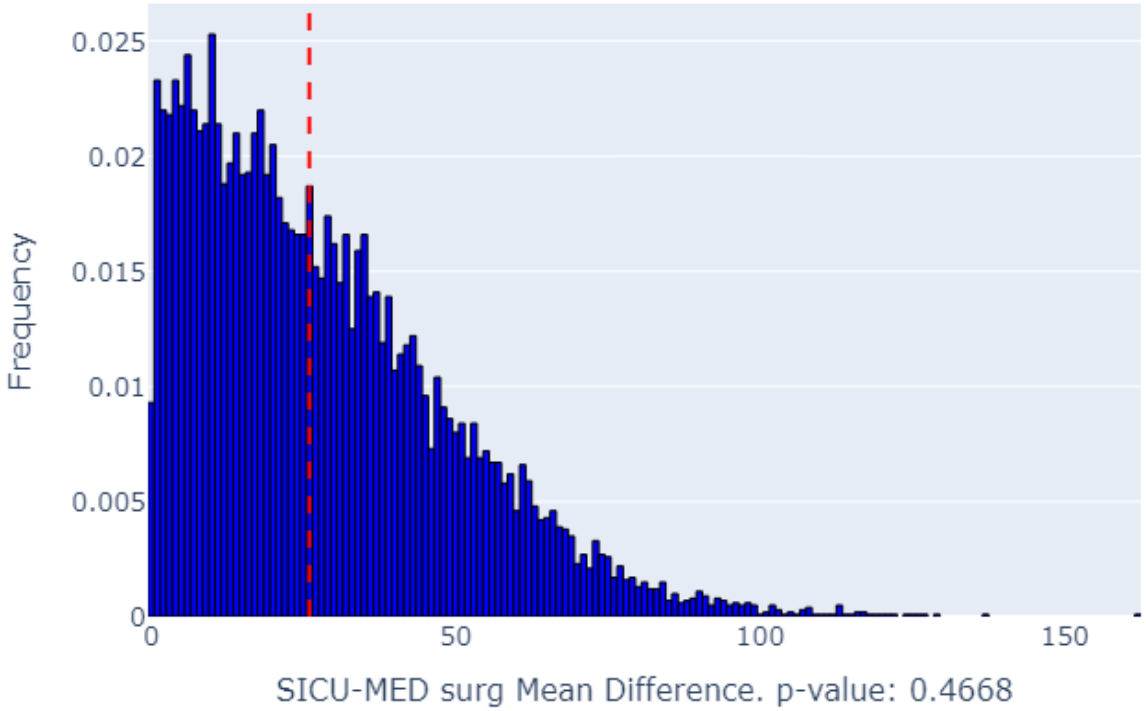

Figure 27: "Time At First NOR" - permutation test of Mean difference SICU-MED Surg ICU

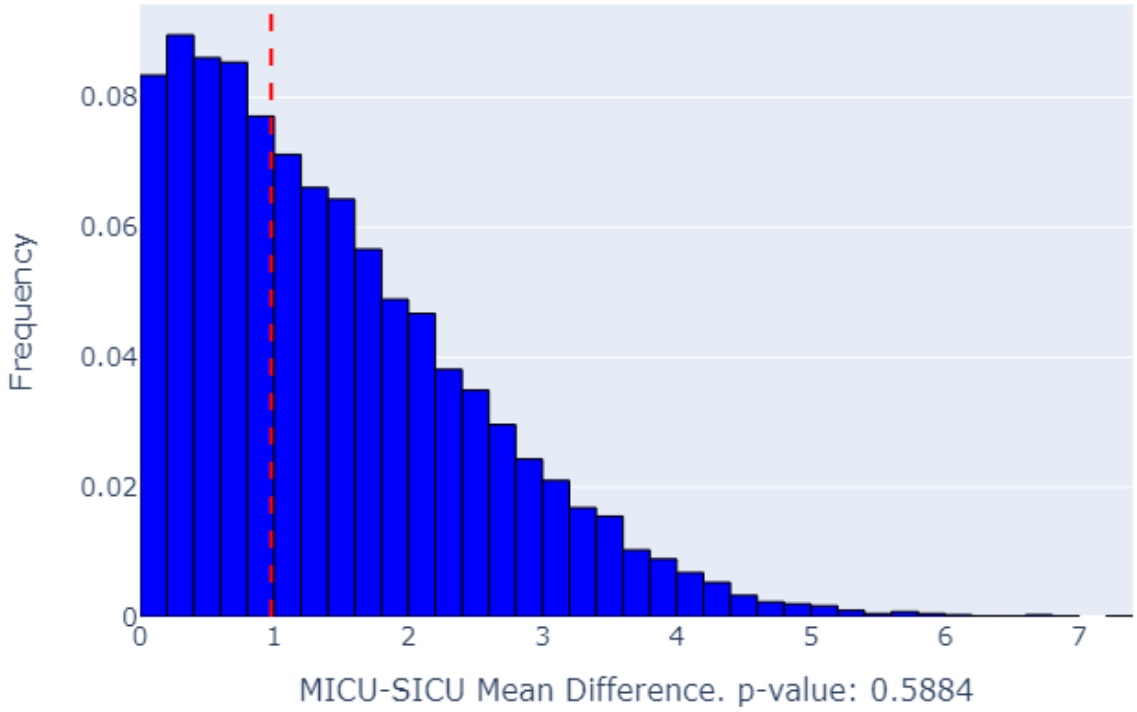

Figure 28: "MAP At First NOR" - permutation test of Mean difference MICU-SICU

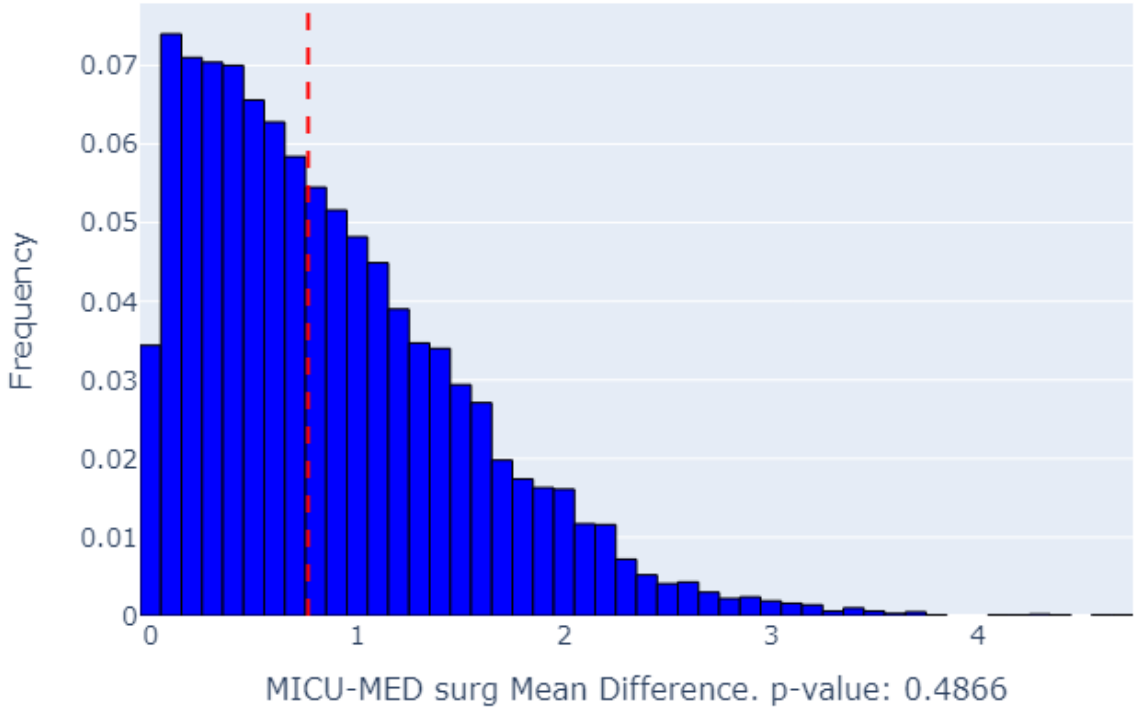

Figure 29: "MAP At First NOR" - permutation test of Mean difference MICU-MED Surg ICU

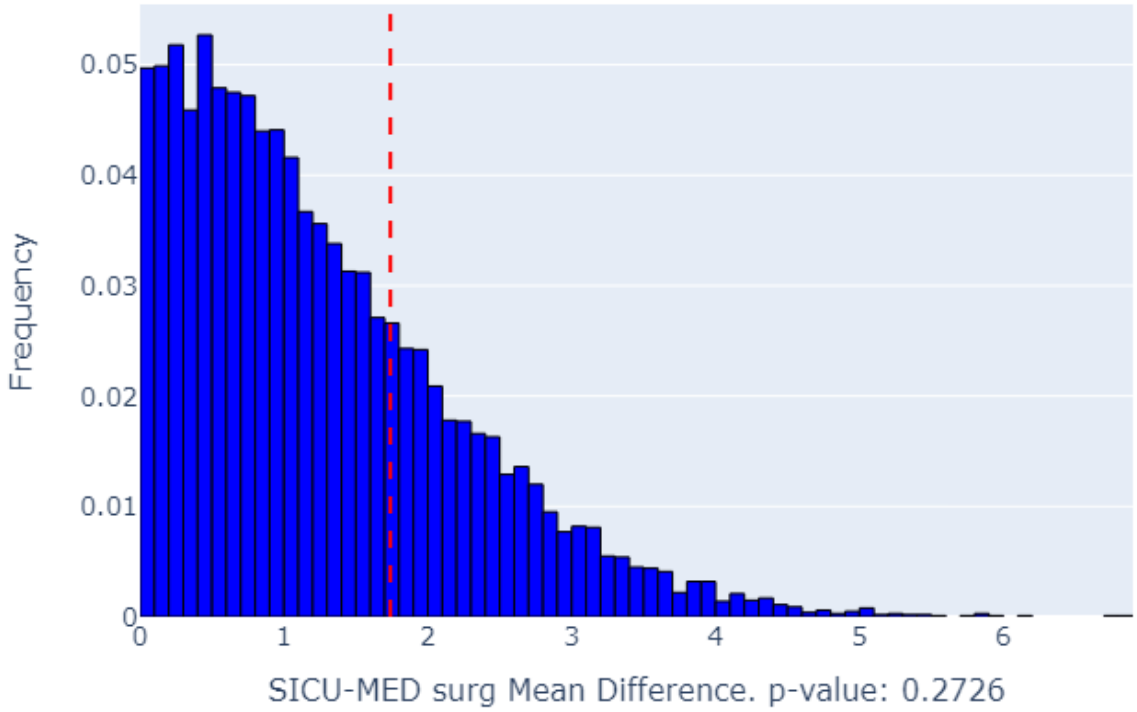

Figure 30: "MAP At First NOR" - permutation test of Mean difference SICU-MED Surg ICU

The statistics were also taken for the SICdb data set. The following graphs are the distributions of the statistics in SICdb data set per patient stay.

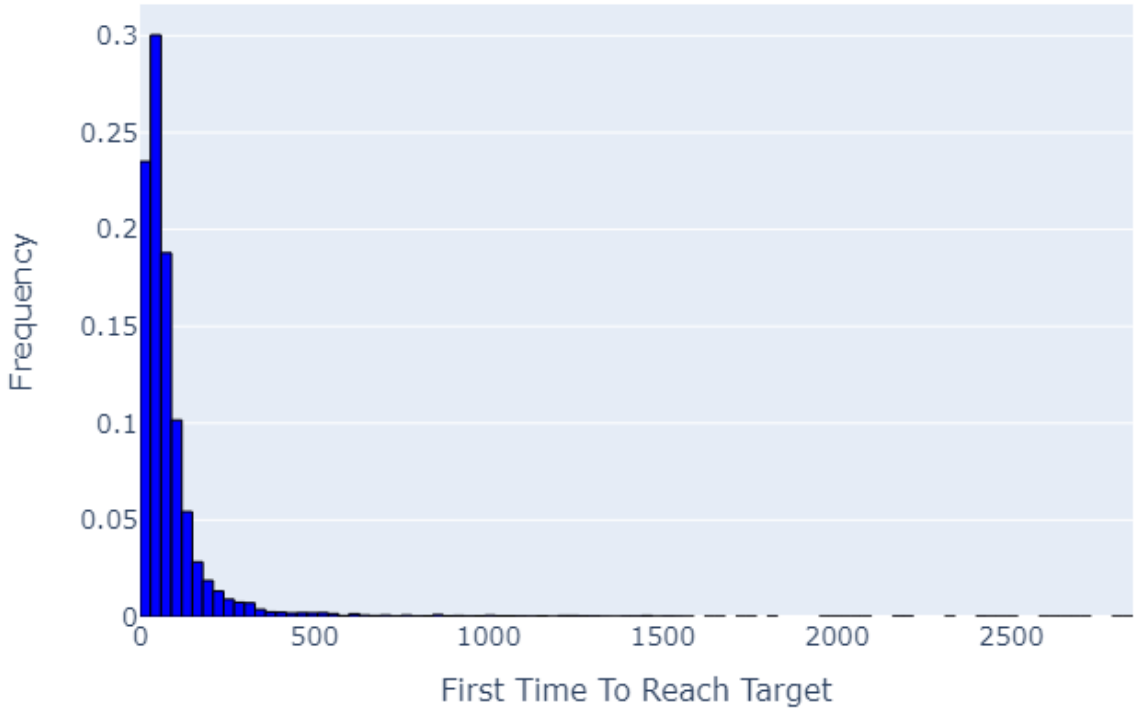

Figure 31: "First Time To Reach Target" distributions in SICdb data set

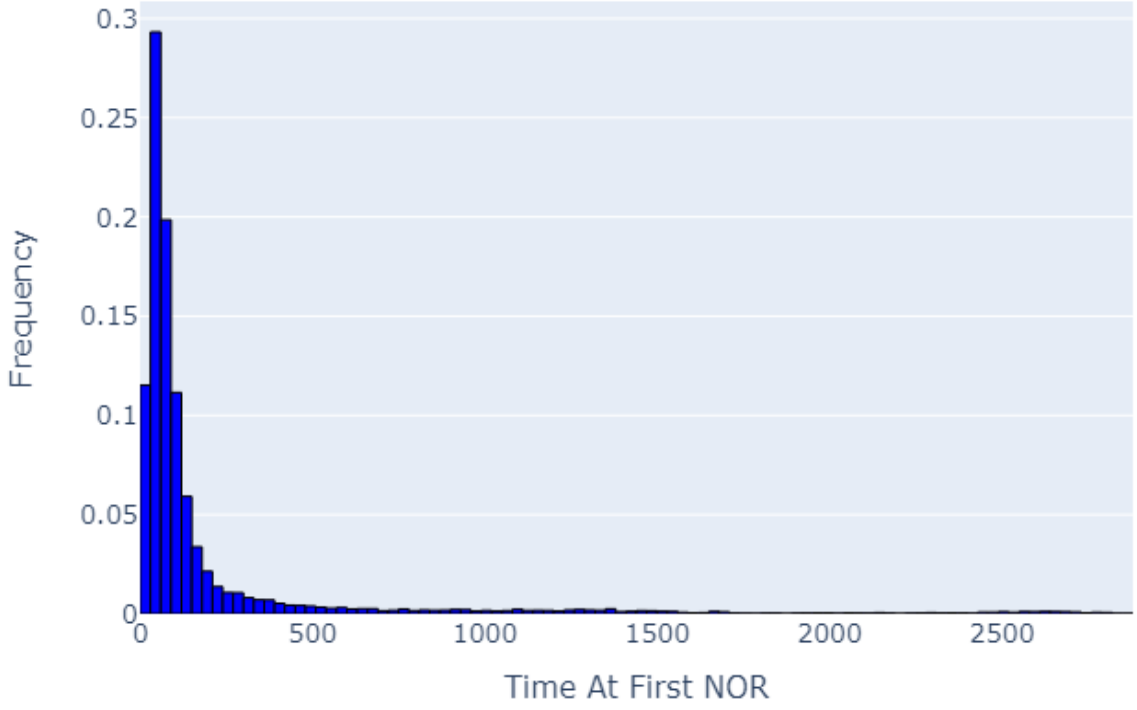

Figure 32: "Time At First NOR" distributions in SICdb data set

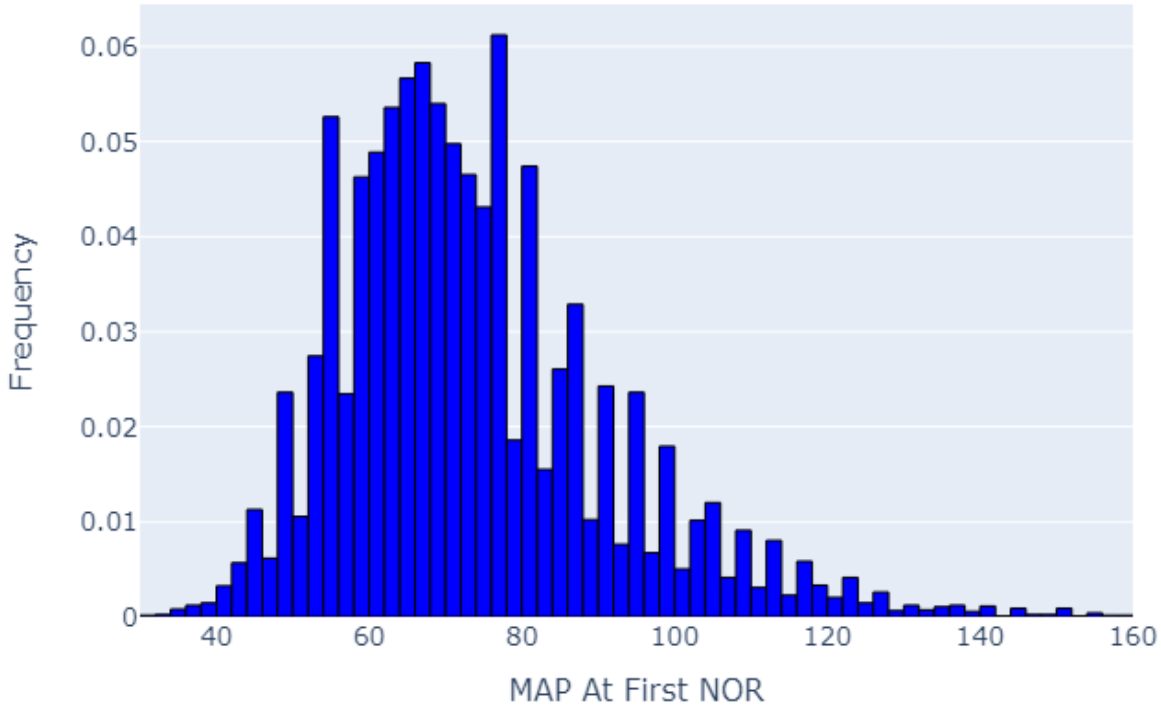

Figure 33: "MAP At First NOR" distributions in SICdb data set

Below are permutation tests of mean difference in the SICdb data set:

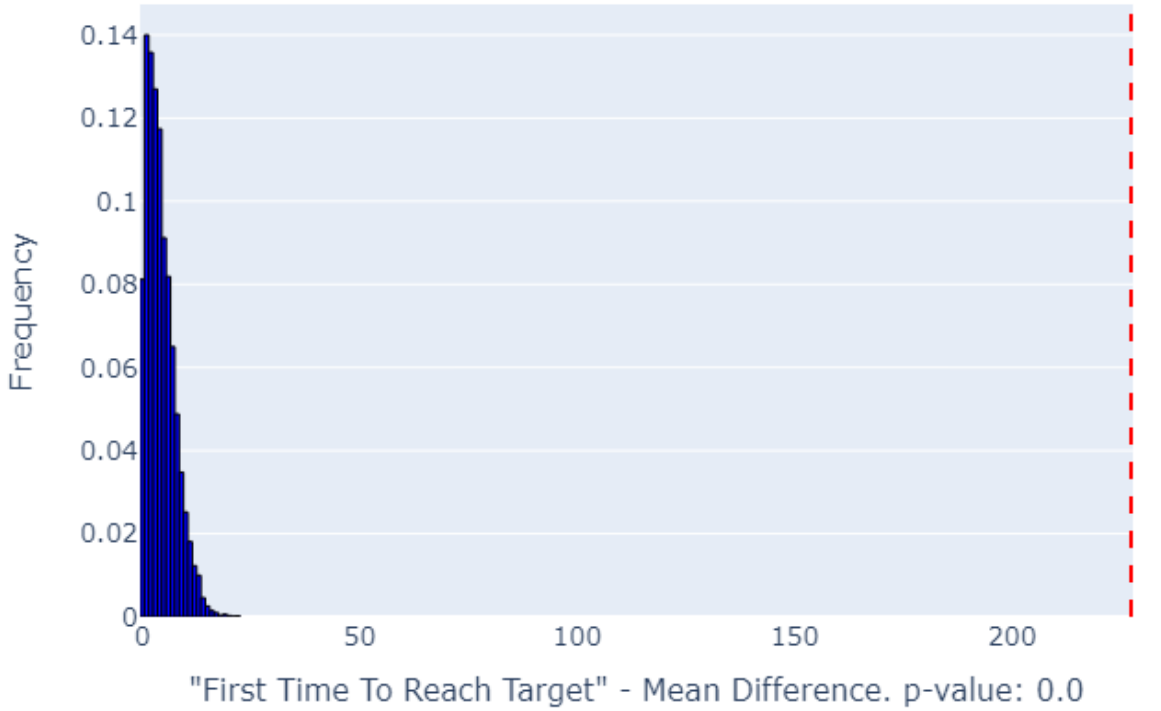

Figure 34: "First Time To Reach Target" - permutation test of Mean difference SICdb-eICU

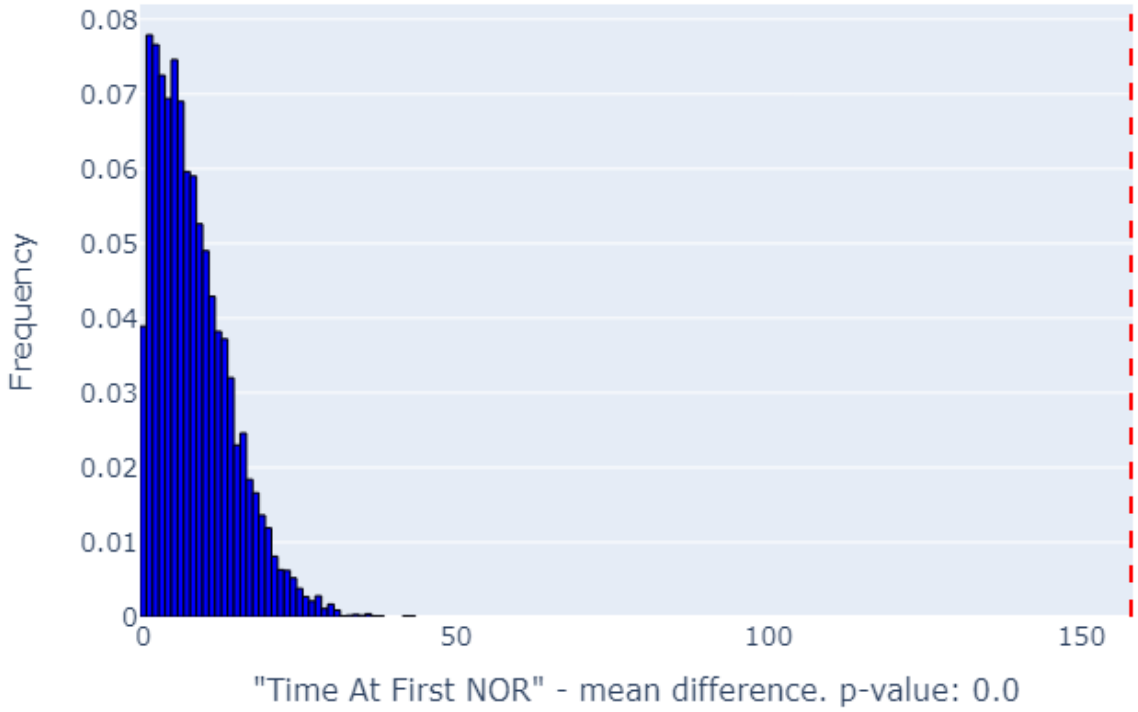

Figure 35: "Time At First NOR" - permutation test of Mean difference SICdb-eICU

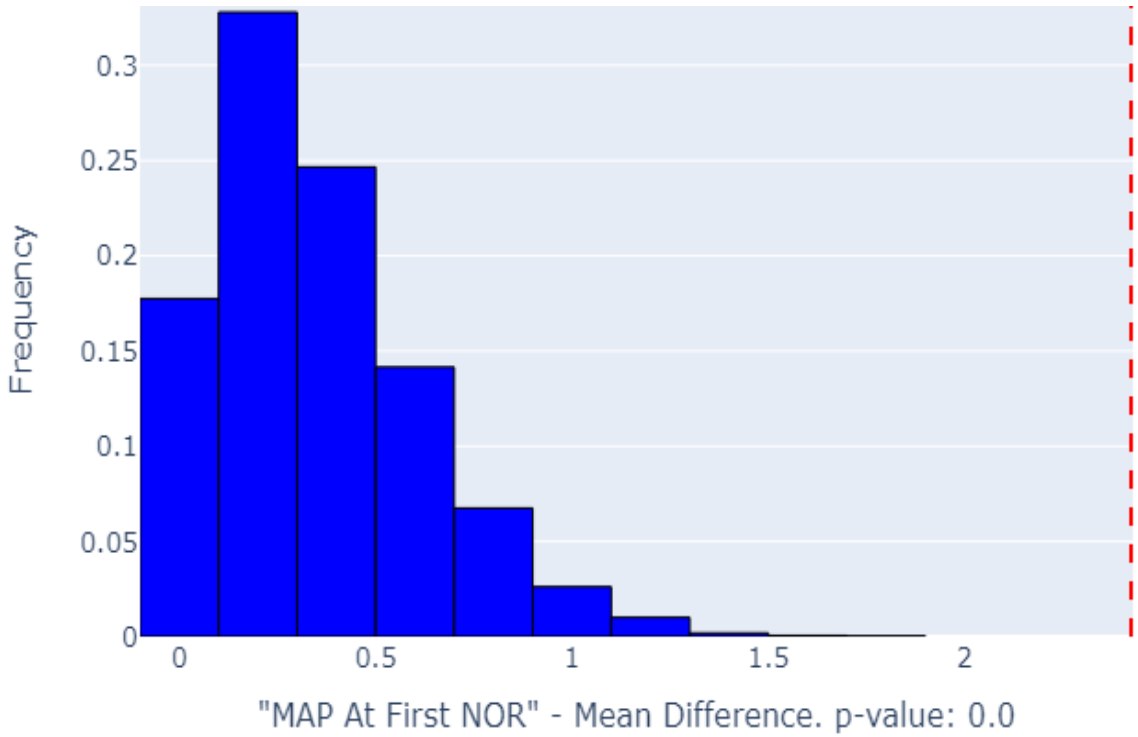

Figure 36: "MAP At First NOR" - permutation test of Mean difference SICdb-eICU

These permutation tests indicate a significant difference between the two data set dis-

tributions. Applying the non-parametric tests Two-sample Kolmogorov-Smirnov and Mann-Whitney U yields the same results.

3.2.2 Patient Stay Wise Pearson Correlation Coefficient as a Statistic

For each patient's stay, the correlation coefficient, denoted as $\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Sd}(X)\cdot\mathrm{Sd}(Y)}$, was estimated between NOR and the rolling statistics of mean, median, max, and min for windows ranging from 5 to 50 minutes before the time of the entry of the NOR value.

This statistic is a nice workaround for the weight normalization problem, with the assumption that the NOR rate is a linear product $Y = w \cdot T(\mathrm{MAP})$, where T the function in MAP determined by the treatment policy, which is set to normalize, and w is the weight, using basic probability we have

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Sd}(X)\cdot\mathrm{Sd}(Y)} = \frac{\mathrm{Cov}(X, w\cdot T)}{\mathrm{Sd}(X)\cdot\mathrm{Sd}(w\cdot T)} = \frac{w\cdot\mathrm{Cov}(X,T)}{w\cdot\mathrm{Sd}(X)\cdot\mathrm{Sd}(T)} = \frac{\mathrm{Cov}(X,T)}{\mathrm{Sd}(X)\cdot\mathrm{Sd}(T)} = \rho_{X,T}$$

This leads us to the conclusion that the MAP-NOR correlation is the same for actual NOR, normalized by the patient's weight, and for the mcg/min value deduced by medical personnel, which is our point of interest.

The formulation of the MAP-NOR relationship as a state-action process leads us to believe that the correlation should be negative. This is because increasing the NOR rate is expected to be the appropriate reaction for a lowered MAP, implying an inverse relationship between the two variables. Another phenomenon that we would like to observe, in order to further distinguish the decision-making process among different units, is a different correlation distribution between the units. For instance, if one of the units exhibits a stronger correlation between the 2-rolling statistics and the others don't, this could imply different practices regarding the time window considered in infusing NOR, the time it takes to execute a decision or documentation delays.

3.2.2.1 Comparison Between units These histograms demonstrates 2 surprising qualities. First, the mean appears to be 0, which is not compatible with our hypothesis that the correlations should be negative. Second, the distributions seem to resemble in different degrees the normal distribution.

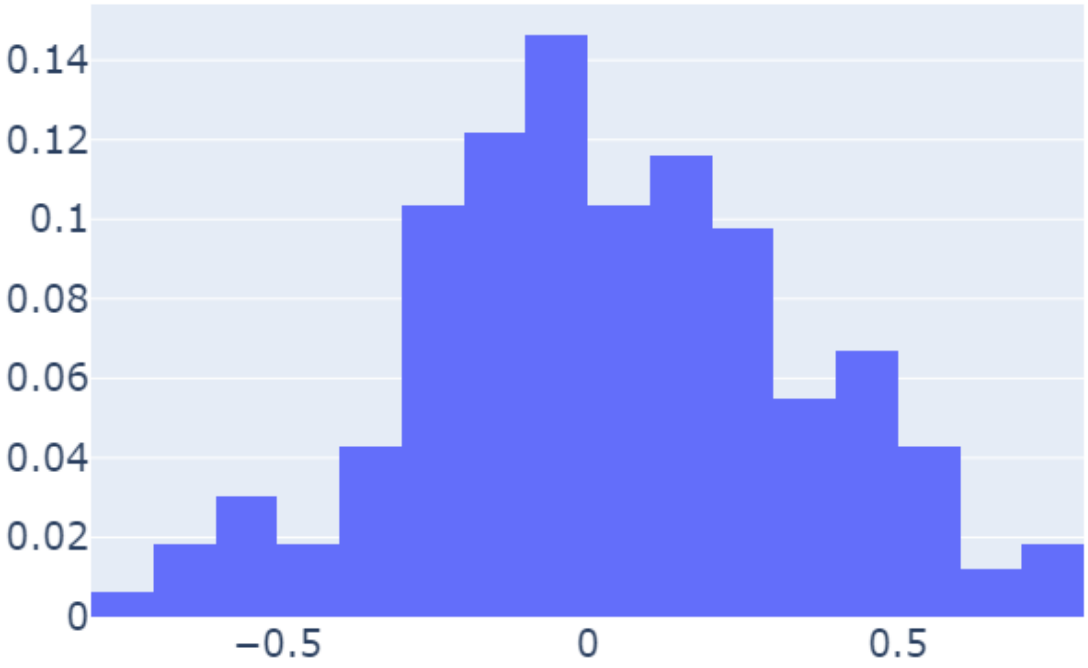

Figure 37: The distribution of the NOR-MAP correlation per patient stay - MICU

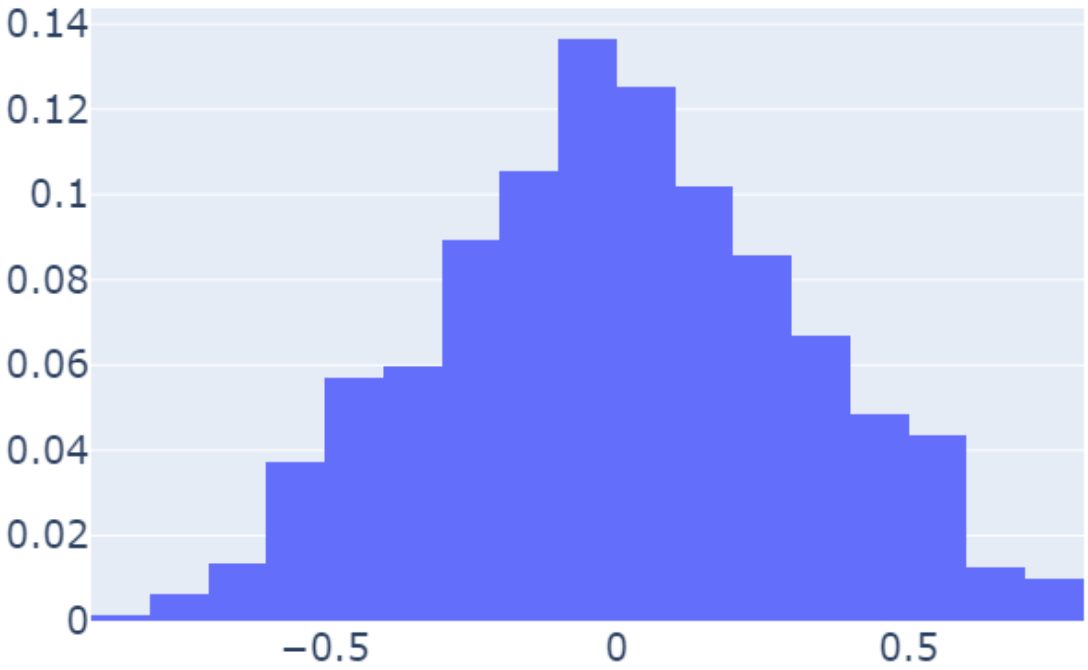

Figure 38: The distribution of the NOR-MAP correlation per patient stay - Med-Surg ICU

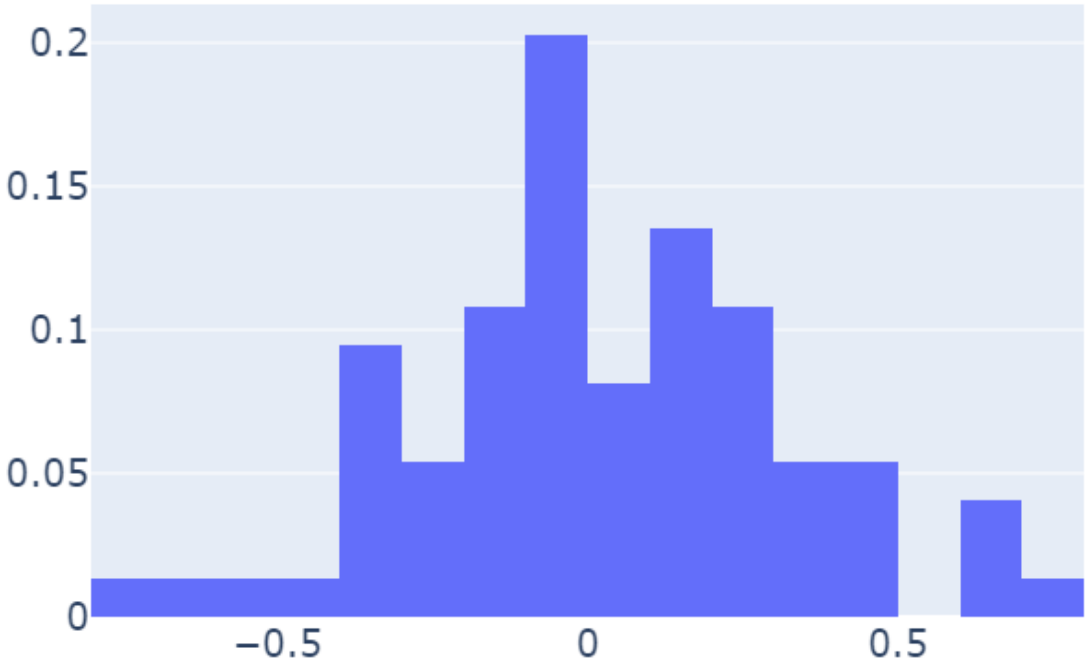

Figure 39: The distribution of the NOR-MAP correlation per patient stay - SICU

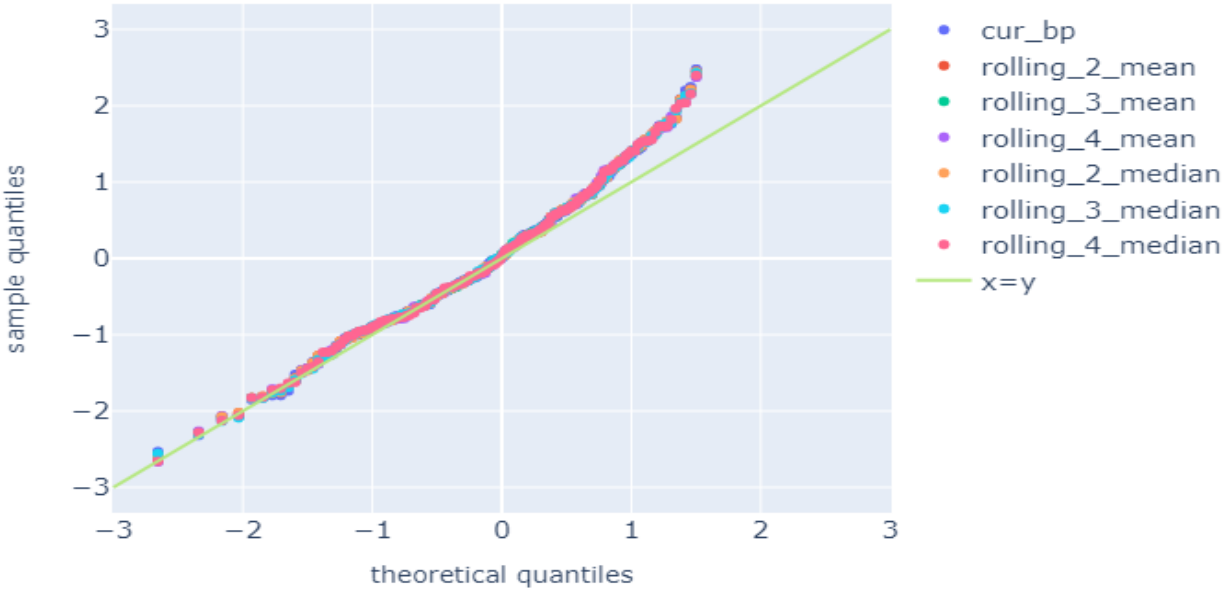

Figure 40: quantile-quantile plots for correlation distributions of the NOR-MAP statistics per MICU patient stay in comparison to the standard normal distribution

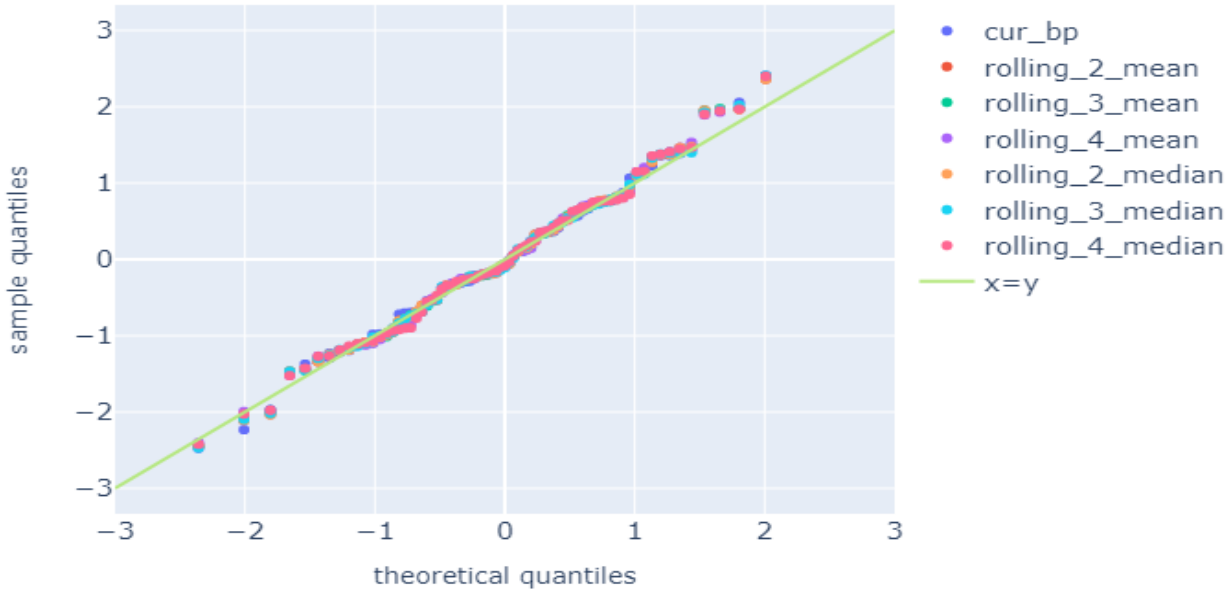

Figure 41: quantile-quantile plots for correlation distributions of the NOR-MAP statistics per SICU patient stay in comparison to the standard normal distribution

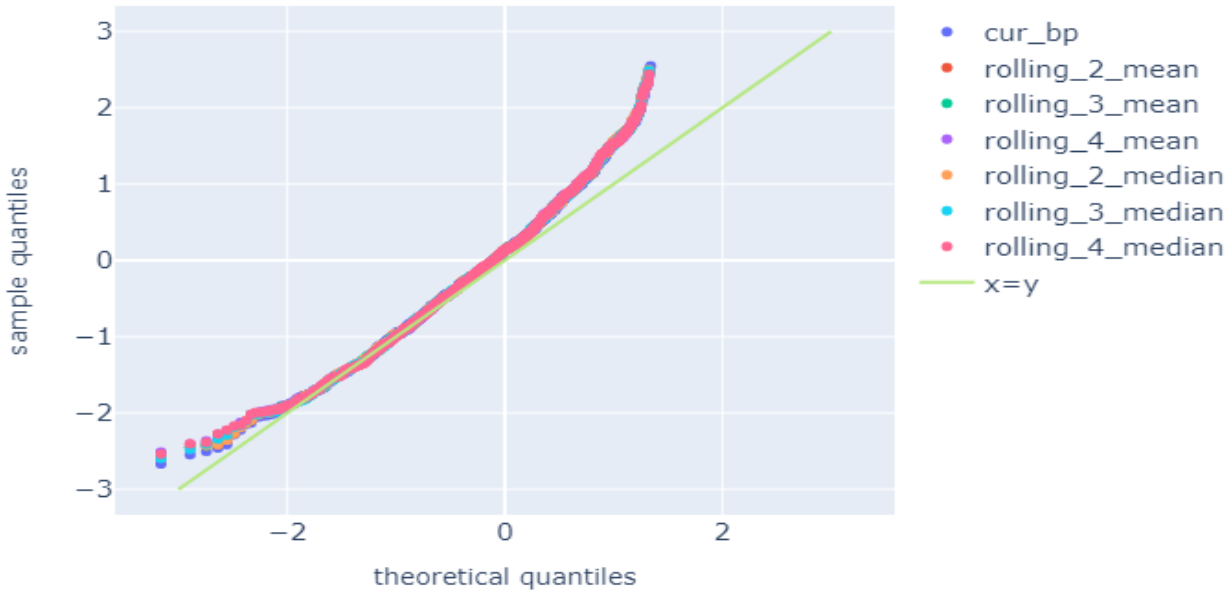

Figure 42: quantile-quantile plots for correlation distributions of the NOR-MAP statistics per Med-Surg ICU patient stay in comparison to the standard normal distribution

The application of the Shapiro-Wilk test for normality yielded results indicating that MICU and SICU are more normally distributed, whereas Med-Surg ICU is not. This plot illustrates the trend that we hope to find statistically significant: a difference between low and high values of the rolling parameter, which can be interpreted as the MAP range considered when making decisions about NOR infusion.

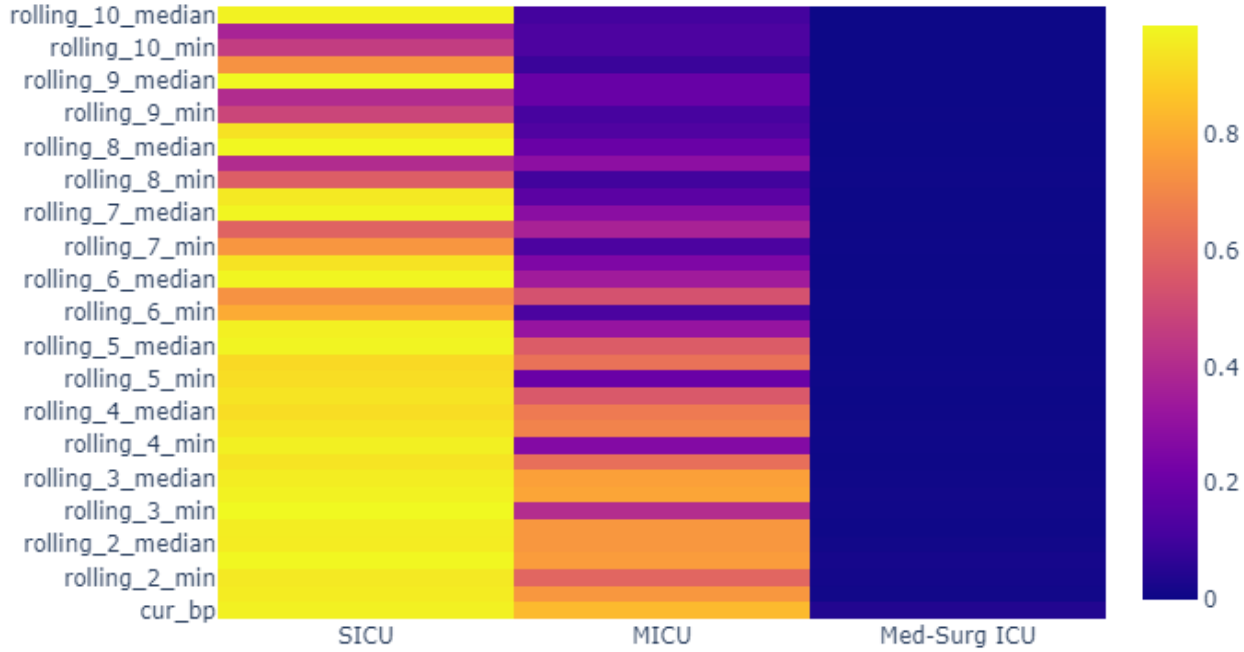

Figure 43: p-values of Shapiro-Wilk test applied to every statistic and every unit type

3.2.2.2 Change in Distribution for different values of k For each pair of statistics, we conducted tests to assess the potential differences between their distributions using both parametric and non-parametric tests. In the next heat maps, the color in each cell shows the p-value of the a test on 2 distribution given by 2 statistics.

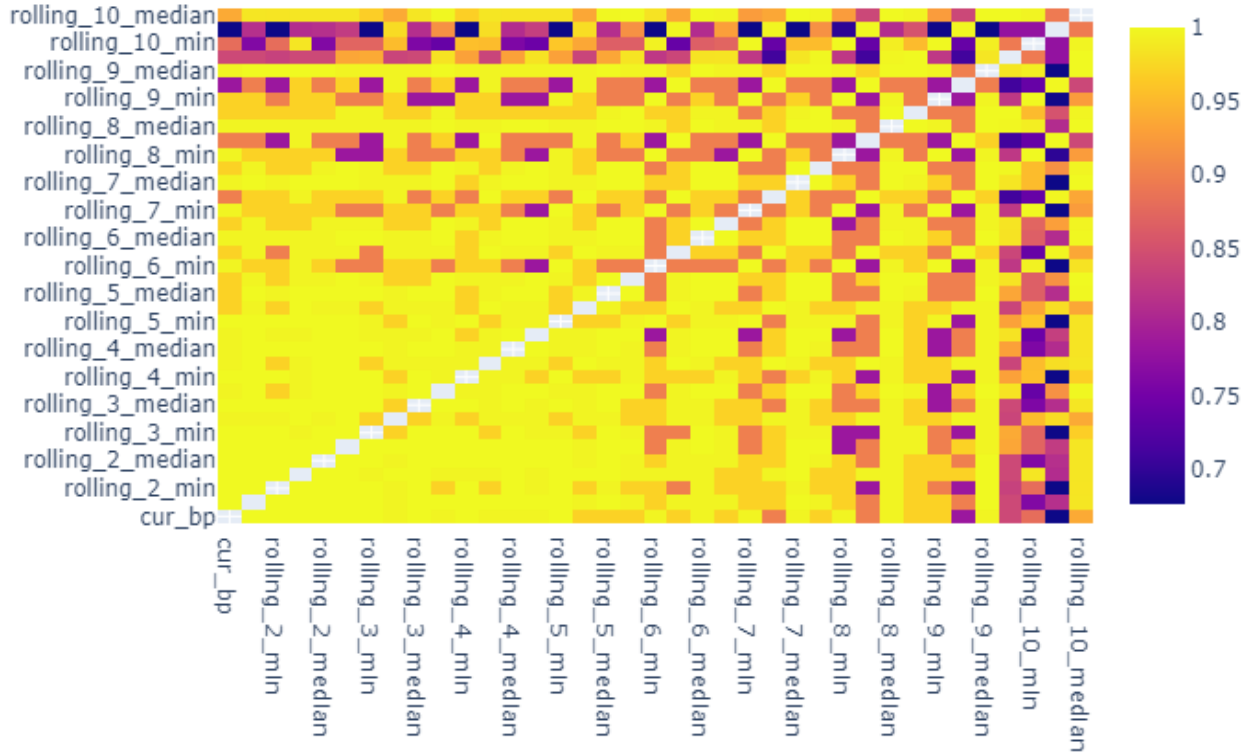

Figure 44: Kolmogorov Smirnoff test p-values for MICU

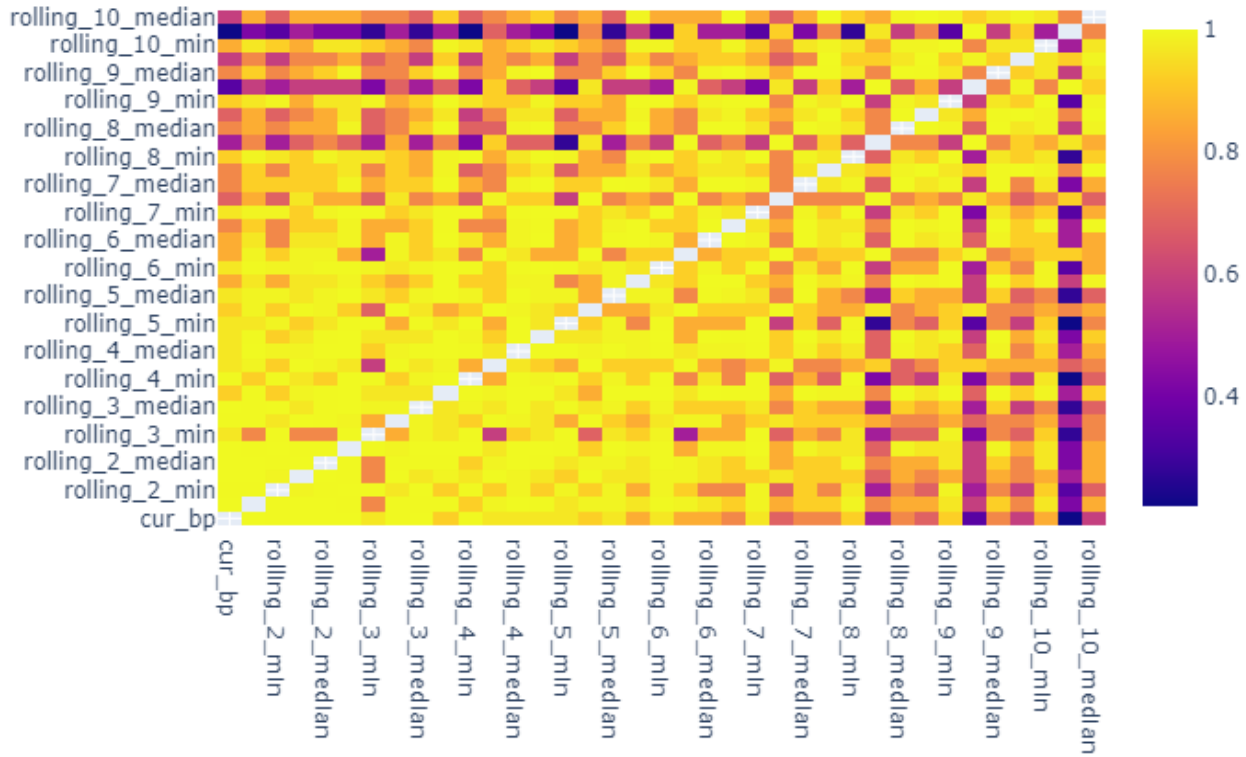

Figure 45: Kolmogorov Smirnoff test p-values for SICU

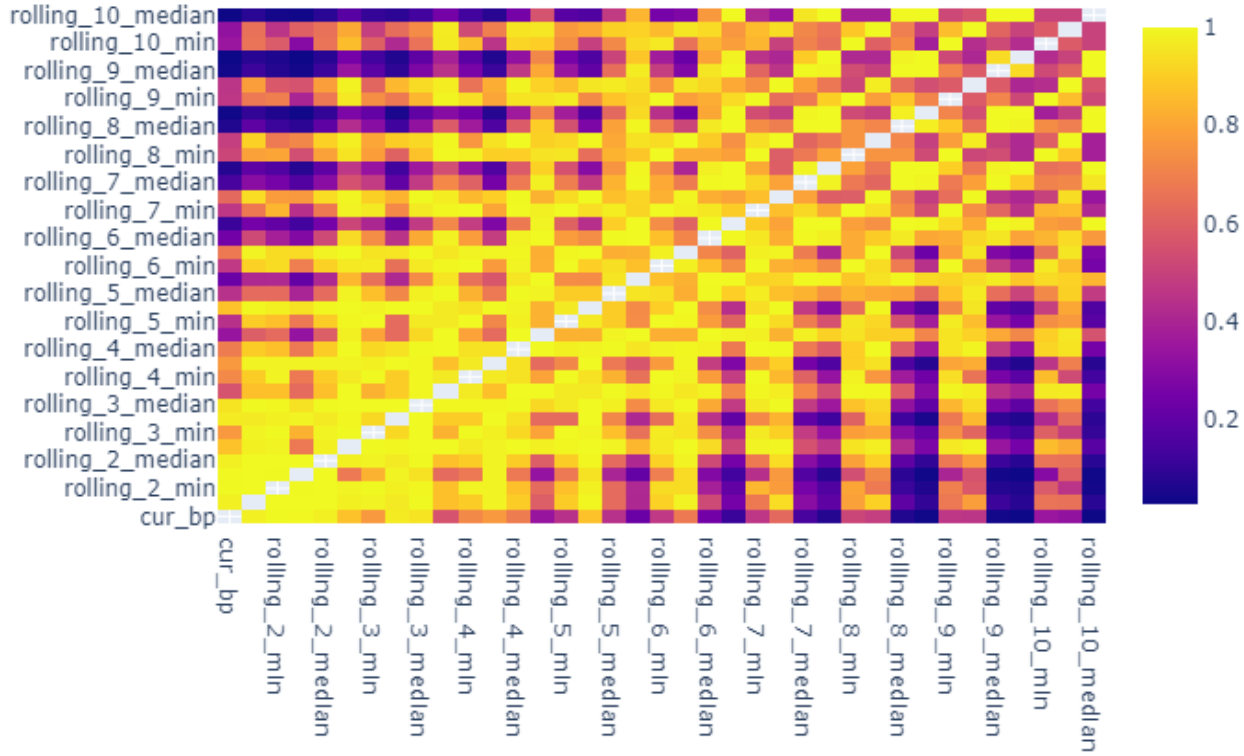

Figure 46: Kolmogorov Smirnoff test p-values for Med-Surg ICU

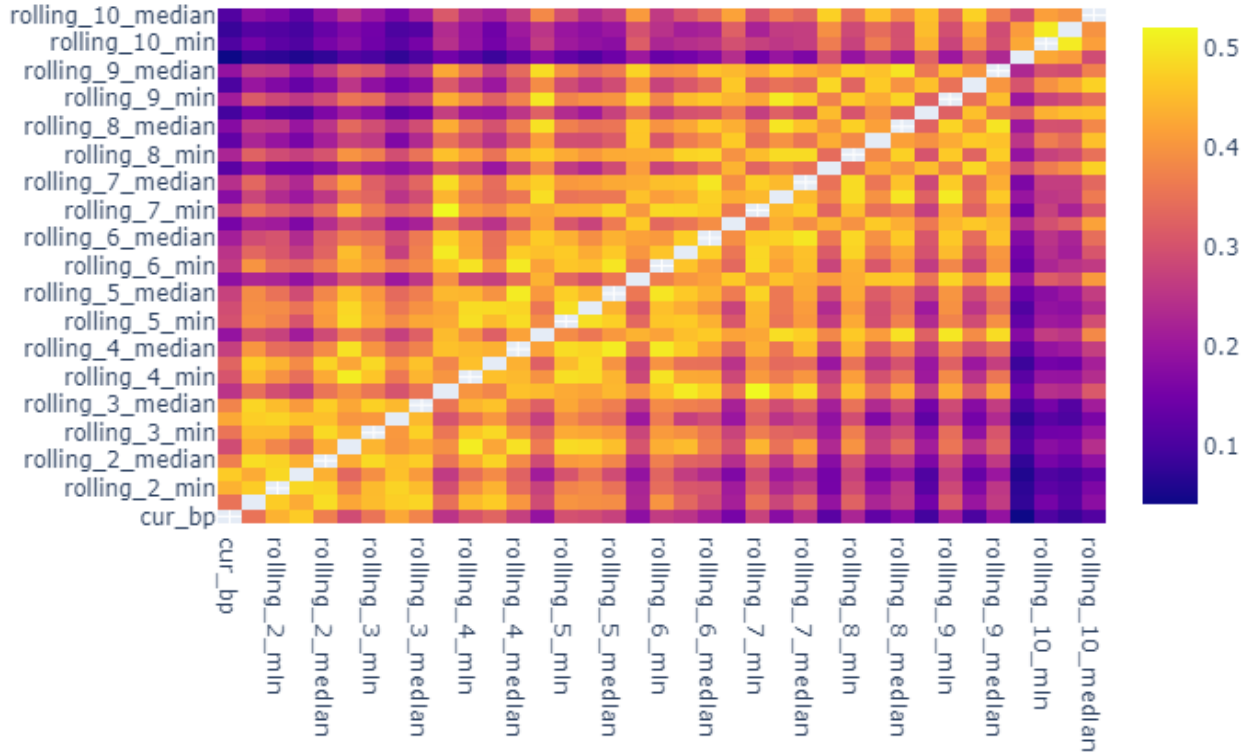

Figure 47: Variance ratio permutation test p-values for MICU

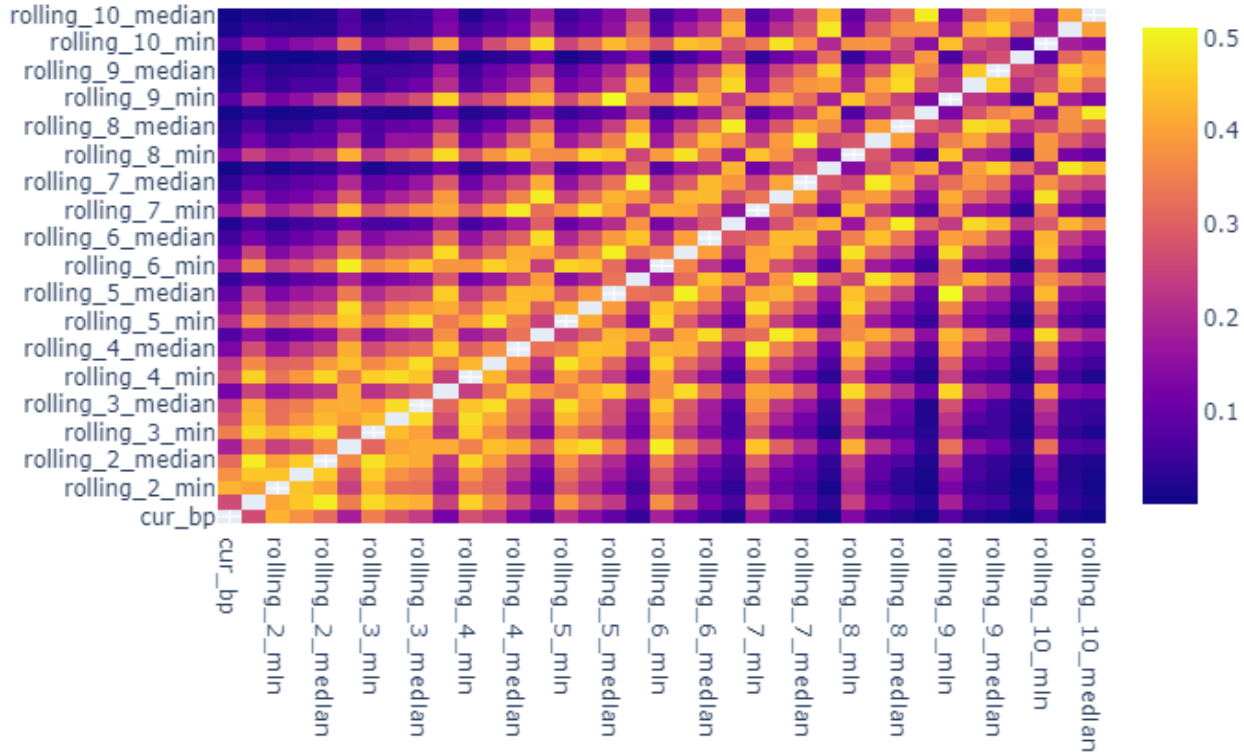

Figure 48: Variance ratio permutation test p-values for SICU

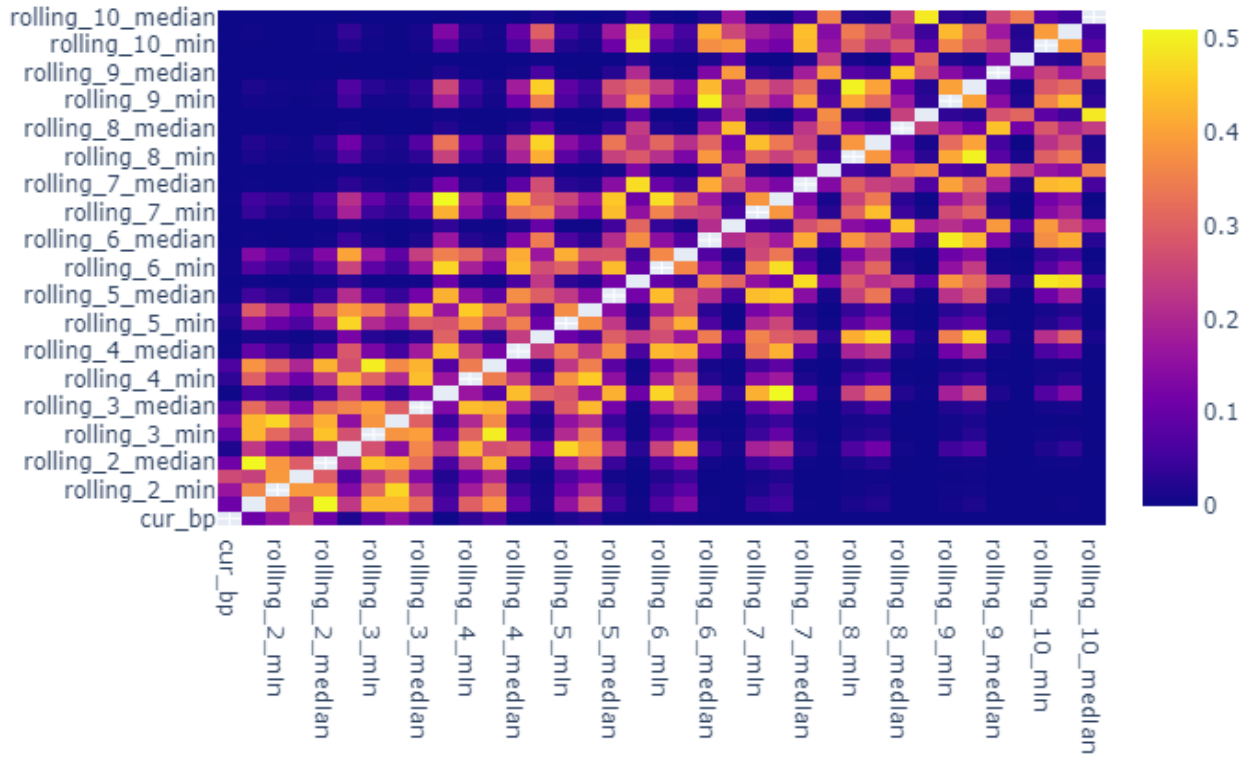

Figure 49: Variance ratio permutation test p-values for Med-Surg ICU

Since MICU and SICU demonstrated little deviation form the normal distribution, as shown in the Shapiro-Wilk test, we used applied difference in variance tests that assume a normal distribution, specifically the Levene and Bertlett tests.

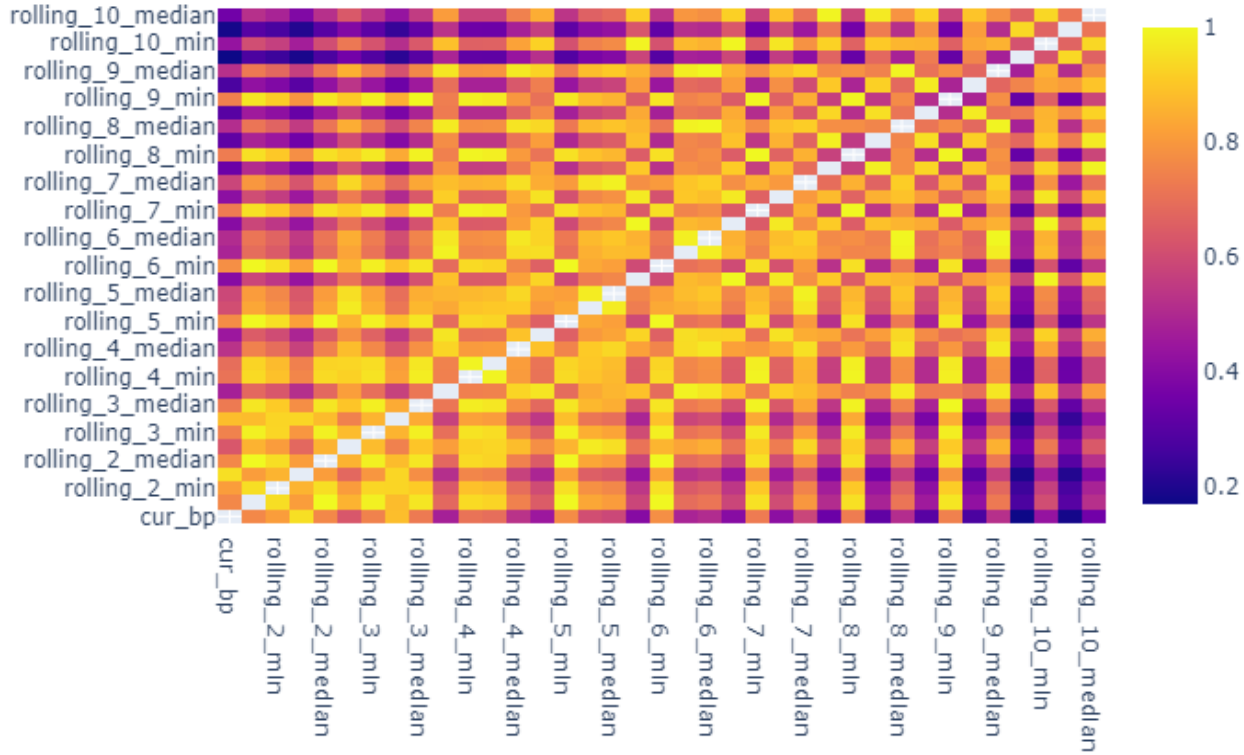

Figure 50: Levene test p-values for MICU

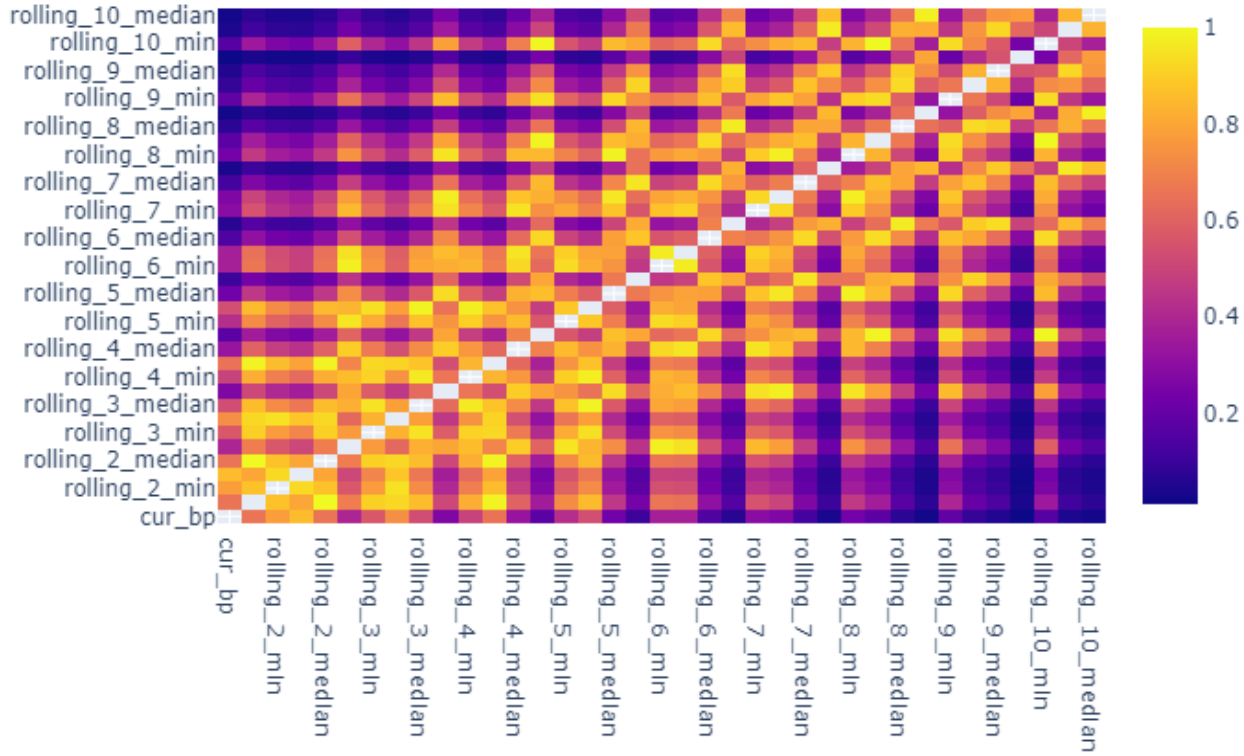

Figure 51: Levene test p-values for SICU

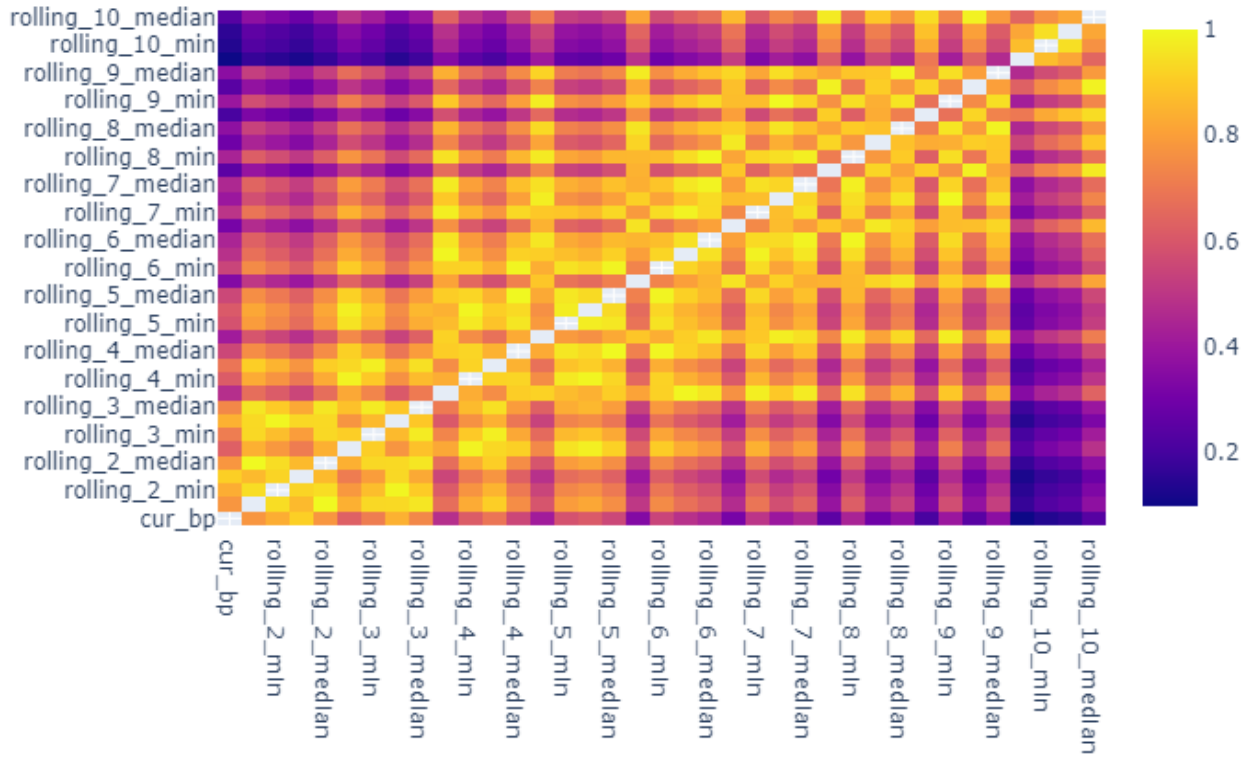

Figure 52: Bartlett test p-values for MICU

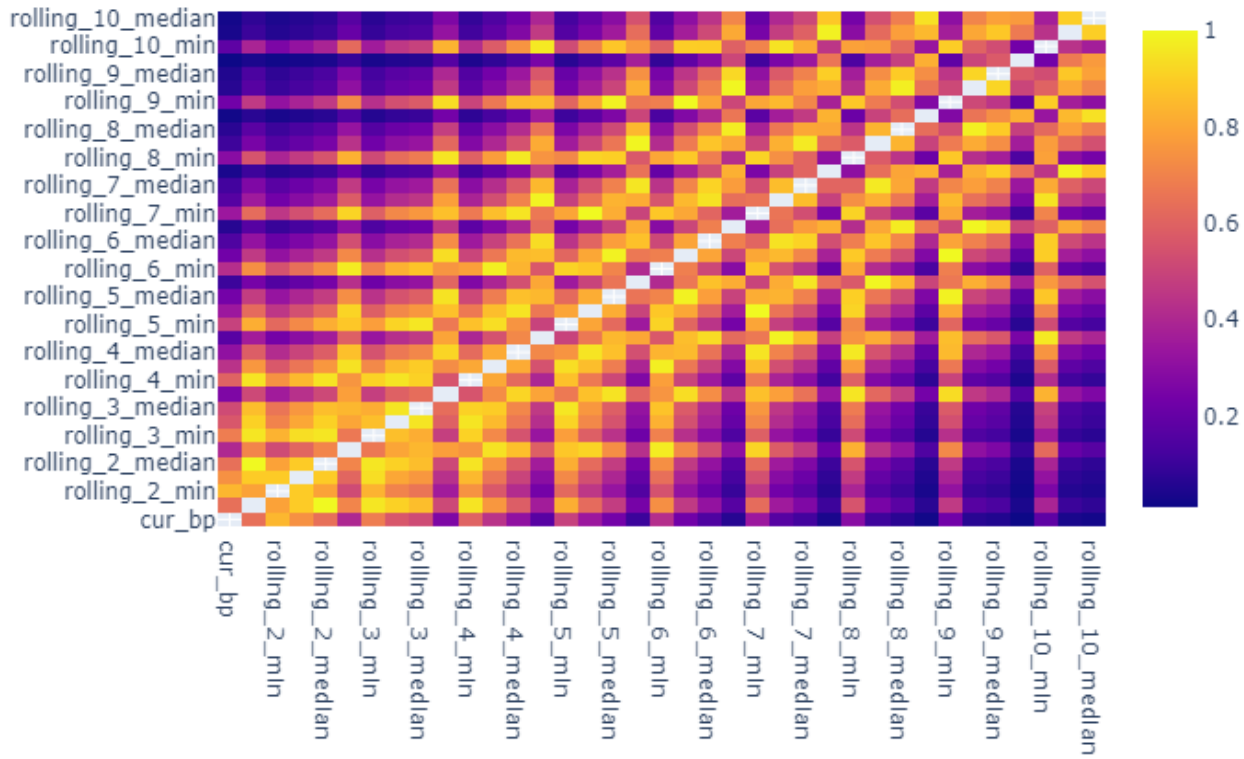

Figure 53: Bartlett test p-values for SICU

All of these heat maps illustrate the same trend: the rolling parameter k significantly influences the changes in the distribution. It appears that as the rolling parameters of the two distributions move further apart, the differences become more apparent. Additionally, it's worth noting that MICU seems to exhibit this effect only when the rolling parameter differences are larger compared to the other units, and it's p-values are higher.

A similar analysis, made only on measurements where NOR rate was altered, gives a slightly different picture:

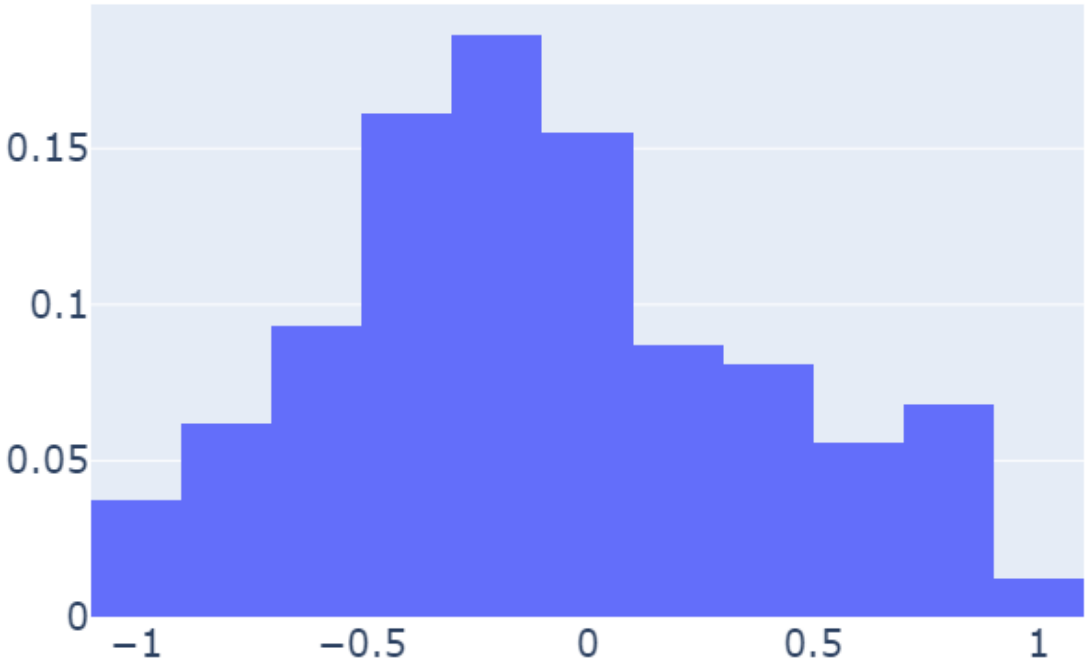

Figure 54: The distribution of the NOR-MAP correlation restricted to NOR change per patient stay - MICU

Figure 55: The distribution of the NOR-MAP correlation restricted to NOR change per patient stay - SICU

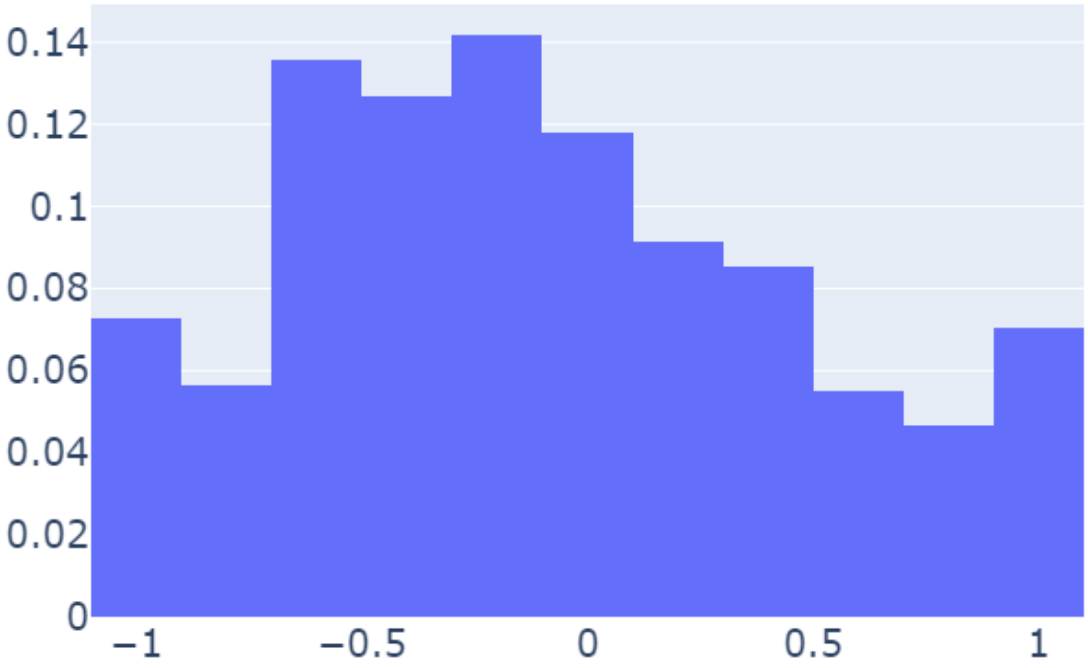

Figure 56: The distribution of the NOR-MAP correlation restricted to NOR change per patient stay - Med-Surg ICU

This distributions align better with the initial modeling of the MAP-NOR relationship, with a significant part of patient stays having negative correlations. However, most of the time the NOR rate stays the same, seemingly regardless to the MAP.

Similar Kolmogorov-Smirnov tests do not exhibit the same trends to the extent observed previously.

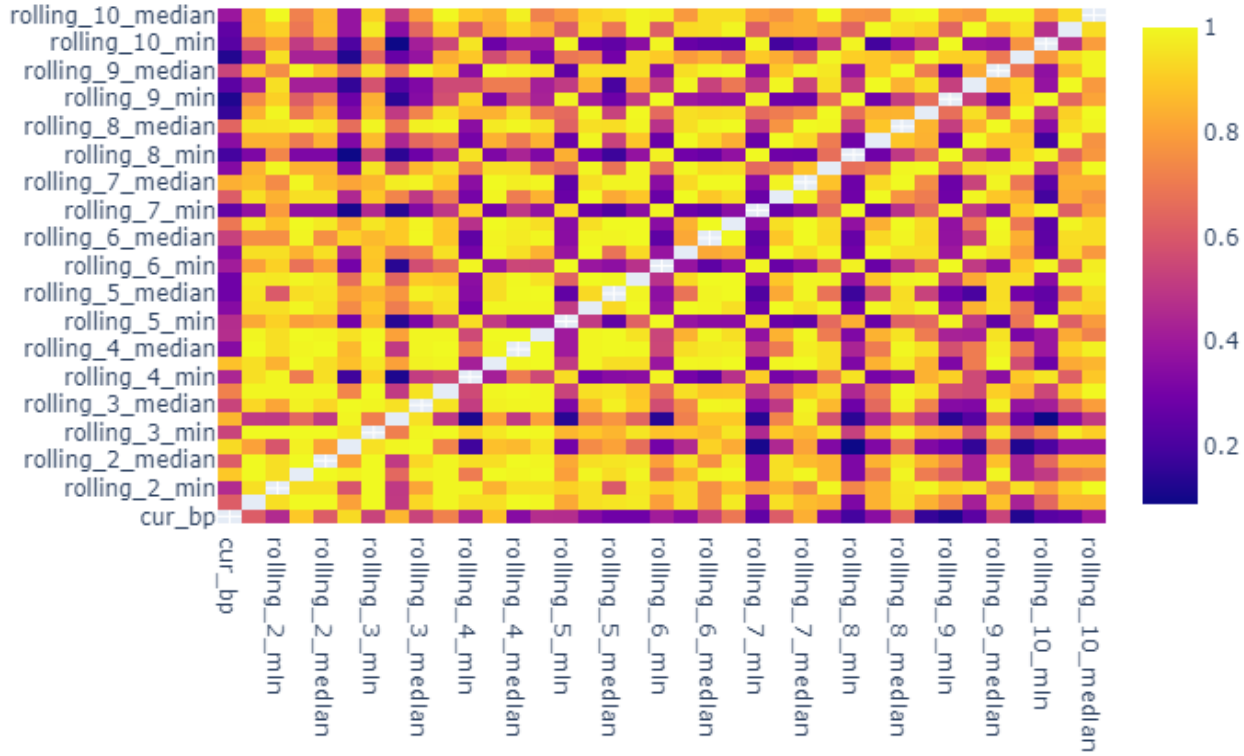

Figure 57: Kolmogorov Smirnoff test restricted to NOR change p-values for MICU

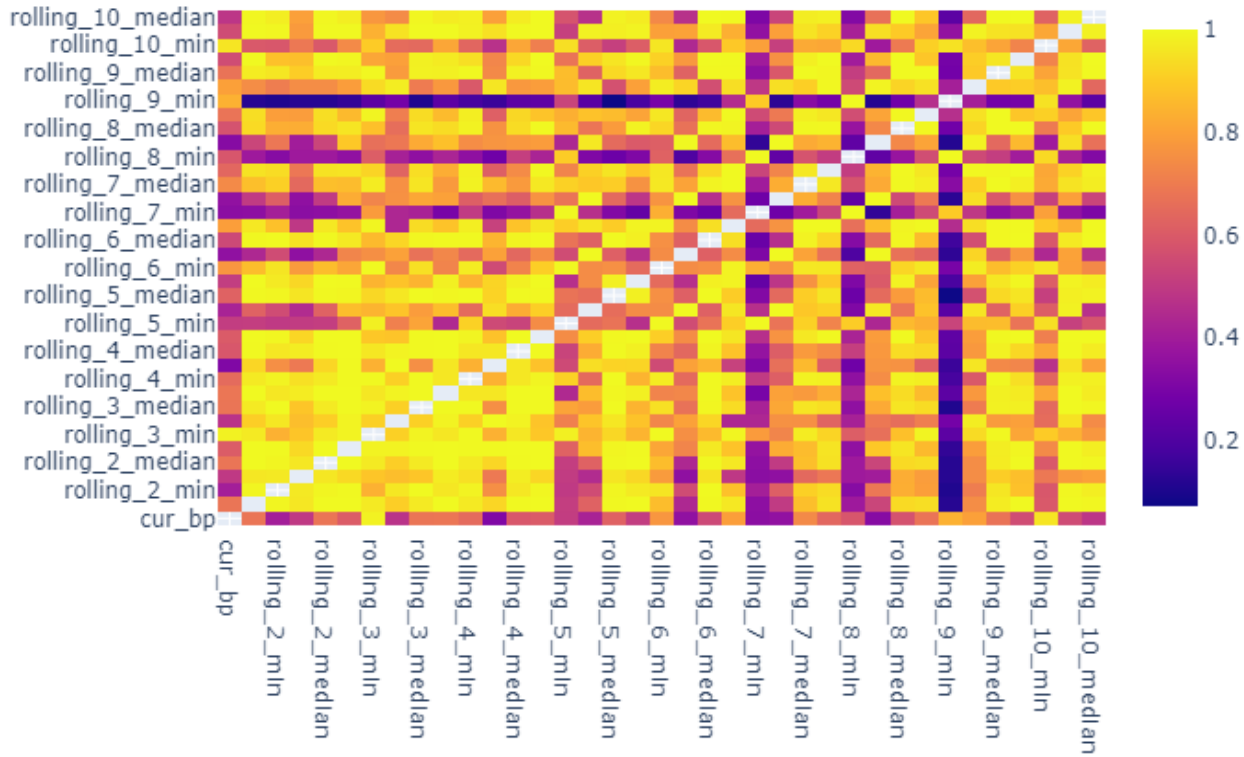

Figure 58: Kolmogorov Smirnoff test restricted to NOR change p-values for SICU

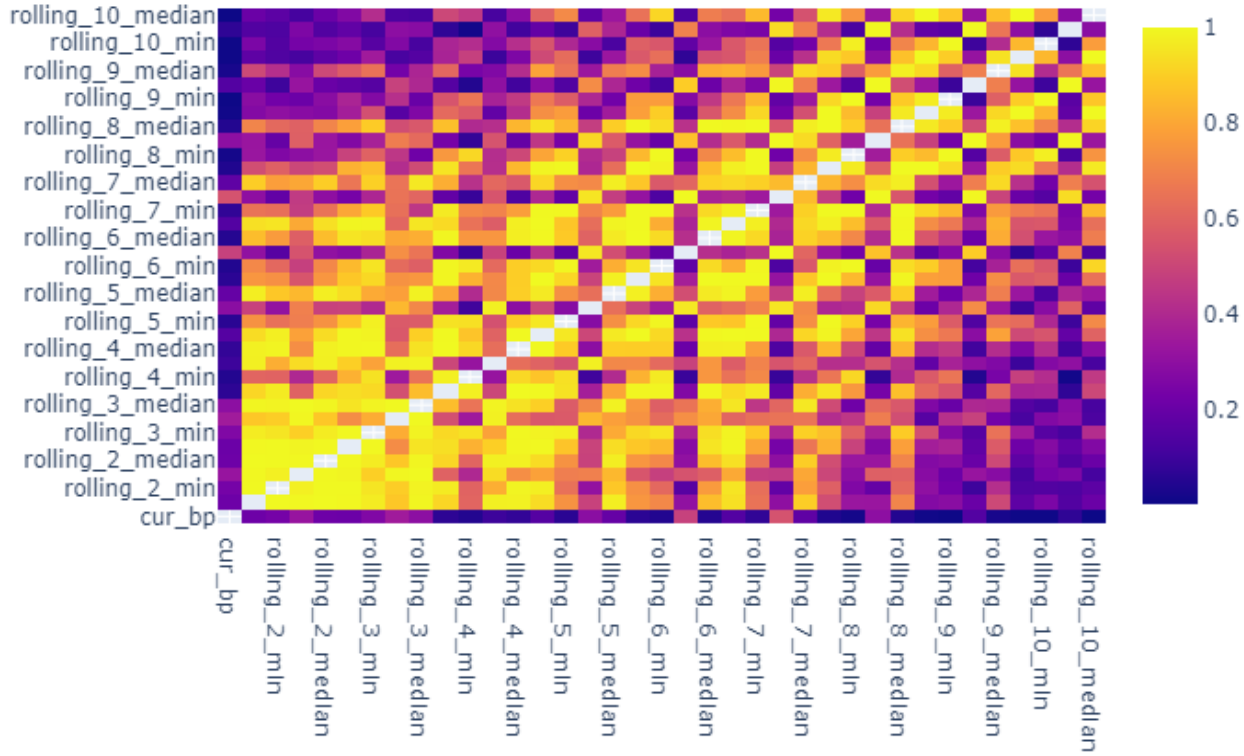

Figure 59: Kolmogorov Smirnoff test restricted to NOR change p-values for Med-Surg ICU

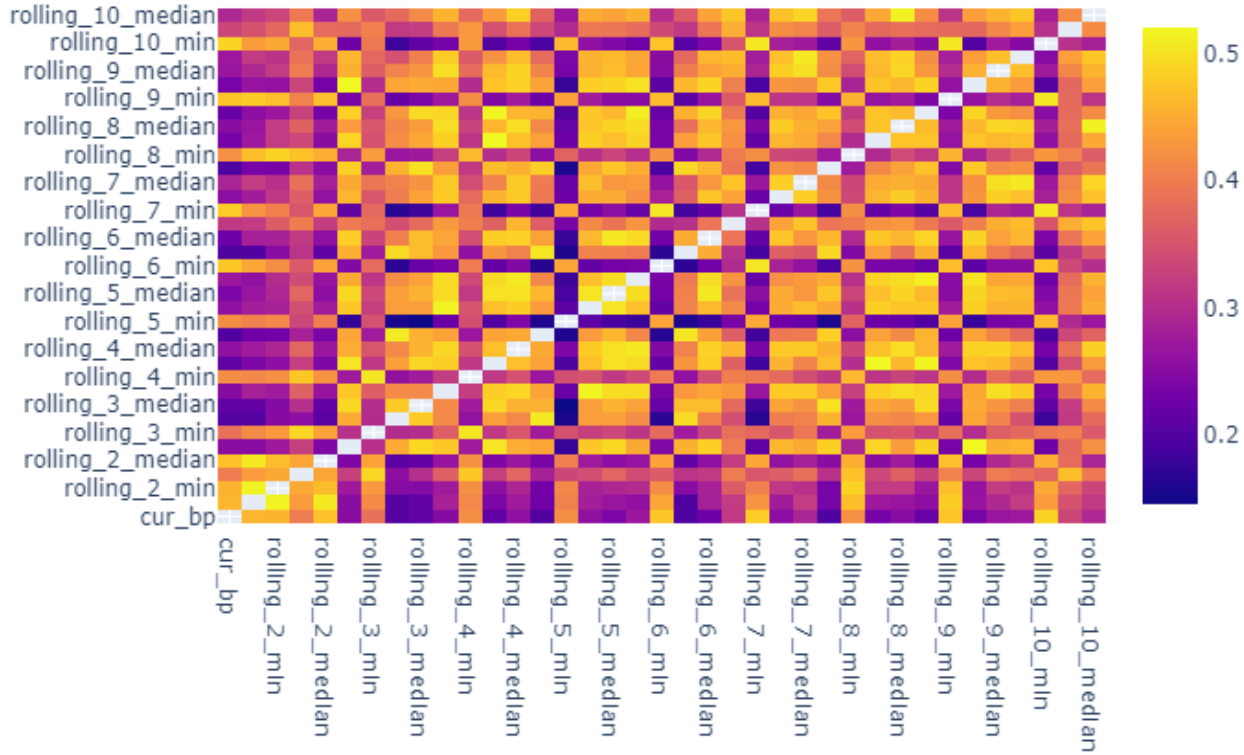

Figure 60: Variance ratio permutation test restricted to NOR change p-values - MICU

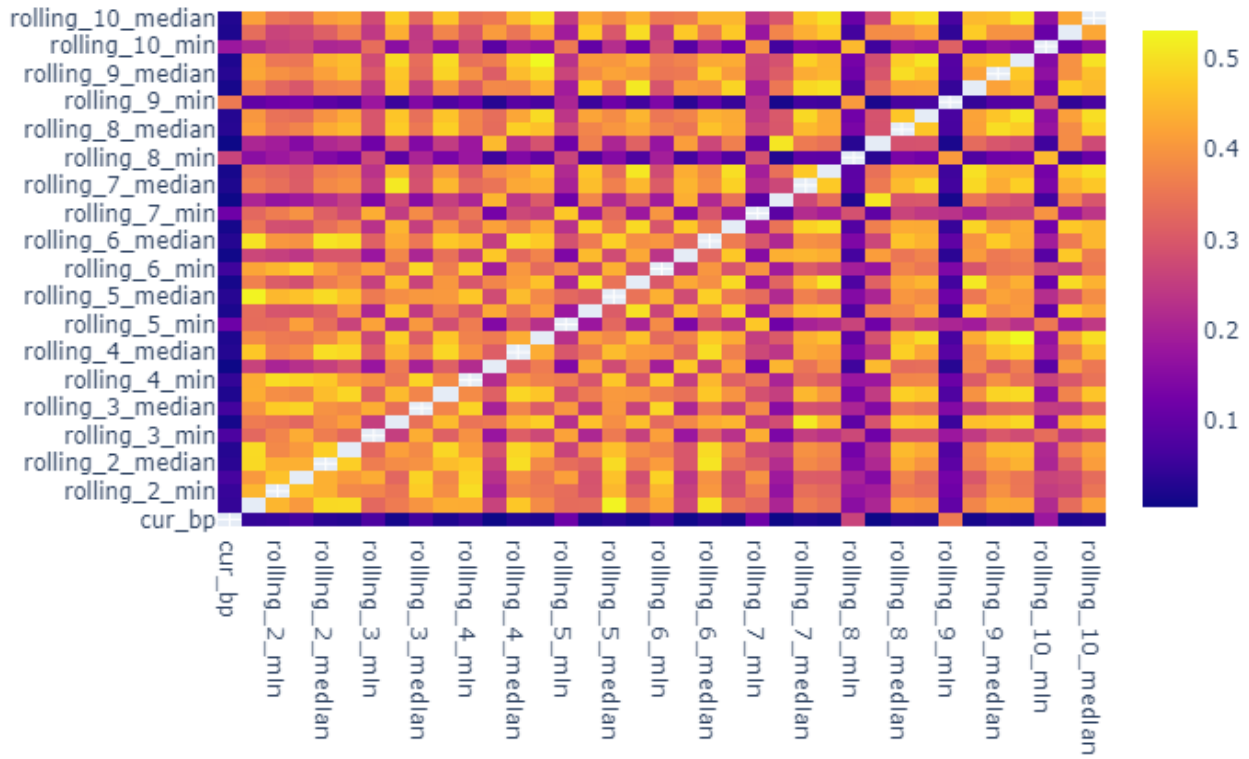

Figure 61: Variance ratio permutation test restricted to NOR change p-values - SICU

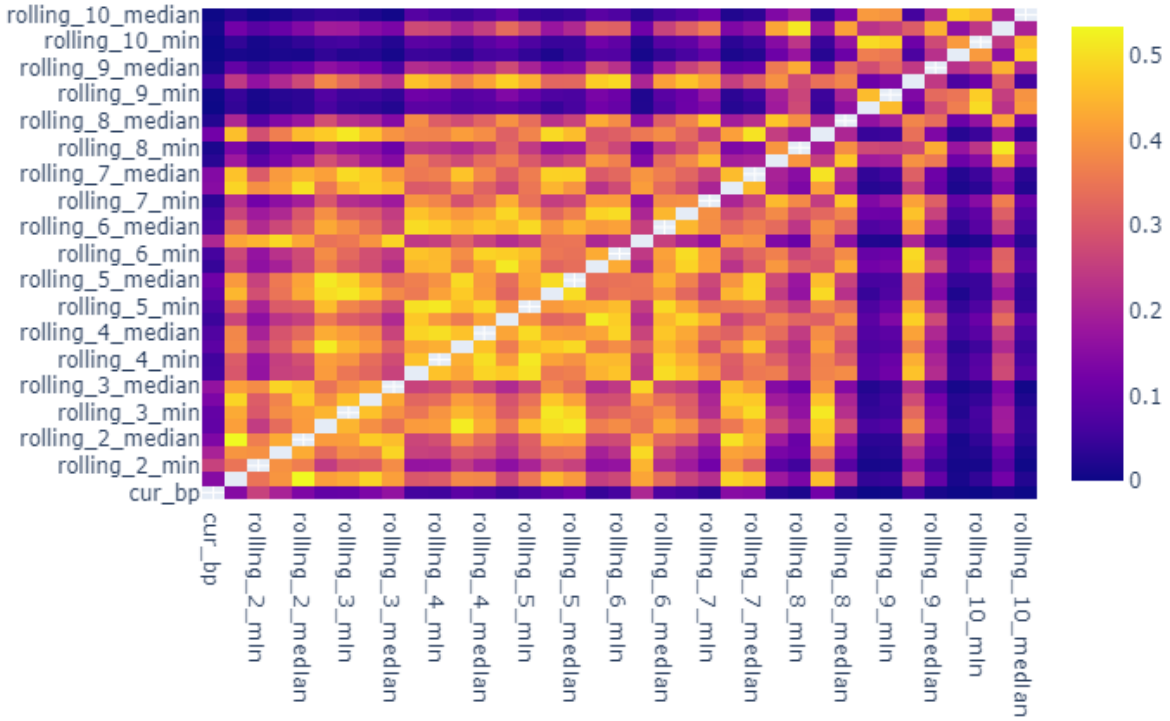

Figure 62: Variance ratio permutation test restricted to NOR change p-values - Med-Surg ICU

The variance ratio permutation test continues to demonstrate a trend that is still influenced by the rolling parameter k. It appears as a box-like pattern in the p-values, suggesting that the changes in distribution are less dependent on the differences in rolling parameters, but rather on the separation of values into two distinct sets, where one set consistently contains larger values than the other. Additionally, it's noteworthy that MICU once again exhibits less sensitivity to larger differences in the rolling parameter k.

3.2.2.3 Comparison Between Data Sets As noted, this analysis can be viewed as a comparison between countries. The first result, the empirical distributions show a significant difference

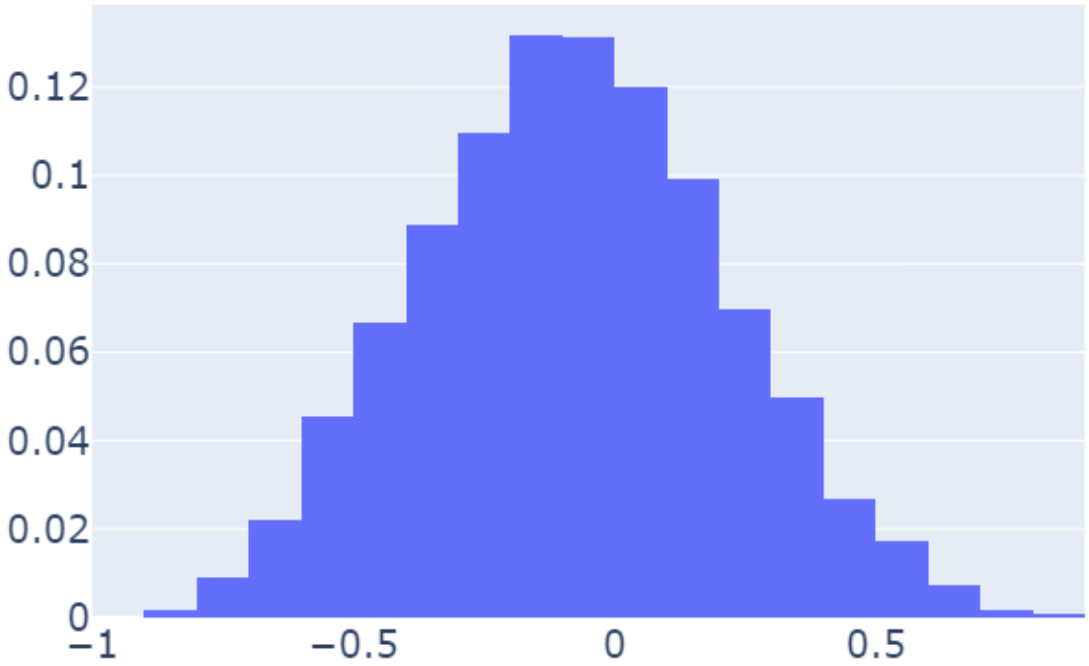

Figure 63: The distribution of the NOR-MAP correlation per patient stay - SICdb

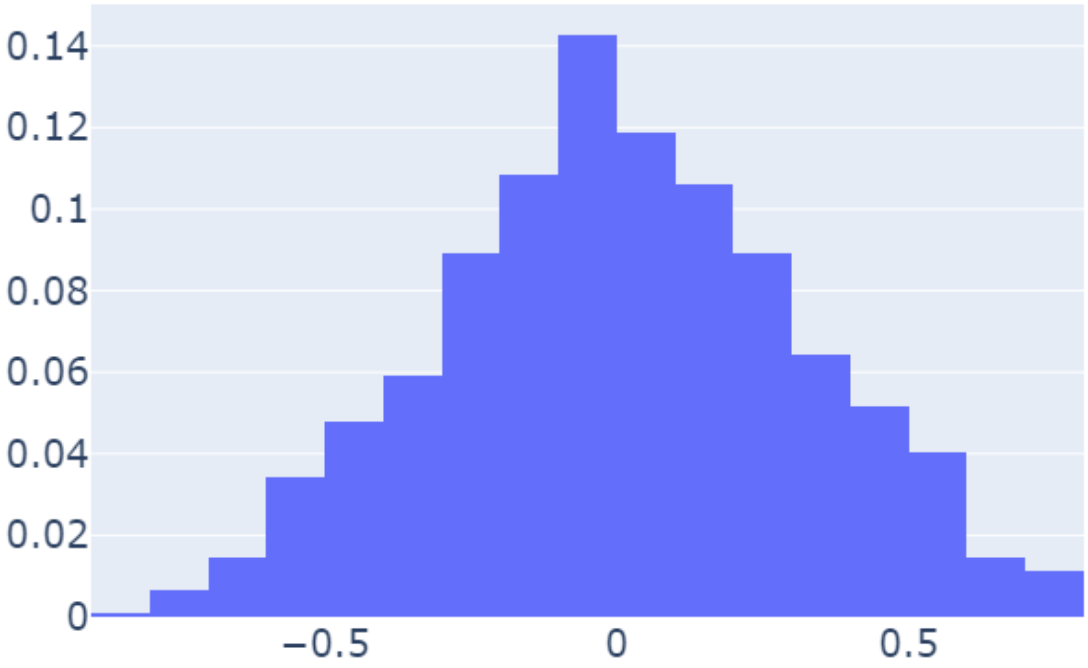

Figure 64: The distribution of the NOR-MAP correlation per patient stay - eICU

The distribution in the eICU dataset retains the normal quality previously observed, while the SICdb data set shows a shift to the left, aligning more closely with the initial expectations. This observation is evident in the trend of the means, where the correlations are significantly higher in SICdb, and the means differ between the two data sets. As well, Salburg seems to have differnt means and trends within the differnt statistics.

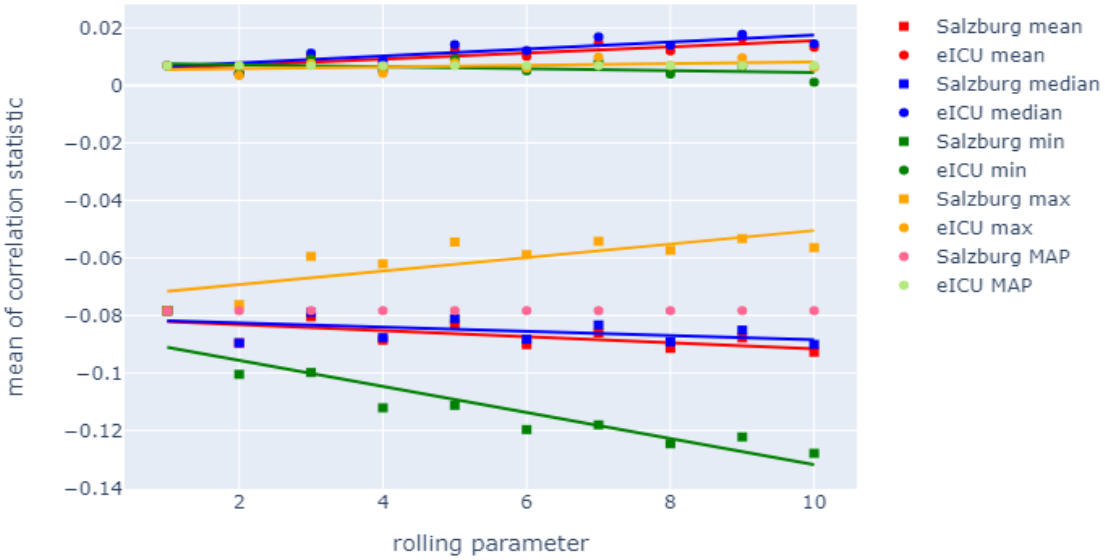

Figure 65: Mean of Correlation Statistic for rolling statistic parameter with Least-Square lines

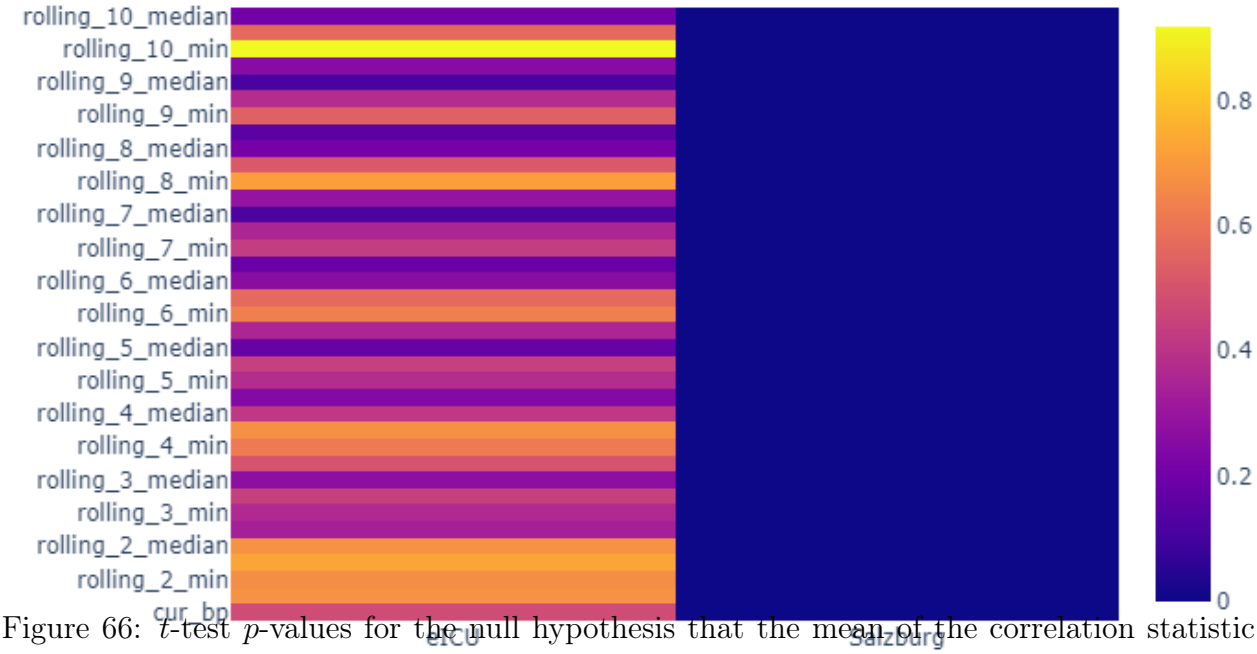

Figure 66: t-test p-values for the null hypothesis that the mean of the correlation statistic distribution for the rolling statistic has a mean value of 0

It's noteworthy that the minimum and maximum statistics show opposite trends in SICdb, while in eICU, they have approximately the same values. Another notable difference between

the two datasets is the significant change in distribution. In the eICU dataset, we observe a slow and gradual change in distribution, similar to what was seen in the unit comparison. However, in the SICdb dataset, the change is more rapid and further distinguishes between the rolling parameters.

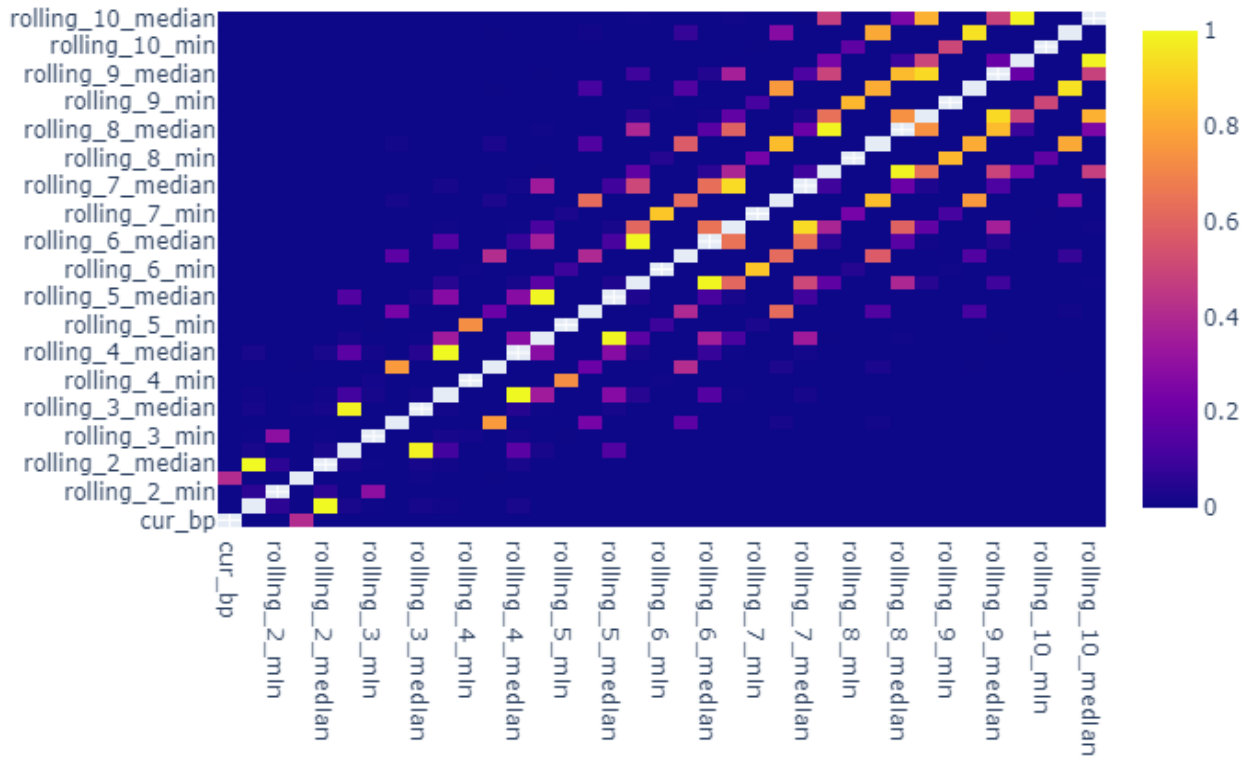

Figure 67: Kolmogorov Smirnoff test p-values for SICdb

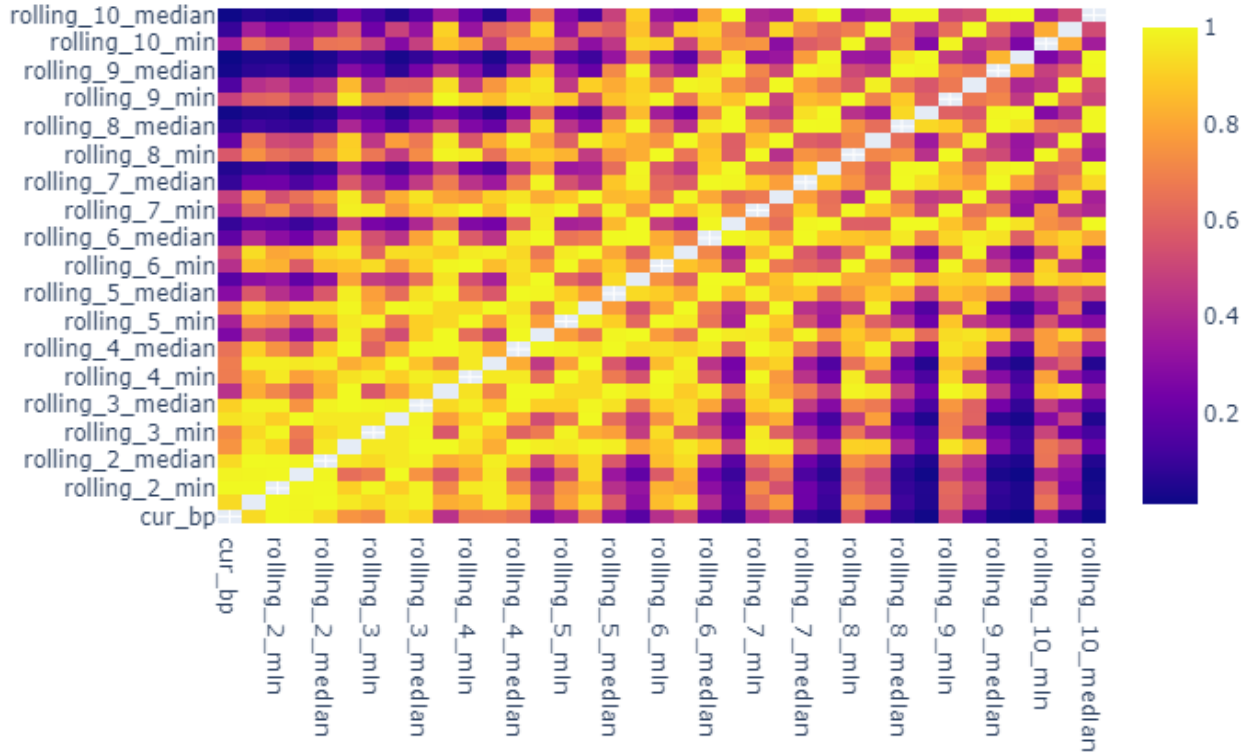

Figure 68: Kolmogorov Smirnoff test p-values for eICU

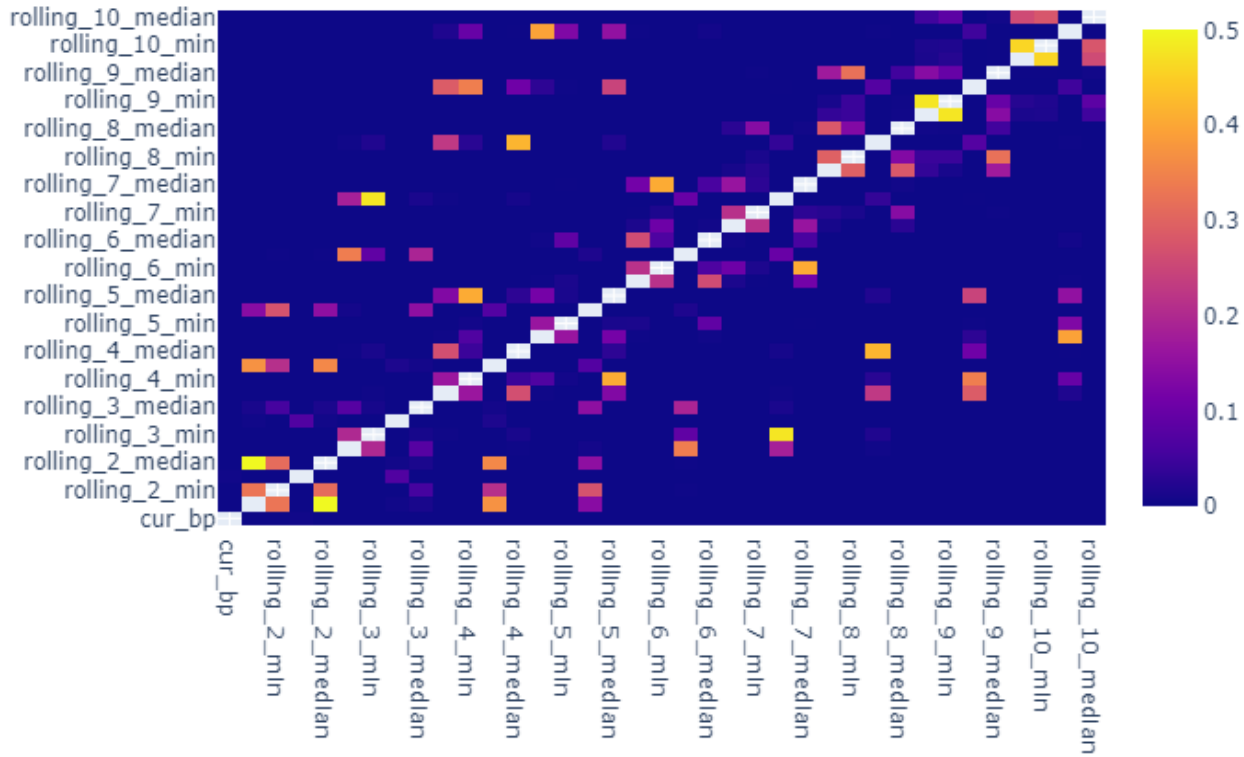

Figure 69: Variance ratio permutation test p-values for SICdb

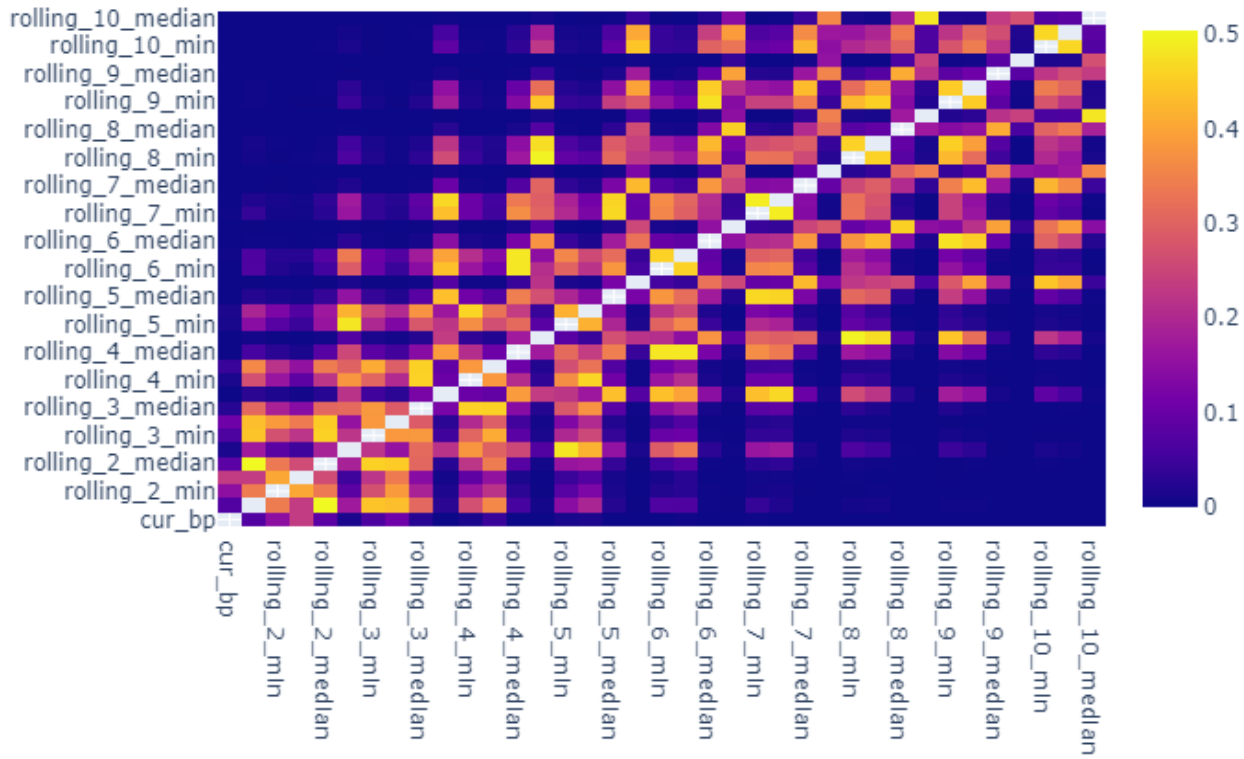

Figure 70: Variance ratio permutation test p-values for eICU

An analysis focused exclusively on NOR changes produces similar outcomes. The shift observed in the unit analysis is once again evident, along with SICdb's lower mean.

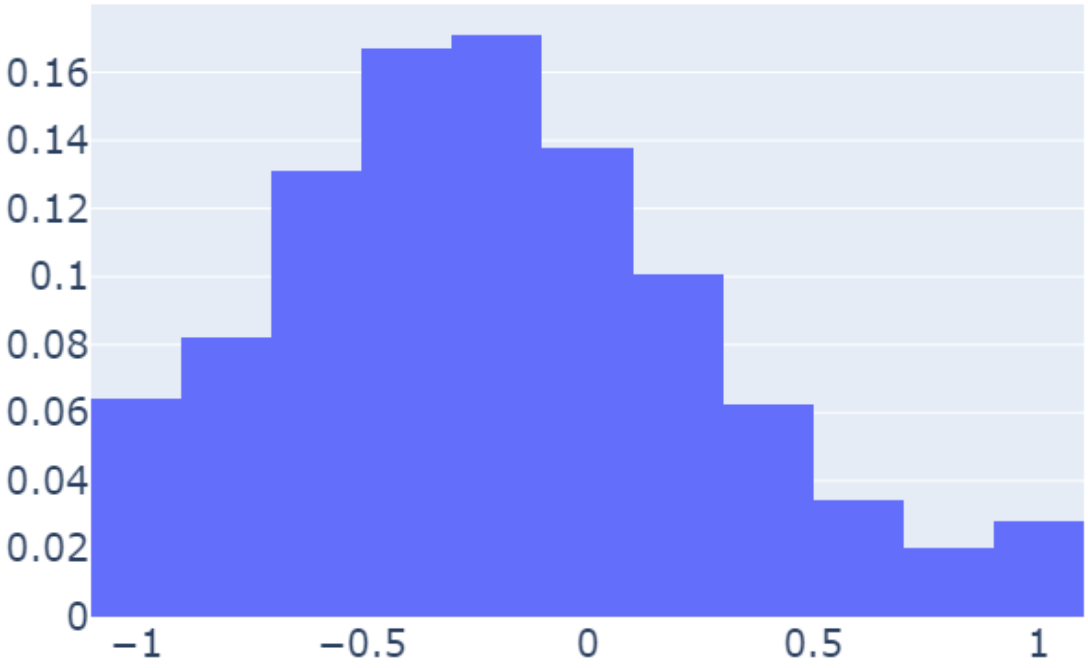

Figure 71: The distribution of the NOR-MAP correlation restricted to NOR change per patient stay - SICdb

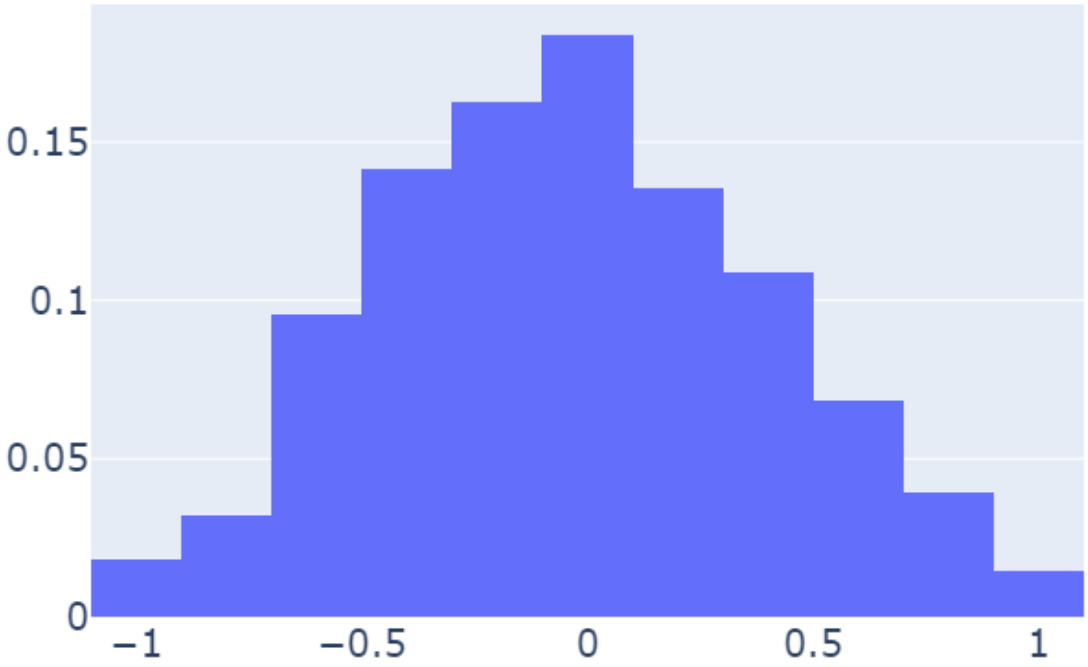

Figure 72: The distribution of the NOR-MAP correlation restricted to NOR change per patient stay - eICU

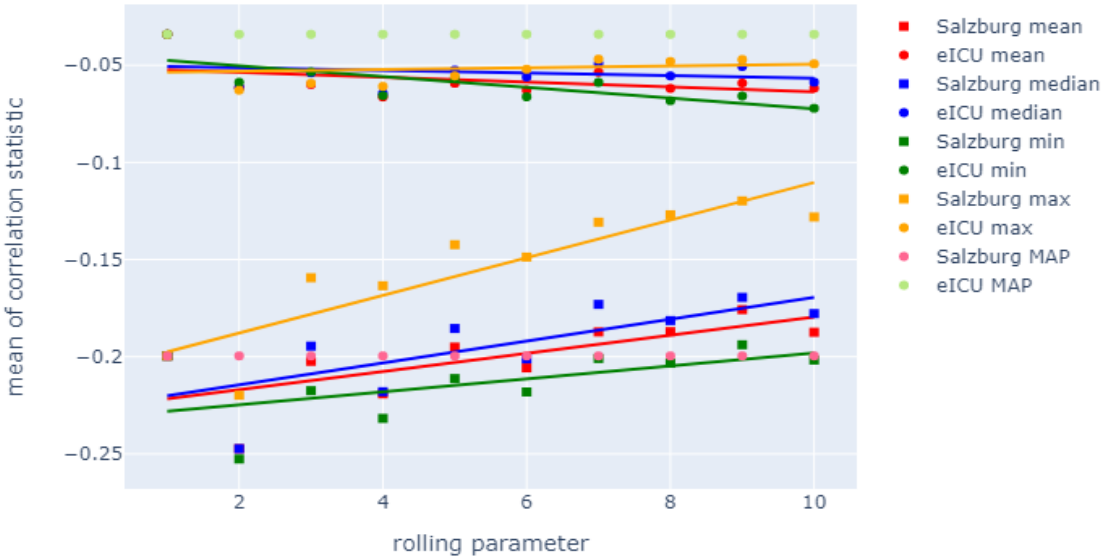

Figure 73: Mean of Correlation Statistic for rolling statistic restricted to NOR change parameter with Least-Square lines

The means are lower, and the trends now show a tendency towards 0, which is different from the previous pattern observed, except in the case of the max statistic where the trend remains consistent. The change in distribution seems to have the same properties seen without the restriction to the NOR change.

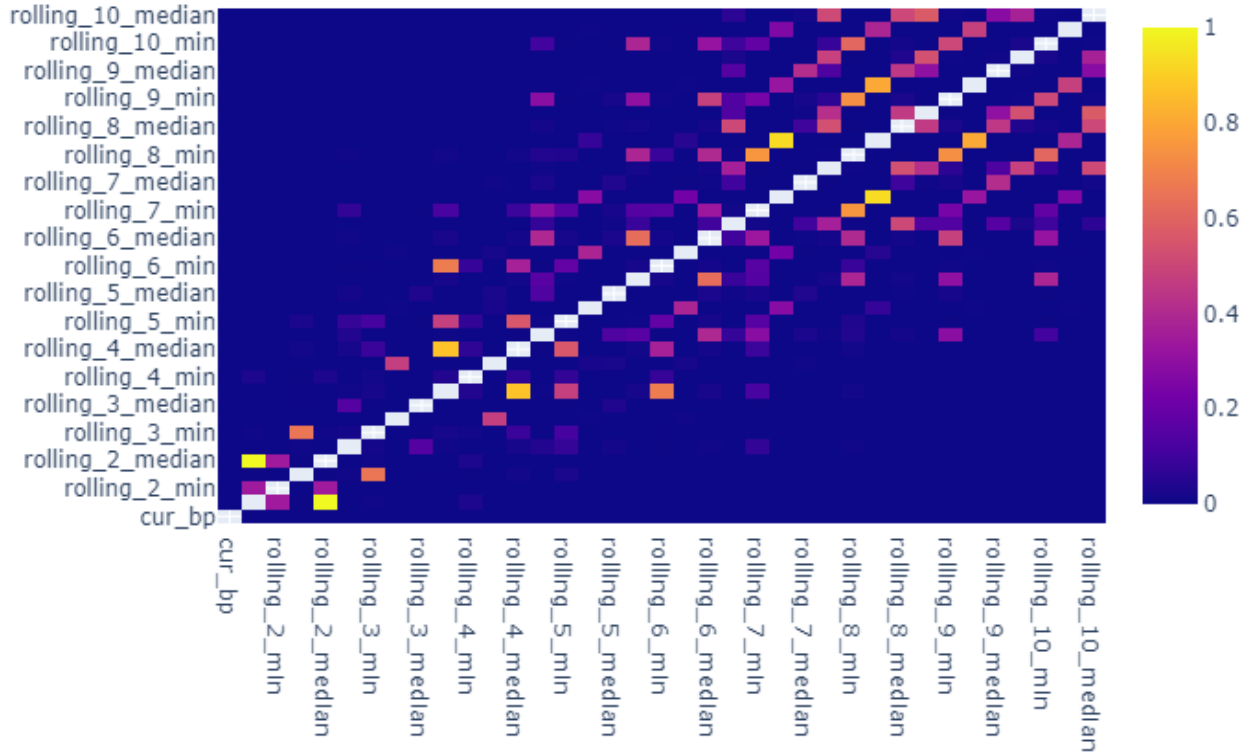

Figure 74: Kolmogorov Smirnoff test restricted to NOR change p-values for SICdb

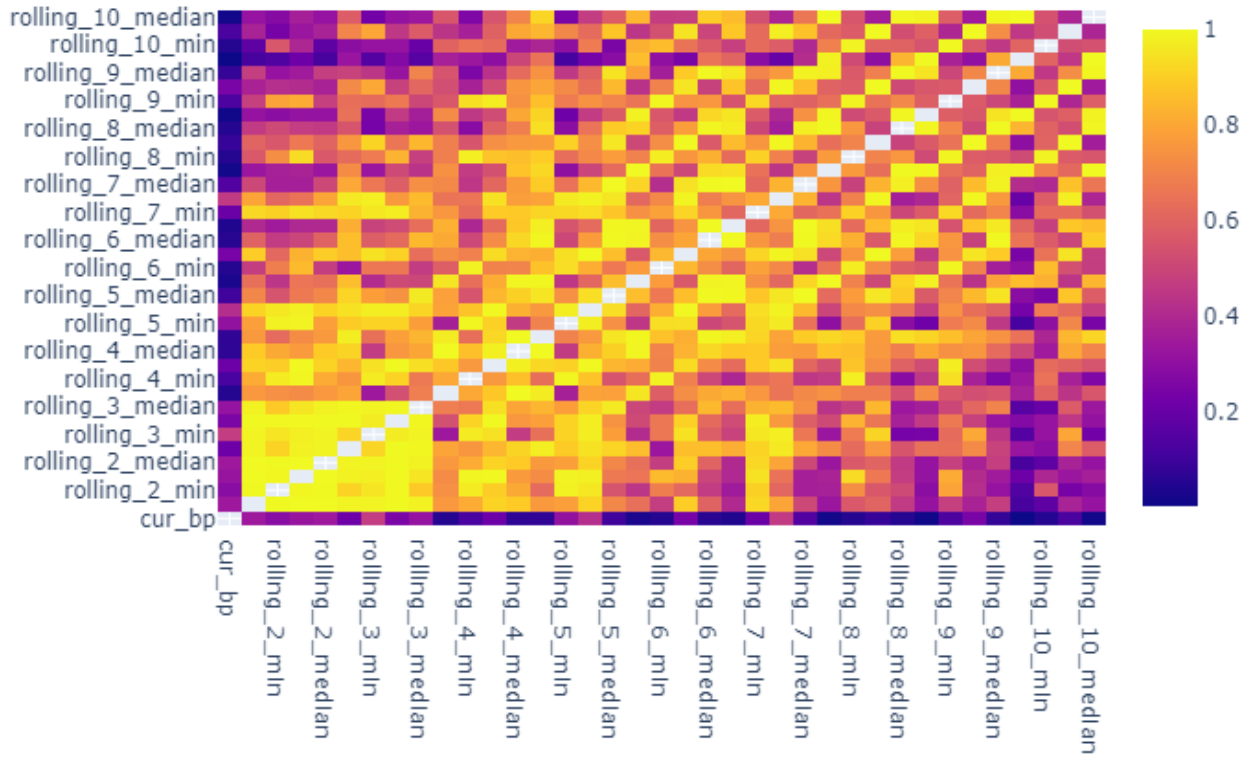

Figure 75: Kolmogorov Smirnoff test restricted to NOR change p-values for eICU

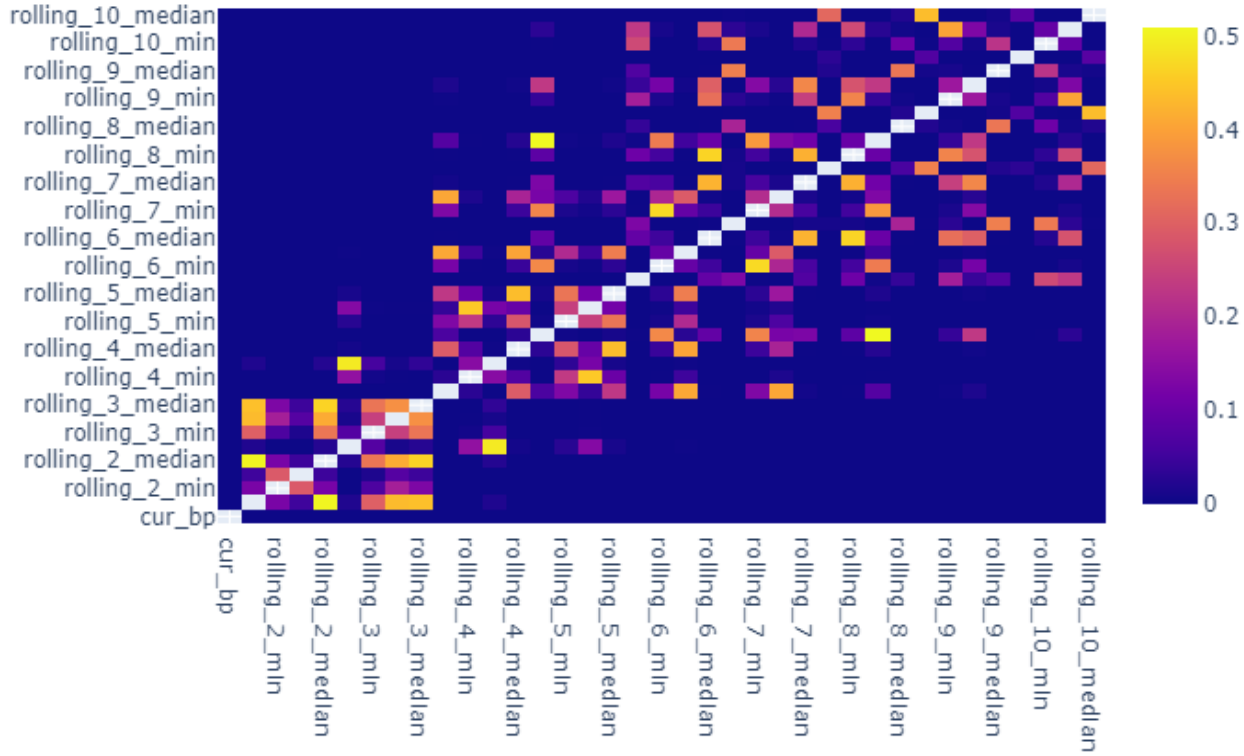

Figure 76: Variance ratio permutation test restricted to NOR change p-values for SICdb

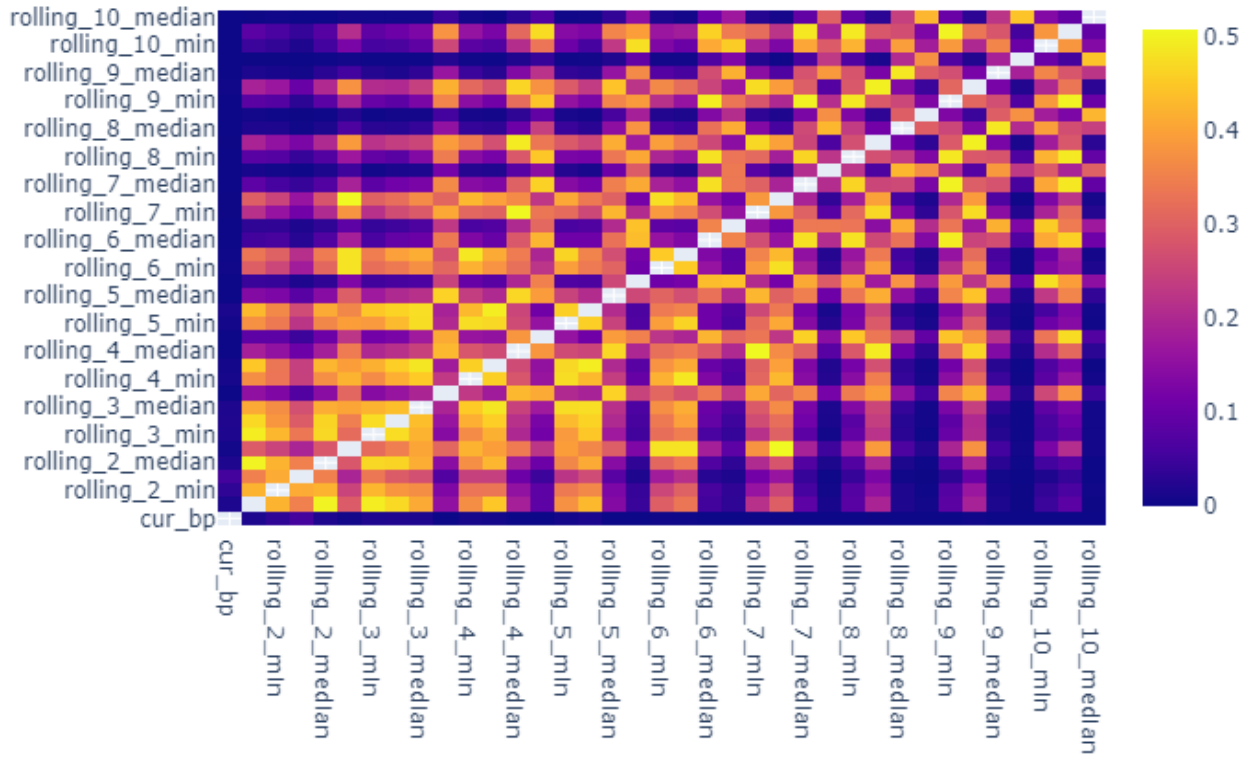

Figure 77: Variance ratio permutation test restricted to NOR change p-values for eICU

4 Discussion

4.1 A Brief Summary of the Results in Retrospective

This work aimed to investigate various forms of variability in the treatment of septic shock in different units and countries, while trying to take advantage of the relatively simple structured recommended practices, , i.e. to target a specific MAP (65 mmHg) with a single drug (NOR). As the results unfolded, particularly in the eICU data set, it became apparent that rather than identifying variability solely based on a policy related to the MAP-NOR relationship, the MAP-NOR relationship is just one of several factors. However, when the analysis is restricted to instances where the medical staff makes a decision to change the NOR infusion rate, the results align more closely with the expected MAP-NOR relationship, suggesting a correlation with recommendations. The changes in distribution imply that more factors than just a simple reading of MAP are taken into consideration. However, when NOR is applied, the correlation between MAP and NOR differs between units and countries, indicating variable policies.[12]

While this finding may pose challenges for the prospect of using AI to assist in decision-making, from a reinforcement learning perspective, the variability is evident. There are clear distinctions between SICdb and eICU, and to a lesser extent among the different unit types, Med-Surg ICU, MICU, and SICU. This implies that there are different policies to review and consider and it remains difficult to determine the optimal policy based solely on human decision-making.

4.2 Conclusions Based on the Results

4.2.1 Existence of MAP-NOR Based Treatment

The MAP-NOR relationship, in the context of the generally recommended MAP target of 65 mmHg, is undoubtedly a component of the decision-making process, although to a lesser extent than initially anticipated.

4.2.1.1 Minimizing NOR Exposure As observed in Figures 3-6 there is a notable adherence to the clinical guidelines though lower then the initial expected significance. This observation leads to an interesting hypothesis: clinicians seem to collectively prioritize minimizing exposure to NOR in their treatment strategies due to its harmful side affects. It's important to note that while this objective is widely agreed upon, there is often a lack of consensus regarding the precise methods and strategies for achieving it. These differences manifest in discussions surrounding the speed and precision required to maintain MAP within the target range. This is a hypothesis worth exploring further.

To fortify this observation, it became evident that when comparing data restricted to changes in NOR versus examining the entire dataset, a substantial difference in adherence patterns became apparent.

The previous observation, in conjunction with the near-zero Pearson correlation, underscore the presence of additional factors influencing the NOR Rate decision-making process. These may include lack of attention to changing MAP because of more urgent problems to attend such as other patients in more severe conditions.

4.2.1.2 Interpretation of MAP measurments by medical staff

One particular aspect delved into, is that treatment decisions take into account a wider range of MAP measurements when assessing alterations in the NOR rate. This pattern is clearly observable in the correlation distributions of the SICdb data set.

4.2.2 Variability in MAP-NOR Based Treatment

In the eICU dataset, there is variability, but the different policies do not seem to converge into a well-defined treatment policy, or at least not one that is mainly based on the MAP-NOR relationship. The normal distribution characteristics found in many of the distributions do not instill confidence that a specific policy can be extracted from the data. Meanwhile, in the SICdb dataset, the treatment appears to align more closely with the proposed MAP-NOR relationship, as evidenced by significant correlation values. The pronounced changes in distributions also suggest that there is potential to discover a more accurate statistic related to MAP that could better represent the MAP-NOR relationship.

A recurring theme in the correlation analysis using rolling statistics is the influence of the rolling parameter k on the changes in distributions and correlation values. This suggests that part of identifying an effective policy may depend on determining the appropriate window of MAP to consider.

4.2.3 Difference in treatment between units

4.2.3.1 Difference In Distribution

When analyzing the correlation between MAP and NOR within patient stays, the only significant finding was the influence of the rolling parameter k which was used in calculating the statistics. Interestingly, MICU demonstrated a slower change in correlation with respect to varying values of k. These trends remained consistent across the results.

4.2.3.2 Difference in Critical Points

Based on the rational but narrow assumptions we made in filtering the data, our findings when fiting polynomial models (Figures 7-10) revealed intriguing disparities between the MedSurg and Med ICU units, particularly concerning decisions to escalate NOR dosages. This hints at the existence of distinct unit-specific policies regarding NOR dosing strategies in cases of shock.

Furthermore, our research ventured into the identification of points within the NOR-MAP landscape with a high probability of occurrence. Here, we uncovered compelling evidence of criticality — situations in which the majority of decision-makers converged on similar choices. However, it is noteworthy that the locations of these critical points exhibited variations between the units, suggesting disparities in how different units perceive and respond to critical scenarios.

To conclude this phase of our research, it is evident that delving deeper into the understanding of variability, holds significant promise for future investigations. This avenue of inquiry promises valuable insights into the complexities of decision-making in critical care settings.

4.2.4 Difference in treatment between Data Sets

The results show that the MAP-NOR relationships are different between the data sets. Although both data sets exhibit a somewhat cautious approach regarding the alteration of NOR infusions, SICdb shows a stronger correlation with MAP, both in the overall patient stay and when restricted to changes in NOR. Another difference is that SICdb demonstrated a high sensitivity to the change in rolling statistic and parameter, while eICU showed slow trends, relying mostly on the window size k. This could be meaningful when trying to train models based on the behavior of the different data sets.

4.3 Limitations

4.3.1 Lack of Alignment to the Original Treatment Model

The most significant limitation, this work used the underlying assumptions that the NOR-based treatment was both dynamic and heavily dependent on the MAP measurements. Both assumptions failed to demonstrate themselves to the extent that would justify many other assumptions used, such as a negative correlation between MAP and NOR, using almost exclusively the MAP in the analyses, and the binning of entries by positive and negative NOR changes.

Restricting the data to entries where NOR was altered improved of alignment, suggesting that the assumptions have some merit. However, to create a more concrete analyses, it is critical to incorporate more data than just MAP and develop a better-fitting treatment model.

4.3.2 Time resolution

Circling back to the issue of time resolution, while 5 minutes was adequate for many analyses and did reveal differences between the units, it's important to recognize that a finer resolution will likely be necessary for the development of a concrete recommendation system, as it can capture more nuanced patterns and variations in the data.

4.3.3 Patient Stay Population Size

It's worth noting that a substantial number of patient stays were filtered out based on strict criteria to maintain a similar setting within each trajectory. While the issue of data size is less prominent in the larger SICdb data set, it's important to acknowledge that this study focused on a specific group of patients who met the criteria for inclusion.

4.4 Future Work

The work conducted in this study aimed to lay the foundation for an AI recommendation system, possibly implemented using Reinforcement Learning frameworks. Initially, the goal was to demonstrate that MAP measurements alone are sufficient to determine the NOR rate and that these measurements are interpreted differently within various groups. While the latter was found to have significant basis, and therefore can be used for further research

specifically in Reinforcement Learning. While the latter goal was achieved, indicating potential for further research, particularly in the field of Reinforcement Learning, future research in this area will need to incorporate data beyond MAP measurements and may benefit from utilizing the better time resolution available in the SICdb dataset.

5 Resources and Methods

5.1 Preprocessing

5.1.1 Choosing Patients and Measurements

The eICU dataset includes around 19,000 patients who received NOR treatment for septic shock. We want each of the patient stays in our population to have some qualities: A consistent and high-quality measurements, particularly concerning MAP and NOR data. Describing a medical journey mainly based on the MAP-NOR relationship. And the most important quality we want, that all of the patient stay describe the same scenario to some extent.

The second quality remains somewhat ambiguous; therefore, we operate under the assumption that of septic shock occurs under closely monitoring of MAP and administering vasopressors. With this perspective in mind, filtering out patients who exclusively received NOR will help retain patient data that demonstrates a strong decision-making process based on the MAP-NOR relationship. Following the previous step, we will proceed to exclude all patients who did not receive Norepinephrine (NOR) within the first 24 hours of admission. This refinement ensures that we retain patients who received a MAP-NOR treatment right from the outset, further enhancing the this quality in out population.

To make a more uniform patient population where medical staff operated under similar conditions, we will exclude all entries occurring after the first 48 hours from admission. While this step is not a direct patient filter, it plays a crucial role in enhancing the comparability of patient stays. This is because, during the initial 48-hour period, most of the critical care that takes place within the ICU involves limited procedures or tests.

Finally, we will exclude all patients with a MAP measurement break of more than 30 minutes occurring within the first 6 hours of their stay. This is crucial because such a break often signifies the possibility for an event that could be a "game changer" in the patient's trajectory, making their case distinct from others in the dataset. For the same reasons, we will implement an entry-wise filter for all entries occurring after a 30-minute break.

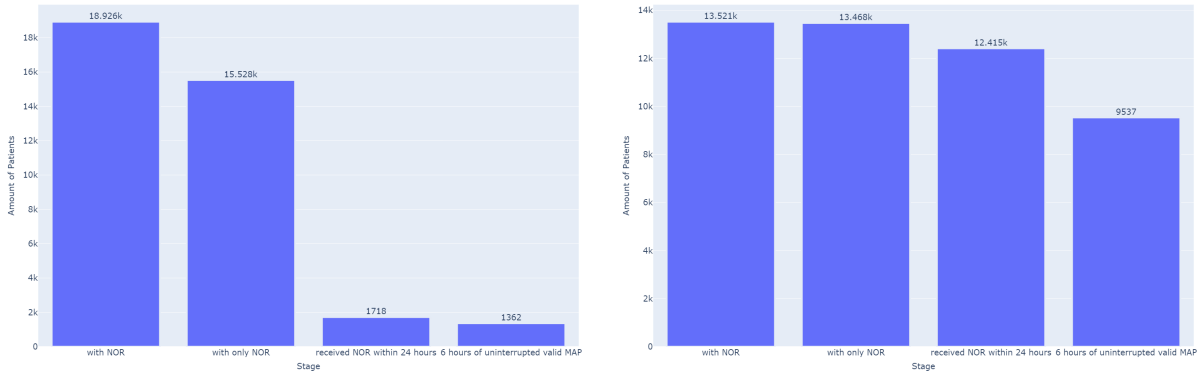

Figure 78: Number of Patient Stays after every stage of filtering, eICU on the left, SICdb om the right

Figure 79: 3

5.1.2 Denoising Time Series Data For MAP

Like any measurement, in addition to the inherent changes in blood pressure, the measurement device as well as the environment introduce a component of errors. In practice, errors can be attributed to two distinct sources: instrumentation noise and artifacts. These two sources differ in terms of their scale and distribution. Furthermore, the presence of artifacts posed a challenge when attempting to remove the instrumentation noise.

5.1.2.1 Instrumentation Noise This noise is layered on a proper reading of MAP, which includes factors such as discretization and tolerance. It is reasonable to assume that this type of error has a mean of 0 since it's primarily controlled by the manufacturer's hardware and software.

The process of denoising this kind of noise can be described as smoothing, as we anticipate the resulting outcome to be a continuous trajectory where changes are gradual and influenced by nearby measurements. Since we want the data to help us determine weather the medical staff used MAP in the decision making process, it is important to imitate the way the medical staff interprets the raw data, which is usually very close to the the raw data they received.

As a result, it is essential to employ conservative techniques, specifically ones that preserve many measurements and upwards and downwards trends in consecutive values. An aggressive smoothing approach, even if it better aligns with the MAP, could potentially distance us further from the data that is relevant for estimating the decision-making process.

We will quickly review the different techniques we used and eventually didn't use. In other related work, which focused on the slightly different objective of predicting MAP, Kalman filter is a popular model.[10][11] In practice, we found that this filter was overly sensitive to artifacts, possibly because of its Markov property, which doesn't take advantage of the broader context available. The desire to leverage more measurements is effectively

incorporated into the Gaussian process, a technique frequently employed for denoising and time series prediction. Although it didn't completely eliminate artifacts, it did not worsen the removal of near measurements affected by instrumentation noise. The primary drawback of this method lay in its heavy reliance on extensive assumptions and its extended runtime, ultimately leading to our decision not to adopt it. Yet, this approach may be used in the future, where we would want to more closely control the denoising process.

Ultimately, the most successful method for removing instrumentation noise was a rolling rank order filter. Under the assumption that the MAP changes should occur gradually, when considering a small value for k, the data points x_i through x_{i+k} should exhibit small deviations from the mean of their values.

We identified outliers by comparing measurements to their k-rolling mean, $\overline{x_i} = \frac{1}{k}\sum_{j=i}^{i+k} x_j$. Outliers were identified as measurements with substantial deviations from their respective k-rolling means and replaced with $\overline{x}_i$. This straightforward method has demonstrated its effectiveness as a smoothing tool while also being highly efficient to compute. It is the chosen method for our analysis.

5.1.2.2 Artifacts During tests, procedures, or instances when the tube connected to the patient's artery is misplaced, the device continues to record and log information into the eICU system. However, it loses its connection to the patient's arterial pressure, resulting in "garbage" data that should be removed. Unlike the measurement noise, this noise is characterized by a significantly larger magnitude and is not commonly encountered.

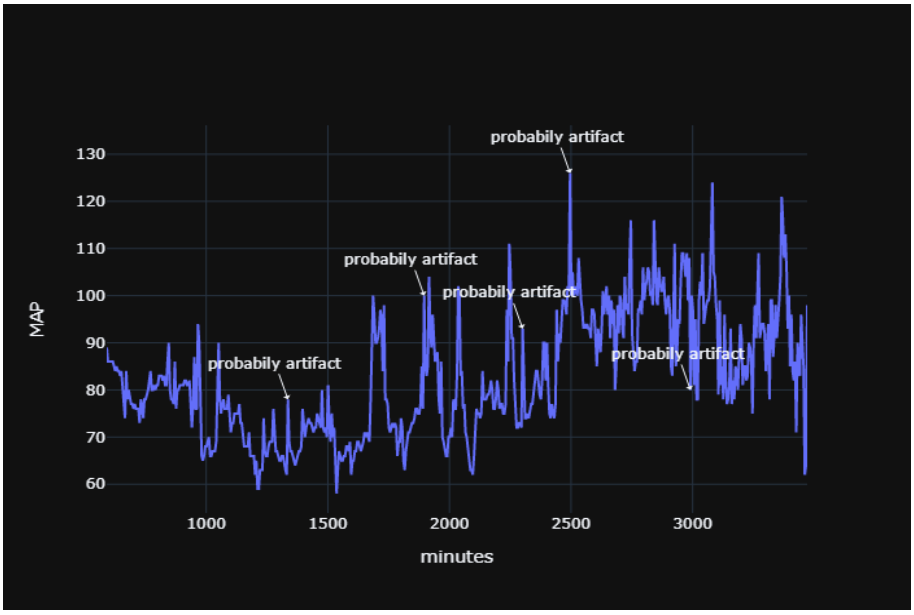

Figure 80: MAP for Time for Patient-Stay ID 2893284

Based on this information, it's reasonable to conclude that measurements with MAP values outside the range of 30 to 160 are likely not outliers resulting from any measurement noise model. Instead, they are more likely to be a false readings. Therefore, filtering out such measurements is an effective initial step.

To address false readings that didn't occur due to active treatment, such as tube misplacement, an assumption was made that these malfunctions should be corrected by the staff within a 10-minute time frame, which aligns with a single eICU measurement interval.

Based on the previously stated assumption that MAP changes occur gradually, it follows that the changes of consecutive MAP values, represented as $(x_i - x_{i-1}, x_{i+1} - x_i)$, should be close to $(0, 0)$. In our false reading model, if x_i is an outlier, we anticipate that x_{i+1} and x_{i-1} will be close, indicating prompt resolution by the staff. This implies that $x_i - x_{i-1} \approx -(x_{i+1} - x_i)$.

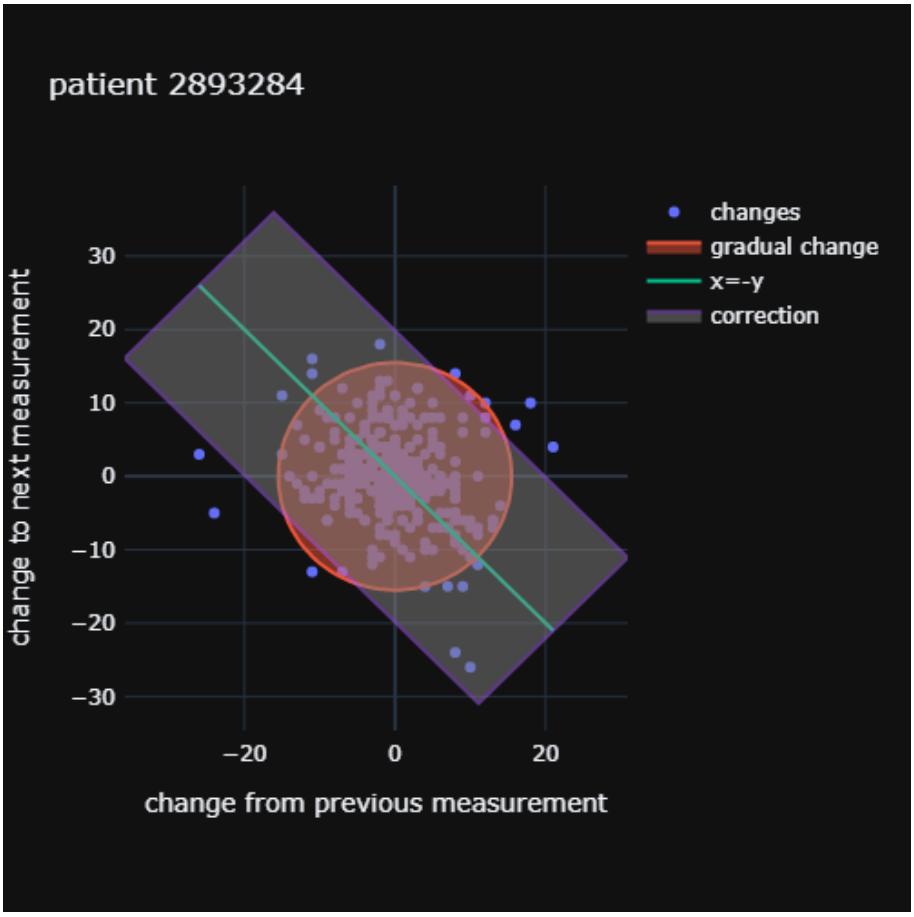

Figure 81: Changes from previous and next measurements per measurement for patient

This leads us to a straightforward outlier detection rule represented by:

$$\|y\|_2 > t_{\text{change}} \ \wedge \ \|y - P_{\text{Span}(1,-1)}y\|_2 < t_{\text{correction}}$$

where $y = (x_i - x_{i-1}, x_{i+1} - x_i)^T \in \mathbb{R}^2$, $P_{\text{Span}(1,-1)} \in \mathbb{R}^{2\times 2}$ denotes the projection matrix to $\text{Span}(1, -1)$ and $t_{\text{change}}, t_{\text{correction}}$ are the thresholds for non-gradual change and corrections in MAP values, respectively.

If needed in future processing, $t_{\text{correction}}$ could be define by a function on y instead as a constant.

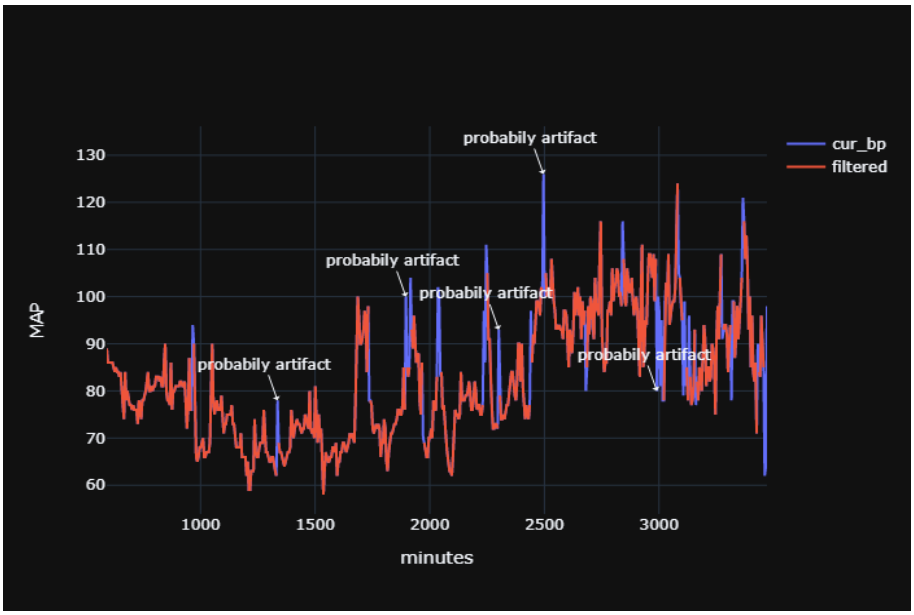

Figure 82: MAP for Time for Patient-Stay ID 2893284 with filtered points

5.2 Statistical Methods and Considerations

In the initial stages of our work, the approach was to exclusively employ non-parametric tests to minimize assumptions. However, as our data collection progressed, certain datasets exhibited distinct distributions. Consequently, we gradually introduced parametric tests, particularly those assuming normality, to our analytical toolkit.

5.2.1 Non Parametric Tests

5.2.1.1 Kolmogorov–Smirnov Test This work used the 2 way KS test, but first will introduce the one way test.

For a given sample $X = \{X_i\}_{i=1}^n$, where $X_i \overset{\text{i.i.d.}}{\sim} F$ are independent and identically distributed (i.i.d.) random variables following a continuous distribution function, we define the empirical distribution function as $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{x_i \leq x}$. To test our null hypothesis that $X_i \sim F$ we will use the statistic $D_n = \sup_x |F(x) - F_n(x)|$. This statistic is valuable for two key reasons: It is as commutable as F, and, more importantly, D_n exhibits the same distribution for every continuous F. Consequently, the only remaining step is to insert D_n into its distribution function to obtain the corresponding p-value.

Converting this test into a two-way test, where the null hypothesis is that the samples $X_i \overset{\text{i.i.d.}}{\sim} F, Y \overset{\text{i.i.d.}}{\sim} G$ hold $F = G$. The statistic in this case is $D_n = \sup_x |F_n(x) - G_n(x)|$, where F_n, G_n are defined the same way as in the last test. Again, the main feature of this statistic is that its distribution is the same for any continuous F, G.

It is important to note that in practical implementations, D_n and F_n is computed using some predefined binning of the the range of X.

5.2.1.2 Permutation test

While testing a null hypothesis that two independent samples, denoted as $X = \{X_i\}_{i=1}^n$ and $Y = \{Y_i\}_{i=1}^m$, are drawn from the same distribution, we can utilize the following property: if the distributions are the same, then the members of the 2 samples are interchangeable.

Suppose we have a statistic $T : \mathcal{X}^n \times \mathcal{Y}^m \to \mathbb{R}$ that measures the difference between the two distributions. This statistic can take various forms, such as the mean difference or the variance ratio, as long as it can yield different results when the distributions are not the same.

Under H_0, T probably wouldn't yield an outlier, as most of the space isn't made of them, and because $X \overset{d}{=} Y$ T is essentially a statistic from $\mathcal{X}^{n+m}$ to R, so a way to reproduce many instances of T is by permuting the members of $X \cup Y$ and assigning the to X', Y' and computing $T(X', Y')$. Though computing all permutations isn't feasible, even a small sample of this new space is sufficient.

Under H_1, our permutation-based scheme for generating T induces a new distribution, but for a fine engineered statistic, $T(X, Y)$ should not comply with that distribution, this enables to compute a p-value with the empiric distribution and reject H_0 if it is low.

5.2.1.3 Mann-Whitney U test

Given 2 samples X, Y as denoted before, in the one sided version the null hypothesis that $\mathbb{P}[X > Y] = \frac{1}{2}$ will be tested using the statistic computed by the next procedure:

1. For each X_i, Y_j compute $\tilde{X}_i = \sum_{j=1}^m \mathbf{1}(X_i > Y_j)$ and $\tilde{Y}_j = \sum_{i=1}^n \mathbf{1}(Y_j > X_i)$
2. Compute $U_x = \sum_{i=1}^n \tilde{X}_i,\ U_Y = \sum_{j=1}^m \tilde{Y}_j$, note that $U_x + U_y = n \cdot m$.
3. Compute $U = \min(U_x, U_y)$

The distribution of U can be computed and used to produce a p-value, and double the probability to obtain the p-value for a two sided test. In big samples, the normal approximation can be used.

5.2.2 Parametric Tests

In this work, we encountered several distributions that appeared to follow a normal pattern, and we leveraged this characteristic in various variance tests.

5.2.2.1 Shapiro–Wilk test Of Normality

Some of the results in this apeared to distibute normally, this test checks the validity of that notion.

Given a sample $\{X_i\}_{i=1}^n$, the null hypothesis that they are i.i.d. from the same normal distribution the next statixtic will be computed:

$$W = \frac{\left(\sum_{i=1}^n a_i X_{(i)}\right)^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}$$

Where $X_{(i)}$ are the ordered sample values, $\bar{X}$ is the sample mean and a_i are constants generated from the means, variances and covariances of the order statistics of a sample of size n from a normal distribution.

The distribution of W isn't known, and the p-value is computed using Monte-Carlo simulations.

5.2.2.2 Levene's Test for Equality of Variances

Consider a set of k i.i.d. sample sets $\{X^{(i)}\}_{i=1}^{k}$, the null hypothesis that all variances are equal for each distribution, $\sigma^2_{X^{(i)}} = \sigma^2_{X^{(j)}}$ for all (i,j), and the alternative hypothesis that for at leat one pair (i,j) it holds that $\sigma^2_{X^{(i)}} \neq \sigma^2_{X^{(j)}}$. The statistic that will be used is

$$W = \frac{N-k}{k-1} \cdot \frac{\sum_{i=1}^{k} N_i (\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (\bar{Z}_{ij} - \bar{Z}_{i\cdot})^2}$$

Where N is the sum of $N_i = |X^{(i)}|$, Z_{ij} is $|X_j^{(i)} - \bar{X}^{(i)}|$, and $\bar{Z}_{i\cdot}$ is the mean of Z_{ij} over j and similarly $\bar{Z}_{\cdot\cdot}$ is the mean of Z_{ij} over i and j.

Under H_0 $W \sim F_{k-1,N-k}$, and the p-value is calculated base on the F distribution, and reject if W has a high value.

5.2.2.3 Bartlett's Test for Equality of Variances

To test the same hypothesis as in the previous paragraph, Bartlett's test offers a stornger test, while being more sensitive to departures from normality.

The statistic used is

$$T = \frac{(N-k)\ln s_p^2 - \sum_{i=1}^{k}(N_i - 1)\ln s_i^2}{1 + \frac{1}{3(k-1)} \cdot \left(\sum_{i=1}^{k} \frac{1}{N_i - 1} - \frac{1}{N-k} \right)}$$

Where N, N_i are the size of the union of the samples and the size of the ith sample, respectively. s_i^2 is the non-bias estimator of the variance in the ith sample and s_p^2 is a weighted average of the estimators $\frac{1}{N-k}\sum_{i=1}^{k}(N_i - 1)s_i^2$.

$T \sim \chi^2_{k-1}$ under H_0, and uunder H_1 T will get high values. The p-value is computed according to the CDF of χ^2_{k-1}.

5.3 Models and ML Methods Used

5.3.1 Kalman Filter

By assuming that the changes in MAP are gradual, that the instrumentation noise is on a smaller scale than the actual MAP change, and additional assumptions about the trend of the MAP changes, we establish a solid foundation for the filter, which also possesses denoising capabilities.

The probabilistic model is described as follows: a measurement x_{i+1} is distributed as

$$x_{i+1} = A_{i+1}x_i + v_{i+1} + u_{i+1} \sim \mathcal{N}\left(Ax_i, \sigma_v^2 + \sigma_u^2\right)$$

Where $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ is the actual change in MAP, and $u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2)$ is the measurement noise. A_i is a linear operator that represent the trend of MAP. This formulation

models MAP measurements as a Markov process updated with changes and noise distributed normally based on the current state.

The estimation precedure has 2 main parts. The prdictiong stage, computing $\hat{x_{i+1}}$, the Minimal Mean Square Error (MMSE) estimator of $x_{i+1}|x_i$, with $\mathbb{E}[x_i|x_i]$. Then the updating stage, where the values of v_{i+1}, u_{i+1} are estimated. and seeing what is each of their contributes to x_{i+1} and eventually removing the noisy contribution of u_{i+1}.

Simplifying A_i to be equal to 1, indicating a constant trend, yielded favorable results. Attempts to dynamically determine A_i were found to be overly influenced by noise and artifacts and did not effectively address the issue of disregarding artifacts in general.

5.3.2 Gaussian Process

Similar to the Kalman filter, this model assumes that the measurement noise is independent and normally distributed. However, this model diverges from that assumption by positing that measurements at times t and s are not independent. Instead, they are jointly modeled as following a normal distribution, with their covariance determined by a kernel that relies on both s and t. For some of kernels, the distance between t and s, denoted as $|t - s|$, is sufficient for computation. In other words, the measurement in time t is highly dependent on the neighboring measurements.

A Gaussian process is denoted by $\mathbf{f} \sim \mathcal{GP}(m, k)$, where $m : \mathbb{R}^d \to \mathbb{R}$ and $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. It is common practice to use $m = 0$, and k needs to be a valid kernel. $\mathbf{f}$ is a mapping between $\mathbb{R}^d$ and random variables that hold $\mathbf{f}(x) \sim \mathcal{N}(m(x), k(x, x))$ and the covariance between $\mathbf{f}(x), \mathbf{f}(x')$ is $k(x, x')$. This effectively defines a Gaussian measure of an infinite-dimensional space over $\mathbb{R}$.

The most useful quality of GP is that for some $\mathbf{x} = \{x_i\}_{i=1}^n$, $\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), C)$ where $C_{ij} = k(x_i, x_j)$. This is well defined as k is a valid kernel, hence C is PSD. This framework provides the foundation for using the Minimum Mean Square Error (MMSE) estimator for $\mathbf{f}(\mathbf{x}')|\mathbf{f}(\mathbf{x})$ through basic probability tools and also accommodates the addition of independent noise.

In practice RBF, $x, x' \mapsto \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right)$ with σ as a hyperparmeter, gave the best results compared to other kernels.

Compared to the Kalman filter, this approach yielded better results. The smoothed trajectories retained the original information while demonstrating robustness against outliers. However, the primary drawback of this filtering method was its extended computation time and the reliance on many of assumptions.

5.4 Resources

5.4.0.1 Git Repository Code files and enlarged plots and graphs: https://github.com/Naftali-A/ds-final-23.git

5.5 Work Pipeline

5.5.1 Data Extraction

Our analysis commenced with the selective extraction of the specific data we intended to utilize from the databases.

5.5.2 Pre-processing

In the preprocessing phase, we focused on the extraction of relevant data. This process unfolded in several key steps:

5.5.2.1 Patient Selection Initially, we identified pertinent patients based on predefined criteria. This step ensured that our analysis focused on the appropriate data subsets.

5.5.2.2 Artifact Removal Subsequently, we conducted artifact removal within these selected patient entries. This meticulous step helped eliminate extraneous data that could distort our analysis.

5.5.2.3 Instrumental Noise Reduction The next critical task was the reduction of instrumental noise within the dataset. This procedure contributed to the refinement of our data for accurate analysis.

5.5.3 Diverging Paths for Analysis

Following the data preprocessing stage, our work pipeline bifurcated into two distinct directions:

5.5.3.1 Patient Stay Based Tests In one direction, we conducted tests primarily based on stay IDs. This approach involved two key phases:

a. Inter-Unit Analysis: Initially, we performed tests comparing different units within the EICU database. This allowed us to discern variations within the same institution.

b. Inter-Database Comparison: Subsequently, we extended our analysis to encompass a comparative evaluation between the EICU dataset in its entirety and the SICdb database. This broader perspective facilitated cross-database insights.

This work pipeline structure ensured a systematic and comprehensive approach to our analysis, enabling us to explore data variations at both the unit and inter-database levels.

5.5.3.2 Group Based Testing In the second direction we conducted tests based on groups of patients.

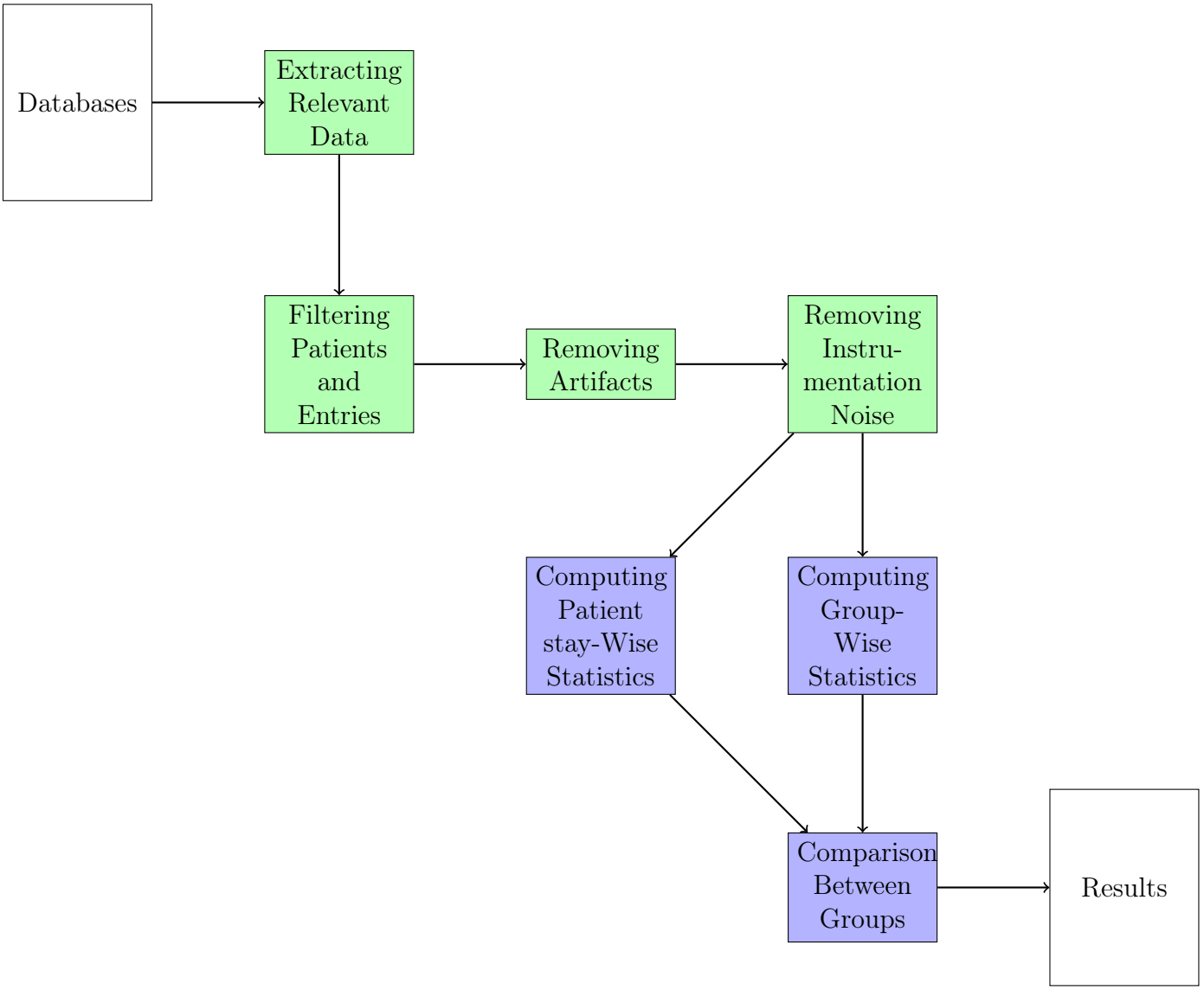

Figure 83: Work pipeline flow divided to Data, preprocessing and statistical analysis

Division of work

- Naftali Arnold: Group wise analysis, patient stay filtering.
- Eytan Slotnik: Stay id correlations analysis, denoising.
- David Orenstien: Saltzburg data extraction, high-level testing, denoising.

References

[1] Komorowski, M., Celi, L.A., Badawi, O. et al. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 24, 1716–1720 (2018).

https://doi.org/10.1038/s41591-018-0213-5

[2] Jeter, R., Josef, C., Shashikumar, S., Nemati, S. (2019). Does the "Artificial Intelligence Clinician" learn optimal treatment strategies for Sepsis in intensive care?. arXivpreprint arXiv:1902.03271.

[3] Peter Edelstein. (2016). Thought Leadership: Addressing the Greatest Threat to Healthcare. ELSEVIER. https://www.elsevier.com/__data/assets/pdf_file/0009/846864/Addressing-the-greatest-threat-to-healthcare_variability.PDF

[4] Beil, M., van Heerden, P.V., Sviri, S. (2022). Decision-Making Under Resource Constraints. In: Flaatten, H., Guidet, B., Vallet, H. (eds) The Very Old Critically Ill Patients. Lessons from the ICU. Springer, Cham. https://doi.org/10.1007/978-3-030-94133-8_16

[5] Evans, L., Rhodes, A., Alhazzani, W. et al. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 47, 1181–1247. https://doi.org/10.1007/s00134-021-06506-y

[6] De Backer, D., Cecconi, M., Chew, M.S. et al. A plea for personalization of the hemodynamic management of septic shock. Crit Care 26, 372 (2022). https://doi.org/10.1186/s13054-022-04255-y

[7] Kerlin MP, Epstein A, Kahn JM, Iwashyna TJ, Asch DA, Harhay MO, Ratcliffe SJ, Halpern SD. Physician-Level Variation in Outcomes of Mechanically Ventilated Patients. Ann Am Thorac Soc. 2018 Mar;15(3):371-379. doi: 10.1513/AnnalsATS.201711-867OC. PMID: 29283699; PMCID: PMC5880524. https://pubmed.ncbi.nlm.nih.gov/29283699/

[8] Nordenskjöld Syrous, A, Malmgren, J, Odenstedt Hergès, H, et al. (2021). Reasons for physician-related variability in end-of-life decision-making in intensive care. Acta Anaesthesiol Scand. 2021; 65: 1102–1108. https://doi.org/10.1111/aas.13842

[9] Pollard, T., Johnson, A., Raffa, J. et al. (2018). The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 5, 180178 (2018). https://doi.org/10.1038/sdata.2018.178

[10] Masnadi-Shirazi MA, Behbehani K, Zhang R. (2004). Kalman filter modeling of cerebral blood flow autoregulation. Conf Proc IEEE Eng Med Biol Soc. 2004;2006:734-7. doi: 10.1109/IEMBS.2004.1403263. PMID: 17271782.

[11] Ujjwal Manikya Nath, Chanchal Dey, Rajani K. Mudi. (2020).Designing of dynamic Kalman filter for prediction of mean arterial blood pressure, Procedia Computer Science, Volume 167, 2020, Pages 2478-2485, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.300.

[12] Stephanie Hunter, Julie Considine, Elizabeth Manias, Nurse decision-making when managing noradrenaline in the intensive care unit: A naturalistic observational study, Intensive and Critical Care Nursing, Volume 77, 2023, 103429, ISSN 0964-3397, https://doi.org/10.1016/j.iccn.2023.103429