2022

A Novel Nonparametric Test for Heterogeneity Detection and Assessment of Fluid Removal Among CRRT Patients in ICU

Shaowli Kabir
University of Kentucky, shaowlikabir@gmail.com
Author ORCID Identifier:
![ORCID icon] https://orcid.org/0000-0002-5839-2460
Digital Object Identifier: https://doi.org/10.13023/etd.2022.051

Right click to open a feedback form in a new tab to let us know how this document benefits you.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

<div align="right">

Shaowli Kabir, Student

Dr. Richard J. Charnigo, Major Professor

Dr. Heather Bush, Director of Graduate Studies

</div>

A NOVEL NONPARAMETRIC TEST FOR HETEROGENEITY DETECTION
AND ASSESSMENT OF FLUID REMOVAL AMONG CRRT PATIENTS IN ICU

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Public Health at
the University of Kentucky

By
Shaowli Kabir
Lexington, Kentucky

Director: Dr. Richard J. Charnigo, Professor of Biostatistics
Lexington, Kentucky 2022

ABSTRACT OF DISSERTATION

A NOVEL NONPARAMETRIC TEST FOR HETEROGENEITY DETECTION

AND ASSESSMENT OF FLUID REMOVAL AMONG CRRT PATIENTS IN ICU

Over the past decade acute kidney injury (AKI) has been occurring among 20%-50% of patients admitted to the intensive care unit (ICU) in United States. Continuous renal replacement therapy (CRRT) has become a popular treatment method among these critically ill patients. But there are multiple complications in implementing this treatment, including discrepancies in practiced and prescribed fluid removal, possibly related to the heterogeneity among these patients. With mixture modeling there have been several techniques in detecting heterogeneity with their specific limitations. In this dissertation a novel nonparametric 'd test' will be used to detect heterogeneity among CRRT patients in ICU. Along with heterogeneity detection, this dissertation will also seek to understand ongoing issues with fluid removal and discrepancy in treatment implementations.

KEYWORDS: CRRT, Heterogeneity, Discrepancy, Biostatistics, Systematic-Review.

Shaowli Kabir

April 25, 2022

A NOVEL NONPARAMETRIC TEST FOR HETEROGENEITY DETECTION
AND ASSESSMENT OF FLUID REMOVAL AMONG CRRT PATIENTS IN ICU

By

Shaowli Kabir

_____
Dr. Richard J. Charnigo

Director of Dissertation

_____
Dr. Heather Bush

Director of Graduate Studies

_____
April 25, 2022

Date

To my parents Maksuda and Rezaul, my sister Nadia, my husband Mobasshir and our beloved Mohi and Chotu.

ACKNOWLEDGMENTS

I am thankful to the Almighty for all the experiences I have had through this journey.

Thank you to my family, who have made countless sacrifices that have provided me with this life. Ammu and Abbu I know how hard it was for you to let me go and study abroad, but you have been strong and always kept me going. To my sister for being there for me every time I needed her and for taking care of our parents when I couldn't. To my Mohi, having you was one of the best things in my life. Thank you for bringing so much joy in our life. To my friends in Lexington, I have enjoyed every bit of time spent with you. Your support, mentorship, guidance, and tolerating my antics boosted me through this journey.

Thank you to all my teachers and mentors who provided valuable guidance and support. To Dr. Richard Charnigo, thank you for being my mentor and advisor. Without you, I would not have imagined surviving this journey. Dr. Erin Abner, Dr. Derek Young and Dr. Mary Lacy-Leigh, this dissertation would not be possible without your guidance and support. Dr. Javier Neyra without your clinical judgement and support for my research, this would not have been possible. I'm thankful for all the years that I've got to work with you. I thank all of my committee members for their support and feedback. To Dr. Heather Bush and Dr. Arnold Stromberg, I cannot thank you enough for helping me decide to do a PhD and for all of the support you have provided from time to time. To my colleagues: thank you for always helping me in need and listening to my complaints. Thank you, Lauren E. Robinson, for helping me within the short amount of time for the literature review. This dissertation would not be possible without you all.

To my husband, Nayan, thank you for being the consistent support, guidance, and for your patience. You have always motivated me and pushed me to keep going through the hard times. Thank you for all the sacrifices you have made so that we can be together. You have always been there for me at every step of this journey, and I would never have imagined being here without your support. I love you so much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Mixture Model

A Mixture model is a probabilistic model that represents the distributions of the sub-populations in an overall population [1]. A distribution f is a mixture of K component distributions $f_1, f_2, ..., f_K$ if,

$$f(x) = \sum_{i=1}^{K} p_i f_i(x)$$

where, p_i are the mixing weights with $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$. A finite Gaussian or normal mixture model is where the ith components are $N(\mu_i, \sigma_i)$. That is,

$$f(x) = \sum_{i=1}^{K} p_i N(x|\mu_i, \sigma_i) \tag{1.1}$$

1.1.1 Application and Example

Mixture models are famous for heterogeneity detection. Normal mixture models have been used in gene filtration in microarray analysis, neuroimaging pattern recognition, disease mapping and numerous other issues in public health [2–7].

For example, birthweight is one of the prognostic factors for fetal-infant mortality. With low and extremely low weights at birth, the mortality risk increases and thus need special attention to these groups. Charnigo et. al. [8] applied a normal mixture model to birthweight and fetal-infant mortality data to build a realistic and flexible framework for birthweight age distribution. The number of components in the mixture model is determined from data using Flexible Information Criterion(FLIC) [9]. The authors used the EM algorithm [10] to apply FLIC by using maximum likelihood

estimate of proportions, mean and standard deviation. They also investigated the relationship between sample size and the number of components selection, where 4 components were chosen for their 50,000 random samples from 202,849 of white singletons. With the comparison of the 4 components model to the contaminated normal model and 2-components model, the 4 components model averted the weaknesses of the other two models [8]. The authors also compared how each component fits the tails of the distributions to the clinical groups of birthweight distribution. That is, they checked how the components model fits the MLBW (medium-low birthweight) range. By comparing the criteria preferences, the authors concluded that, a multiple components (> 2) normal mixture model detects the heterogeneity and reasonably describes the birthweight distribution when compared to a contaminated normal or a 2-components normal mixture model.

1.2   Implementation of Mixture Model

Numerous methods for estimating the parameters of a mixture model (p_i, θ_i), where $\theta \epsilon \Theta$(parameter space), have been proposed. Some suffer from theoretical complications, some suffer from computational challenges. Moreover, the log-likelihood function of a mixture distribution does not have a closed form solution and regarded as incomplete without prior knowledge [10, 11]. Below are some of the popular methods used in implementing the mixture model.

1.2.1   EM algorithm

The EM (Expectation-Maximization) algorithm is a popular approach for finding maximum likelihood estimates with missing or incomplete data. The EM algorithm approaches to find the value of the parameter using sampling density depending on the parameter and the data, which maximizes the sampling density given the observed response. With the popularity of the maximum likelihood method [12], Dempster et

al. (1977) [10] explored the EM algorithm's properties in detail. They showed that, along with the maximum likelihood framework, the EM algorithm can be applied in a Bayesian setting. They also investigated different approaches of the EM algorithm in different scenarios (different types of exponential families and non-exponential families), which can be simplified as follows-

E-step (Expectation): After choosing appropriate initial values, find the appropriate expected value (e.g., likelihood, quasi-likelihood, etc.) given the current estimate of the parameter.

M-step (Maximization): Find the parameter estimate that maximizes the above-mentioned expected value.

Convergence (with the increase of iterations) to maximum likelihood is ensured with the assumption of continuity and differentiability conditions. Also, differentiation and expectation operations are assumed to be interchangeable.

Redner & Walker (1984) [12] investigated the theoretical and iterative properties of the EM algorithm for mixture densities. They used exponential families as the point of interest due to the majority of literature involving mixture densities, including component densities, being members of exponential families. The performance of the EM algorithm was satisfactory as the authors mentioned that process has "good global convergence characteristics" (p-231).

Although the EM algorithm is thought to obtain global maxima with carefully chosen initial values, this might not be the case always. Some initial values might lead the EM algorithm to get stuck at local maxima. Thus, careful consideration should be given in choosing initial values. The EM algorithm can be used initially, and later, Newton-Raphson's iterative approach can be used for rapid local convergence. [12]

1.2.2   Likelihood Ratio Test

G. J. McLachlan (1987) [13] proposed using LRT (Likelihood Ratio Test) to find the smallest number of components in a normal mixture model when there is no prior knowledge. He used bootstrapping of the log-likelihood ratio statistic and performed simulations to test the simplest situation- one component vs two components. Because there is no way to be sure that the largest of the local maxima will be found in the log-likelihood for mixtures, the likelihood ratio statistic ($-2loglambda$) may be biased downard, where λ is the likelihood ratio under the null and alternative hypotheses. The EM algorithm [10] was used to limit this bias through a systematic search for all local maxima using multiple sets of initial values. Also, to avoid the situation of having two local maxima because of the components belonging to the same parametric family, it was conditioned that the mean of the first component should be less than the mean of the second component. The simulation results showed that, if the distances between the components were larger then the test would be more powerful. The author also tested 2 components vs 3 components with the condition of a consecutive increase in the means of the components and found similar results supported by bootstrap simulations.

Hanfeng Chen and Jiahua Chen (2001) [11] studied the large sample behavior of LRT for testing homogeneity under a two-components mixture model. To have a limiting distribution for LRT, the parameter space needs to be bounded, otherwise, the LRT goes to infinity with probability 1 [14]. After describing the regularity conditions, the authors stated that the asymptotic null distribution of the LRT for the two-component mixture model is the squared supremum of a truncated gaussian process. They mentioned that this asymptotic distribution is complex because the null asymptotic distribution depends on how large the parameter space is, thus the bounded condition was implied. They also did bootstrap simulation [McLachlan, G.

1987] for testing homogeneity with different kernel functions (Normal, Binomial, and Poisson). The result concluded that at a 5% level of significance, the approximation improves with the increase of sample size, and the simulation study agreed with the consistent performance of LRT.

Since the likelihood ratio test (LRT) for homogeneity in finite mixture models has complicated asymptotic properties [11] Chen et al. [15] proposed a modified likelihood ratio test (MLRT) with similar power and easier asymptotic properties. They showed that the asymptotic null distribution is a weighted mixture of central χ_1^2 and χ_0^2 distribution. χ_0^2 is the degenerate distribution with all of it's values are at 0. The MLRT has a penalty term that affects the maximized likelihood only under the alternative hypothesis. The estimated weights of the mixture components in the heterogeneous model are forced to be away from 0 by the penalty term, which helps estimate the parameters of the different components.

1.2.3 D Test

Richard Charnigo and Jiayang Sun (2004) [16] presented a new method, the D test, for testing the homogeneity of the finite mixture distribution. The D-test statistic has a closed-form expression for mixture components from standard parametric families in terms of parameter estimators, whereas likelihood ratio type test statistics do not. The D test employs an "L^2" distance between a fitted homogeneous model and a fitted heterogeneous model, based on Scott's (1998) [17] L_2E method for model selection. The reason for choosing L^2 distance is that it puts more emphasis on the larger separations between the homogeneous and heterogeneous density curves and also leads to a simple closed-form expression so that the test is sensitive to the separation between the null and the specific alternative under consideration.

The D test statistic is defined as,

$$d(k, n) := \int \big[\sum_{i=1}^{k} \hat{p}_i f(x|\hat{\theta}_i) - f(x|\hat{\theta}_0) \big]^2 dx \qquad (1.2)$$

where k is the number of distinct components under the alternative hypothesis and n is the sample size. The authors found that the D-test has more power to detect homogeneity compared to the MLRT in simulation study when the mixture components come from a normal location family, but the generalization of MLRT to a two-parameter family tends to perform better than D test when mixture components come from a normal location/scale family. Because L^2 distance increases the smallest distance between shapes, it is easy to detect subtle changes when there is heterogeneity. Moreover, with a small sample size, it becomes difficult to detect heterogeneity in exponential mixture densities. To overcome this weakness, the authors also proposed a weighted D test [16] where the L^2 distance is changed to accentuate the disparities. They also found another equivalent way by changing the measure or transforming the data before conducting the D test. To check the D test's feasibility, the authors applied it to the bankruptcy data from Johnson and Wichern (2002) [18] and found satisfactory results.

Later, the authors [19] investigated the asymptotic equivalences between the D-test and three other likelihood ratio type tests (LR test, MLRT [Chen, Chen and Kalbfleisch 2001] [15] and EM test [Li, Chen and Marriott 2009] [20]). With the same mixture from a regular exponential family using Chen, Chen and Kalbfleich's (2001) [15] penalized likelihood bayesian framework, the authors showed an asymptotic critical value related to the upper 2α quantile of the chi-square distribution on one degree of freedom. Thus, under contiguous local alternatives, the power should stabilize to some larger amount close to 100% and this simple limiting null distribution makes the D test asymptotically locally most powerful. To investigate finite-sample

accuracy of critical values, the authors used MLRT as a benchmark for simulation studies on the mixture from the one-dimensional exponential family and the EM test as a benchmark for normal location mixtures with unknown structural parameters, i.e. variance in this case. Compared to LRT and MLRT, which need both full dataset and parameter estimates, the D test depends on the data only through the applicable framework of parameter estimation.

1.3 D-Test in Application

1.3.1 Real Data Heterogeneity Detection

Charnigo and Sun [19] applied the empirical Bayesian framework D-test to the SLC dataset (Roeder 1994) [21] and found that both EM test and empirical Bayesian framework D test suggest heterogeneity in the form of what the authors called "a high-low probability mixture", where the first component gets much higher probability and second components get much lower probability.

1.3.2 Contamination Detection

In a mixture model when one component's parameters are known, then it is defined as a contaminated density model [22]. In a regression model to test for contamination to describe a subpopulation, Dai and Charnigo explored the asymptotic and finite sample performance of the MLRT (Modified Likelihood Ratio Test) and the D test. The authors are interested in testing zero contamination vs. non-zero contamination and have developed easily applicable inferential methods for contaminated density and regression models. For the contaminated density model, the authors considered using a weighted D-test, and for the contaminated regression model, when the probability density function for the explanatory variable is unknown, the authors considered using an empirical D test. Their empirical assessment showed that the tests distinguish contamination easily when the contamination fraction is large, even if the component

means are not separated enough. Power can be retained with a small contamination fraction if the sample size is large or the component means differ greatly. For the contaminated density model, the authors concluded that using an appropriate weight function improves the power to detect contamination. The authors pointed out that MLRT and D tests can be used to check whether apparent outliers are a part of subpopulation, which is a paramount aspect of the data analysis.

1.3.3 Gene Filtration:

To detect gene expression alterations, Dai and Charnigo (2008) [2] used the Modified Likelihood Ratio Test (MLRT) and the D test to adjust for the large number of simultaneous tests corresponding to the number of genes. Because, when we try to control the false positive rate in large scale multiple testing, the false-negative rates become extremely large. Thus, in large-scale testing, traditional methods have low power to detect differential gene expression, failing to compare groups efficiently.

From previous work of Allison et al.(2002) [23], in large scale hypothesis testing, P-values for genes without expression alterations are distributed as iid $U(0,1) = \beta(1,1)$ and differentially expressed genes' P-values are assumed to be distributed as a Beta distribution with different parameter values. Thus, Dai and Charnigo proposed a beta contamination model to define the marginal distribution of the P-values obtained from large scale gene studies. Because of the abundance of less important genes in such studies, the authors suggested omnibus homogeneity testing as part of a gene filtration process which separates the genes by their differential expression rates using the following hypotheses:

$$H_0 : P_i \sim^{iid} Beta(\alpha_0, \beta_0)$$

$$H_1 : P_i \sim^{iid} (1 - \pi)Beta(\alpha_0, \beta_0) + \pi Beta(\alpha, \beta) \neq Beta(\alpha_0, \beta_0)$$

where α_0 and β_0 are known and $\pi \in [0,1], \alpha > 0, \beta > 0$.

To test for goodness of fit for the null model, i.e. Beta(1,1) model MLRT and D test were used. A maximum modified likelihood estimate (MMLE) was used to test the hypothesis of whether a uniform model was appropriate or not. To test the performance of the D test and MLRT, two empirical studies were conducted where one was concerned with the actual rejection rates under the null hypothesis that the distribution is from a beta(1, 1) model and the other was concentrated on the power under the alternative hypothesis when critical points from asymptotic theory were used. The simulation study confirmed that the D test and MLRT have advantages in both scenarios. The authors assumed independence of P-values and concluded that contaminating beta distribution helps to detect substantial differential expressions under certain assumptions of the parameters.

The contaminated beta (CB) model is useful for describing the distribution of P-values resulting from a microarray experiment. Dai and Charnigo (2010) proposed using the contaminated normal (CN) model instead of the CB model to describe the distribution using Z-statistics instead of P-values. Balancing type I and type II error rates with a large number of hypothesis tests is challenging. The authors suggested that if unnecessary genes can be filtered out through omnibus testing, with fewer genes under consideration, greater power can be achieved in hypothesis testing while maintaining the type I error rate. Thus, they investigated the asymptotic behavior of MLRT (Modified Likelihood Ratio Test) and D test for omnibus testing using the following hypothesis tests:

$$H_0 : \gamma\mu = 0$$

$$vs H_1 : \gamma\mu \neq 0$$

where γ is the proportion of genes that are differentially expressed and μ is the mean Z-statistic of genes that are differentially expressed. Maximum modified likelihood reduces the non-identifiability problem by pushing γ away from 0 using a penalty term

(Chen, Chen and Kalbfleisch 2001) [15]. Dai & Charnigo also used Kolmogorov-Smirnov for simulation study to compare CN and CB [24]. The simulation result showed CN to be more powerful in detecting differential expression than the CB model for some specific cases: when either overexpression or underexpression of genes is more prevalent than the other (that is, asymmetric), when the ratio of $|\mu|$ and σ is moderate, and when two-sided tests are preferred to one-sided tests. That is when there is no assumption on the sign of μ given that $\mu > 0$ corresponds to overexpression and $\mu < 0$ corresponds to underexpression, CN is preferable. But if there is prior knowledge about the direction of μ then a right-sided test for detecting overexpression and a left-sided test for detecting underexpression is preferred. In these cases, CB is superior to CN. CN and CB can also be used to estimate the fraction of differentially expressed genes.

To understand real-life applicability, the authors used real microarray data of 10 SARS patients and 4 healthy controls from a study on expression levels of immune response genes [25]. The data is also available at the Gene Expression Omnibus of the National Center for Biotechnology Information. Charnigo and Dai compared the performances of CN and CB. They found that $\hat{\gamma}$ is much larger in CB than in CN because, unlike CN, CB does not assume differential expression to lie in mainly one direction. They compared the performances of CB and CN by computing maximized modified log-likelihoods and concluded that the CB model performs better for this data set. One advantage in comparing CB and CN is that these two models have the same number of parameters and thus, can be compared based on a BIC-type criterion. The authors pondered the limitations of CN in the case of a symmetrical distribution of Z-statistics and concluded that it can be overcome by using extended two contaminated components in the CN model.

1.4 Nonparametric Approach in Mixture Model

Nonparametric methods are becoming quite popular now in identifying clusters. With all the parametric, semi-parametric approaches discussed in the previous sections, we explored some ideas in the field of nonparametric clustering and classification techniques (where one is an unsupervised learning method and the other is a supervised learning method) for detecting heterogeneity without applying mixture modeling. We will discuss both

Hastie et. al. described a kernel or weight function technique that provides flexibility in estimating a regression function with multiple inputs in chapter 6 "Kernel Smoothing Method" [26]. The technique fits different simple models at each target point separately, along with assigning weights to chosen points based on their distance from the corresponding target points. This method is dependent on training data only and requires only determining the neighborhood distance indexed by λ. For example, Epanechnikov quadratic kernel is given as follows:

$$K_\lambda(x_0, x) = d(\frac{|x - x_0|}{\lambda})$$

with $d(t) = \frac{3}{4}(1 - t^2)$ if $|t| < 1$ else, 0 otherwise.

For one-dimensional kernel smoothers, the authors suggested local linear regression and local polynomial regression. Local linear regression is a nonparametric approach in which the f(x) in a regular fitted regression model (y = f(x) + ϵ) is nonlinear but behaves linearly when divided into small regions. For local linear regression, locally weighted averaging has a bias problem at or near the boundaries of the domain, but with less variance. On the other hand, local polynomial regression reduces bias at boundaries but has increased variance. Also, there is a bias-variance trade-off with the change of the width of the averaging window. This is very important as the

choice of width, either as fixed or variable, has a big impact on kernel smoothing. The authors suggested that if extrapolation is of interest then, local linear fits are probably more trustworthy than local polynomial fits of higher order.

This local regression technique comes with its own characteristic limitations. When the dimension is more than 2 or 3, the difficulty in visualization and the choice of kernel make local regression less useful. The authors suggested that when the dimension to sample size ratio is very large, structural assumptions about the kernel or the regression function make local regression helpful. For structured kernel, when we have multiple inputs, we have the option of choosing multiple bandwidths (we could also choose one tuning parameter instead of two, in this case we could standardize each input variable to unit standard deviation). For structured regression function, we can use one-dimensional local regression to estimate analysis of variance (ANOVA) decompositions.

With the broad concept of local regression, the authors suggest that if a parametric model is fitted by accommodating observation weights, then it will be called local. The local likelihood estimation technique is a smoothing technique based on local polynomials in non-gaussian regression models. This is useful because, the local likelihood does not restrict the data analyst to a parametric model, which is a globally linear or generalized linear model, which may be unreasonable versus a locally linear or generalized linear model. The authors also suggested avoiding the dimensionality problem by assuming an additive structure of the regression function, as generalized additive models using kernel smoothing methods are closely related to multiclass linear logistic regression models.

The authors also talked about kernel density estimation, which is an unsupervised learning procedure with an application for nonparametric classification. They discussed how a discontinuous histogram estimate can be smoothed by using the Parzen

estimate from Loader (1999) [27]. They suggested using nonparametric estimated densities for classification, using Bayes' theorem. Also, when classification is the ultimate goal, we only need to estimate close to the decision boundary. They have discussed a naïve Bayes classifier where a naïve Bayes model assumes features are independent given a class. It becomes easy to estimate class-conditional marginal densities. Even though the estimated marginal class densities are biased, it doesn't affect the posterior probabilities near the decision boundaries. When we are not sure about integrating the estimated density function to 1, the radial basis function uses a kernel type argument for localization. A basis function by itself is not flexible enough to show local behavior. That is, we can use the basis function without positivity constraints.

The authors discussed how all this information can be used to create a mixture model for density estimation and classification. Even though the Gaussian mixture model is very popular, the authors popularized the concept of the mixture model: "mixture models can use any component densities in place of the Gaussian" (page-214). They also mentioned two special cases when the covariance matrices are restricted to a scalar multiple of the diagonal than the gaussian mixture model is related to a radial basis function. Also, when component variances are equal and the number of components increases with the sample size, the maximum likelihood estimates for the Gaussian mixture model approach Kernel density estimate.

1.5   Future Direction with Heterogeneity Detection

In this chapter, we have explored the wide applicability of mixture modeling and its limitations in implementations and nonparametric approaches. Keeping in mind the limitations of both parametric and nonparametric approaches, we aim to find a feasible test for heterogeneity detection that improves the limitations of parametric tests.

As discussed in the previous section, when the number of mixture components is large with a large sample size, the maximum likelihood estimates approach the kernel density estimate. This encouraged us to use kernel density as a alternative fit for detecting heterogeneity. Thus, we would explore how a nonparamteric density estimate can be used to obtain a heterogeneity detection test without worrying about the actual number of components that exist.

1.6 Acute Kidney Injury and Continuous Renal Replacement Therapy

Over the past decade, one of the most frequent complications among patients who are admitted to the intensive care unit (ICU) acute kidney disease (AKI) [28–30]. With injured kidney the human body cannot filter out enough waste and toxic substances and as a result, the body starts to accumulate excessive fluids [31]. Continuous renal replacement therapy (CRRT) has become popular among critically ill patients as an efficient treatment for removing excess fluid from the body [32, 33].

Since the health condition among ICU patients are not always stable, the application of CRRT is always challenging, from finding an optimal fluid removal rate to implementing the treatment [34–36]. With several complications in the ICU, it may be beneficial for clinicians to identify clusters of patients to provide specific treatments and obtain optimal results. Also, identifying other problems and consequences, will be beneficial to finding better healthcare solutions.

1.6.1 Fluid overload

Fluid overload (FO) is the medical term for the condition when the blood contains a higher liquid portion (plasma) [37, 38]. A healthy body contains a certain amount of fluid based on age and weight, because their kidney removes excess fluid from the body. But with a dysfunctional kidney body can not remove excess fluid and causes FO. It can also be caused by heart or liver failure, hormonal imbalance, or

14

excessive intravenous (IV) fluid transfer [38]. The majority of the cases are associated with kidney disease [39]. Most of the AKI patients in the ICU suffer from end stage kidney failure or injured kidney. Thus, FO is a common problem among people with acute [40, 41] or chronic kidney disease [42].

To quantify FO, Goldstein et. al. [43] proposed the following formula adjusted for the body weight,

$$FO = (\text{fluid intake} - \text{fluid output})/\text{admission weight} \times 100\%$$

1.6.2 NUF rate and Mortality

The net ultrafiltration (NUF) is used to measure the amount of fluid needed to be removed to attain fluid balance (FB). It is defined as the volume of fluid removed per hour adjusted for patients body weight. NUF rate or intensity is measured [44] as follows:

$$\text{NUF rate (ml/kg/h)} = \frac{\text{Total NUF volume (ml)}}{\text{weight(kg)} \times \text{treatment duration (h)}}$$

In general NUF has been categorized in three groups.

- Low: <1.01 ml/kg/h

- Moderate: 1.01 to 1.75 mL/kg/h

- High: >1.75 ml/kg/h

It has been more than 70 years since NUF has been used in controlling FO [45], but the optimum rate has still not been determined. Some studies [46, 47] suggested that less intensive (slower rate or smaller volume) NUF may be associated with tissue and organ swelling and increased morbidity and mortality, whereas some studies

15

suggested [48,49] faster rate or larger volume causes hemodynamic and cardiovascular stress, which also leads to increased morbidity and mortality.

Given the inability of patients to tolerate therapy and contradicting findings in multiple articles on the association between greater NUF rates and mortality, it appears that there may be a discrepancy in recommended and implemented NUF rates. Murugan, Ostermann et. al. conducted a multinational internet-assisted survey in 80 countries to understand the attitudes and practices of practitioners with respect to NUF prescription [50].

They have pointed out major issues that are causing the discrepancy summarized below.

- Patient intolerance (hypotension, diabetes, health condition etc.)

- Frequent interruption (visit for lab tests, machine malfunction etc.)

- Undertrained Nursing staff.

- Unavailability of the machines and cost associated with treatment.

There are various challenges in assessing human error in the implementation of CRRT. However, we may investigate if the treatment's execution is also influenced by other, controllable phenomena. It's probable that the patient's intolerance is the greatest obstacle to achieving the prescribed fluid removal. Identification of some other modifiable factors that play even a small role in creating a discrepancy in treatment implementation, is also crucial in improving patient healthcare. We propose to identify groups of patients by applying heterogeneity detection techniques and how they can be utilized to understand the treatment discrepancy.

More investigation is required to understand the disparities in previous literature findings. We proposed a systematic literature review to explore the key findings and

identify reasons for variation in results. We will also explore fluid removal issues among CRRT patients to identify key patient characteristics that lead to increased mortality risk, as well as how the CRRT discrepancy relates to other ICU parameters.

Chapter 2 Nonparametric D test

2.1 Introduction

Mixture modeling is a scientific tool for clustering observed data based on some unobserved characteristics. From public health to astronomy, finite mixture models have been widely used to detect heterogeneity [51–53]. Image segmentation [54], disease mapping [55], genetics trait mapping [56, 57] and so many other fields of study have been applying mixture modeling. With the versatile applicability, mixture models are also used in large scale hypothesis testing to adjust for increased numbers of false positives. In large scale hypothesis testing each test results becomes a data point for mixture modeling to cluster: those of scientific interest vs those not [58]. When increased number of false positives makes the multiplicity adjustments difficult to apply, mixture models can discard numerous results through a so called omnibus test [2].

Likelihood ratio test (LRT) [11] [59] has been initially proposed to test for homogeneity in mixture models. Considering complicated asymptotic properties and critical value bootstrap issues, several other methods have been proposed for heterogeneity detection [13,15]. For example, Charnigo and Sun used a L^2 distance based technique called d-test in mixture distribution [16]. They compared the null hypothesis to an alternative hypothesis with specified number of components. But this comes with the challenge of being confident on the number of components that need to be tested.

With the uncertainty of the number of components in alternative hypothesis, multiple testing and adjusting for type I error in sequential testing becomes difficult and time consuming. Also, d test as originally proposed may not have the flexibility to detect variety of departures from homogeneity. Thus we propose a nonparametric version

of the d test which accounts for the various possible departures from homogeneity based on Kernel smoothing and empirical null hypothesis [58]. In particular our test is intended to capture departures from homogeneity that can occur with model misspecification. That is, our robust test will help to reject the incorrect null hypothesis and detect heterogeneity not only without specifying the number of components but also without requiring a finite number of components from same parametric family.

The nonparametric d test (NpD test) compares a parametric distribution to a nonparametric distribution,

$$H_0 : X \sim f(x|\theta_0) \quad vs. \quad H_1 : X \sim f_h(x)$$

Where, θ_0 is a scaler or vector belonging to a parameter space, Θ and $f(x|\theta_0)$ is corresponding probability density function. Also, $f_h(x)$ is a kernel density function with h being the bandwidth.

To visualize this method, we simulated data from two component contaminated distribution $0.8N(0,1) + 0.2T(4)$ with different mixing proportions showed in figure 2.1. The original d test computes the L^2 distance between the upper two red(one component fit) and blue(two-components fit) curves, where the two-component fit is not representing the data well. The NpD test computes the L^2 distance between the bottom two red(Empirical fit) and blue(Nonparametric fit) curves fitted on the data [60]. The empirical fit emphasized on the 80% data coming from one single component (N(0,1)) which helps in understanding the data better, and the nonparametric fit represents a smooth fit of the data. Thus, it is evident from our method that the distance between empirical and kernel fit is more prominent compared to the distance computed in regular d test. Thus, we get better chance at detecting departures from homogeneity when data is coming from a contaminated distribution with huge difference in mixing proportion, and do not need to worry about misspecification of the

hypotheses.
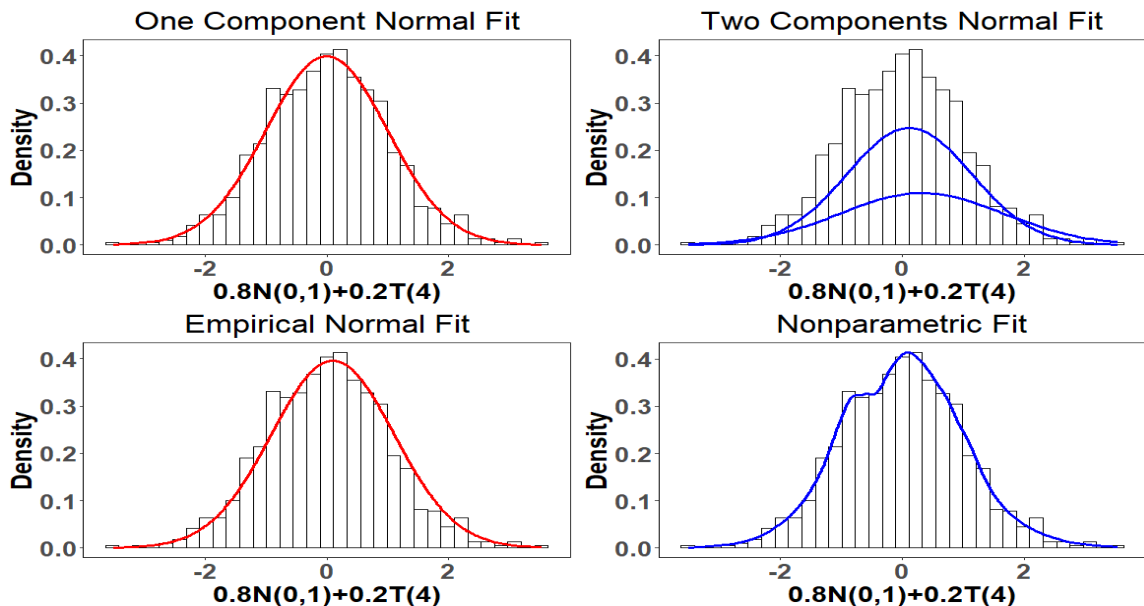
Figure 2.1: Nonparametric d test procedure

The L^2 distance puts more emphasis on the larger separations between the homogeneous and heterogeneous density curve. It also leads to a simple closed form expression so that, the test is sensitive to the separation between the null and specific alternative under consideration. The bigger the distance the more evidence of data coming from a mixture distribution. For cases where the component mixtures are equal and the distance between the components are small, we proposed weighted NpD test in section 2.3. The weight function emphasizes the discrepancies and helps the L^2 distance to detect heterogeneity.

2.2 Nonparametric D test (NpD test)

2.2.1 Definition of non-parametric d test

NpD test measures the L^2 distance between fitted Gaussian kernel density and fitted empirical null density [58]. The empirical null distribution is estimated by obtaining the center and half-width of the central peak of the standardized data distribution.

That is, empirical null uses the data to define homogeneity. It protects a user if there is homogeneity but not the one in expecting (misspecification of the null). Also, Efron established that, empirical null is better at identifying smaller percentages of interesting cases compared to theoretical null [58]. Thus we are testing,

$$H_0 : Z \sim f(z|\mu_0, \sigma_0^2) \quad vs. \quad H_1 : Z \sim f(z)$$

Where Z is the standardized score of the data. μ_0 is defined as the center of the z score distribution and σ_0 is defined as the half width of the central peak or curvature of the center.

$$\mu_0 = argmax(f(z)); \quad \sigma_0 = \left[-\frac{d^2}{dz^2} \log f(z)\right]^{-\frac{1}{2}}$$

f(z) is defined as $\frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{z-z_i}{h})}$ with, n as the sample size, h as the bandwidth.

The NpD statistic is defined as follows[1]:

$$d = \int \left[\hat{f}(z) - f(z|\hat{\mu}_0, \hat{\sigma}_0)\right]^2 dz$$

$$= \int \left[\frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{Z-Z_i}{h})^2} - \frac{1}{\sqrt{2\pi}\hat{\sigma}_0^2} e^{-\frac{1}{2\hat{\sigma}_0^2}(Z-\hat{\mu}_0)^2}\right]^2 dz$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2n^2 h\sqrt{\pi}} e^{-\frac{1}{4h^2}(Z_i-Z_j)^2}$$

$$- 2\sum_{i=1}^{n} \frac{1}{n\sqrt{2\pi(h^2+\hat{\sigma}_0^2)}} e^{-\frac{1}{2(h^2+\hat{\sigma}_0^2)}(Z_i-\hat{\mu}_0)^2} + \frac{1}{2\hat{\sigma}_0\sqrt{\pi}}$$

In our test, the user can choose their prefered bandwidth. For our further analysis, we chose an optimal data-based bandwidth by Sheather et.al [61], beacuse of better theoretical performance and computational advantage [62]. We also explored the power of the test with some other options of bandwidth selection which is discussed

---
[1]Detailed derivation of the NpD statistic in appendix 5.3.

in section 2.2.3. All the analyses are preformed in R version 4.1 [63].

2.2.2 Critical Values and relationship with sample size

To understand the asymptotic characteristics of the test, we approached with bootstrap method. We simulated data from N(0,1), N(3,5) and N(7,3)[2] distribution. Sample sizes in between 50 to 1000 were checked for 1%, 5% and 10% levels of significance (α). The critical values for the respective significance levels and sample sizes were all very similar for all three distribution and fairly linear in logarithmic scale relationship(Appendix Figure 1). Since obtaining critical values from bootstrap was time consuming and they were similar regardless of the different means and standard-deviations, we used the average of the critical values with respect to their sample size and α. Table2.1 shows the fixed critical values obtained from the average bootstrap simulations to use it for evaluating out NpD test.

Table 2.1: Critical values for nonparametric d test

Sample Size	Level of Significance (α)		
	10%	5%	1%
50	0.148	0.176	0.238
100	0.093	0.112	0.152
200	0.058	0.071	0.097
400	0.039	0.047	0.066
600	0.031	0.038	0.052
800	0.027	0.033	0.045
1000	0.025	0.030	0.040

The linear declining trend in Appendix figure 1 shows us the relationship of critical values and sample size. Theoretically for large number of sample size the null density and alternative density will be similar and the critical value will go to 0 to keep the same area under the curve. We further investigated the relationship between sample size and critical values (Appendix table 1 & figure 2) and found significant linear relationship depicted in figure 2.2.

---

[2]Standard-deviation

Log Critical values vs Log Sample Size
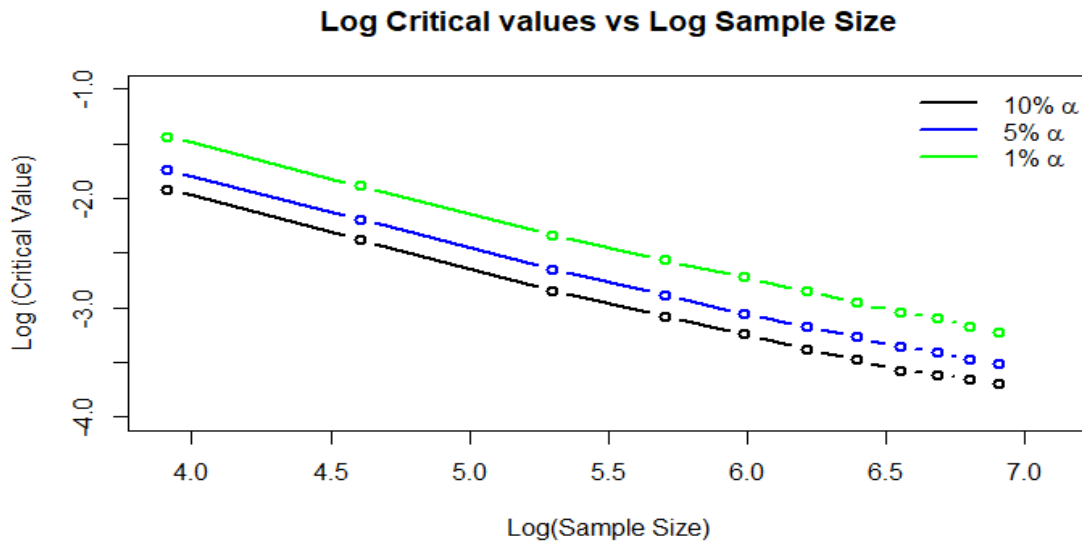
Figure 2.2: Linear relationship between Log Critical value and Log Sample size

Based on the linear relationship between the log values, linear regression is applied (Appendix table 2.1) to calculate critical values for any sample sizes based on their significance level. Table 2.2 shows the necessary coefficients needed to calculate critical values for any sample size. For example, a sample size of n=272, the critical value at 10% significance level is $exp(-0.59702 \times log(272) + 0.3639) = 0.051$, which is in between the critical values of n=200 and n=400 from table 2.1.

Table 2.2: Coefficients to obtain Critical values

Significance Level	Intercept	Slope
10%	0.3639	-0.59702
5%	0.53482	-0.59290
1%	0.844563	-0.592534

The theoretical evidence of the critical values is still under investigations.

2.2.3 Power of the test:

To analyze the power (probability of not making type-II error), data were generated from normal mixture distributions with different mean and standard deviations for 10%, 5% and 1% level of significance. Different bandwidth options for example, 'bw.SJ', 'bw.bcv', 'bw.ucv', 'bw.nrd', 'bw.nrd0' and default '0.7' available in R 'stats' package, were explored to see how power fluctuates which is represented in the following figure 2.3 [62, 64–66]. Figure 2.3 shows no severe fluctuations in power when
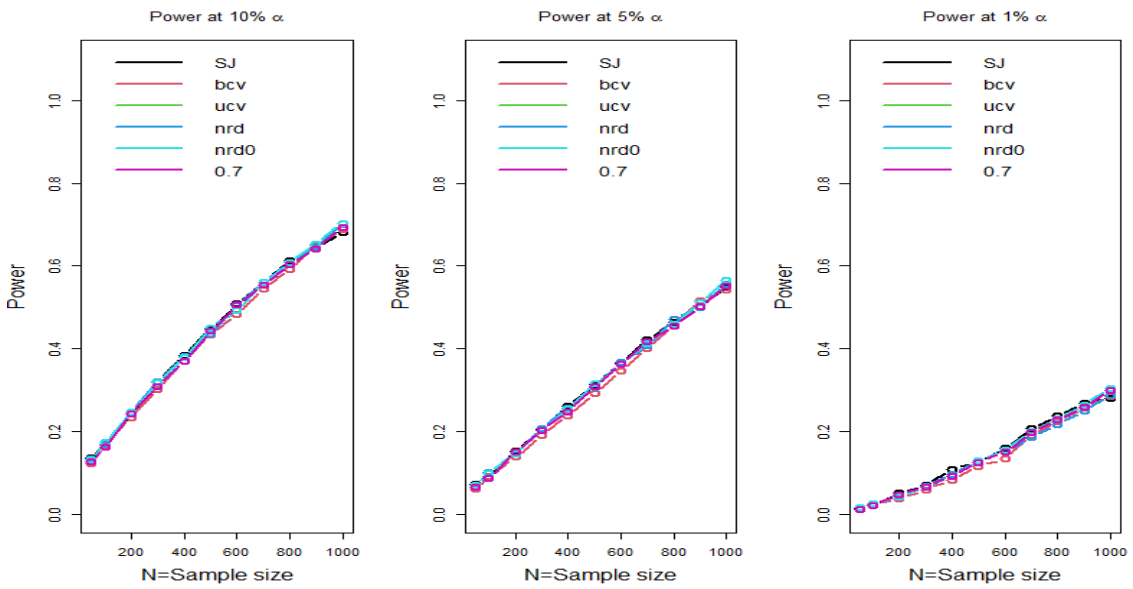
Figure 2.3: Power curves of 0.5 N(0,1)+ 0.5 N(3,1)for different bandwidths

different bandwidth is used in data simulated from 0.5N(0,1)+ 0.5N(3,1). In fact they are really close to each other which gives the flexibility to the user to choose any bandwidth without worrying about fluctuations in results. Thus, the NpD test is robust to bandwidth selection.

Similar to all other heterogeneity detection tests, NpD test shows increase in power with the increase in distance among the means of the mixture components. From figure 2.4, it is evident that the power is higher when the distance among the components mean are 3 and 4 standard deviations respectively, compared to first plot

where the distance among the components are 2 standard deviation. With different variance the power did not change much, as we are using z score in our test which reduces the overall effect of variance.
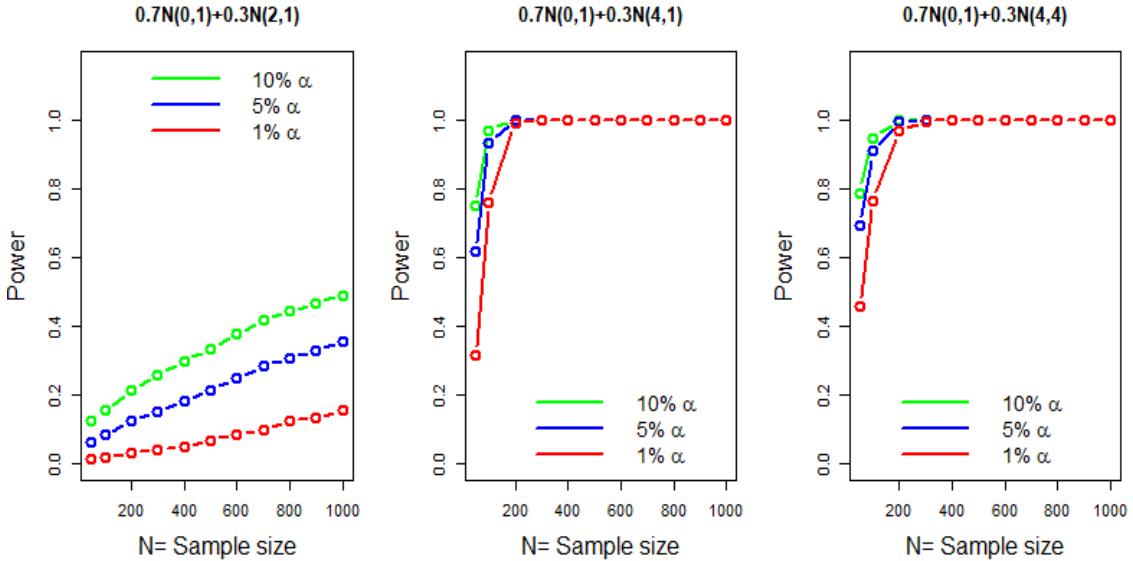
Figure 2.4: Power curves of different 2-component mixing distributions with different distance in mean among the component

In regard to using the empirical null which is based on the data, we also investigated how the mixing proportion affect power in detecting heterogeneity. Figure 3 shows that, NpD test's power is higher for mixing proportion in between 0.2 and 0.8, but not when the mixing proportions are extremely different or extremely similar.

NpD test uses maximum of the density values of z scores as the empirical mean($\mu_0 = argmax(f(z))$), which largely depends on the skewness of the data. For a 2-component mixture model, when the mixing proportion is 0.5-0.5, it technically represents a unimodal situation. Similarly when the mixing proportion is 0.1-0.9 there are fewer values from one component making it harder to detect. When the mixing proportions are different, for example 0.3 and 0.7, the empirical mean will shift toward the second component with 0.7 mixing proportion and fit a nonparametric curve which
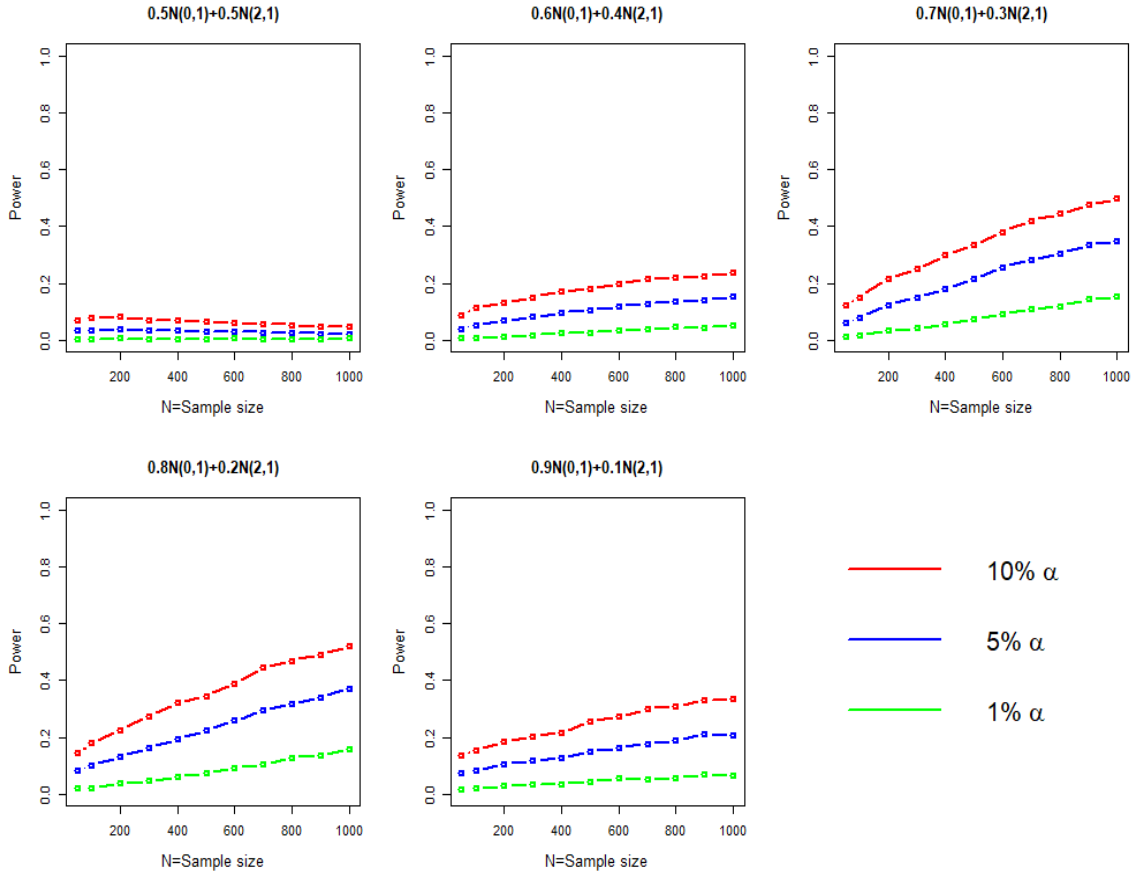
Figure 2.5: Power curves of 0.5 N(0,1)+ 0.5 N(2,1)for different mixing proportions

will be more skewed compared to a 0.5 mixing proportion (Appendix figure 3).

Thus, the NpD test not only depends on the distance of the means but also the mixing proportions of the components. To improve the test for extremely similar or drastic mixing proportions, we proposed a weighted version of the NpD test in section 2.3.

The NpD test was also compared with bootstrap LRT heterogeneity detection method in different type of mixture or contaminated data. Nonparamteric d test performed was faster at detecting heterogenetity compared to bootstrap LRT. Power calculation was attempted but was not completed due to coding issue and time limitation.

2.3 Weighted nonparametric d test:

2.3.1 Definition and motivation:

Charnigo and Sun [16] used a weight function for normal location/scale case. However the mean and the standard deviation defined is not similar to our empirical null and standard deviation. Since we are using z score for our NpD test it is reasonable to use the weight function, $w(x) = exp[-\frac{c(x-\hat{\mu}_0)^2}{\hat{\sigma}_0^2}]$. This weight function places more weights in the central region around $\hat{\mu}_0$ when $c > 0$.

Based on the chosen weight function and our z score of interest, the weighted NpD test statistic is defined as[3]:

$$
\begin{aligned}
d_{NW} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\hat{\sigma}_0}{2hn^2\sqrt{\pi(\hat{\sigma}_0^2 + ch^2)}} e^{\left[\frac{(\hat{\sigma}_0^2 z_i + \hat{\sigma}_0^2 z_j + 2ch^2\hat{\mu}_0)^2}{4h^2\hat{\sigma}_0^2(\hat{\sigma}_0^2 + ch^2)} - \frac{\hat{\sigma}_0^2 z_i^2 + \hat{\sigma}_0^2 z_j^2 + 2ch^2\hat{\mu}_0^2}{2h^2\hat{\sigma}_0^2}\right]} \\
&\quad - 2\sum_{i=1}^{n} \frac{1}{n\sqrt{2\pi(\hat{\sigma}_0^2 + h^2 + 2ch^2)}} e^{\left[\frac{(\hat{\sigma}_0^2 z_i + h^2\hat{\mu}_0 + 2ch^2\hat{\mu}_0)^2}{2h^2\hat{\sigma}_0^2(\hat{\sigma}_0^2 + h^2 + 2ch^2)} - \frac{\hat{\sigma}_0^2 z_i^2 + h^2\hat{\mu}_0^2 + 2ch^2\hat{\mu}_0^2}{2h^2\hat{\sigma}_0^2}\right]} \\
&\quad + \frac{1}{\hat{\sigma}_0^2\sqrt{\pi(1 + c)}}
\end{aligned}
$$

Similar investigations were run and the critical values were obtained using Bootstrap. For the nature of the weighted nonparametric test statistic, various weight values were observed and found to have similar relationship compared to NpD test.

Figure 2.6 depicts the linear relationship between Log Critical value and Log sample size for regardless of the weight values. Thus, the critical values performs similarly to unweighted NpD test. Performance for detecting heterogeneity was also compared and represented in the following figure 2.7.

The above figure 2.7 shows increase in power when the c is higher compared to the

---

[3]Detailed derivation discussed in appendix 5.3
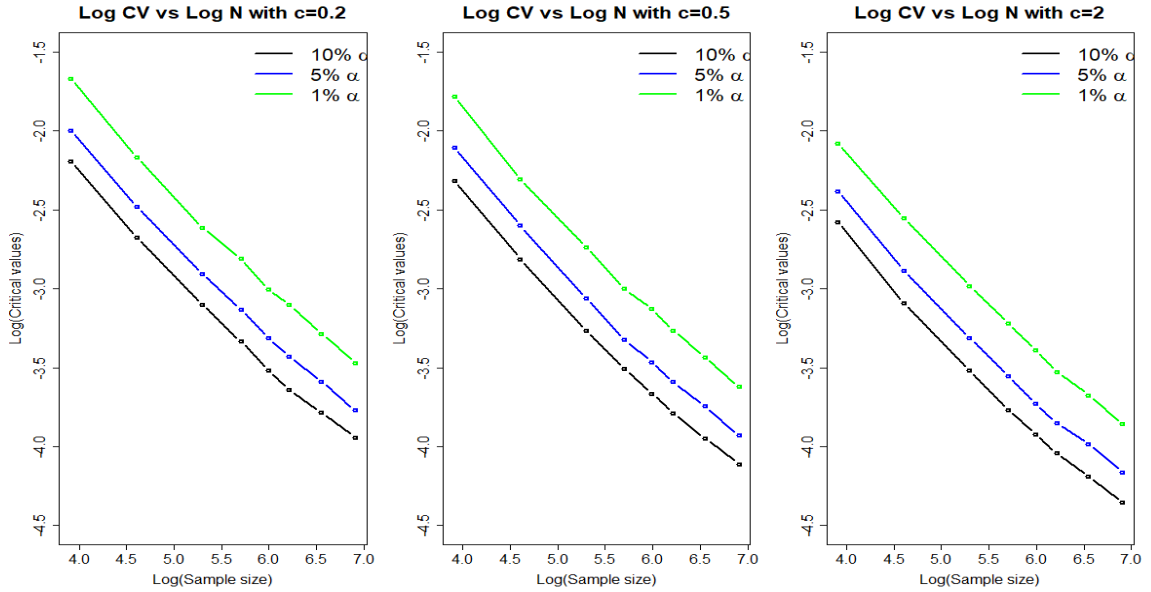
Figure 2.6: Weighted NpD test Linear relationship between Log Critical value and Log Sample size
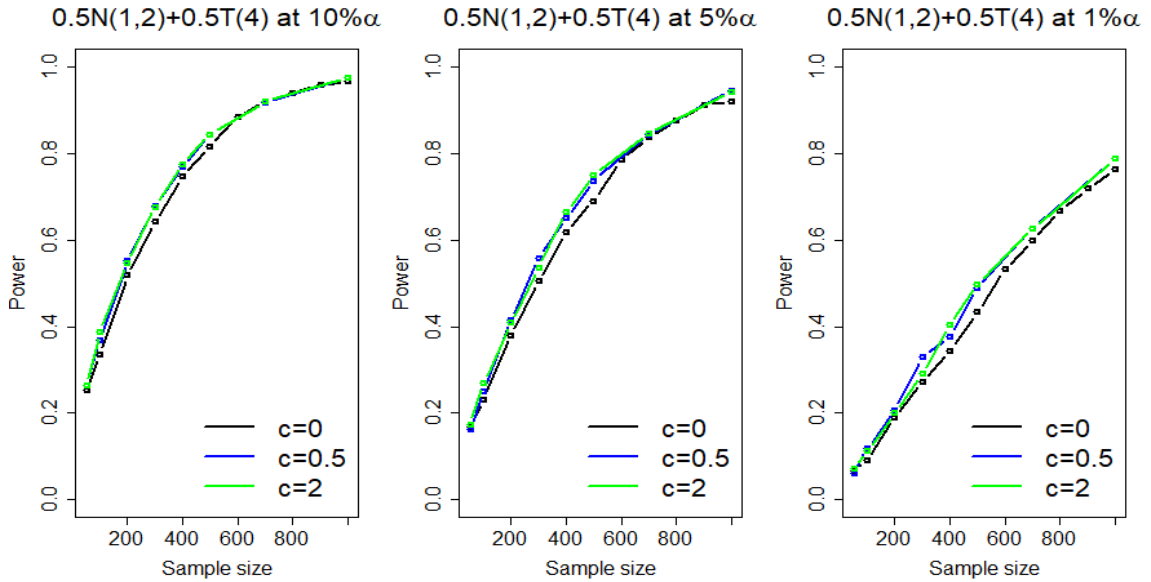
Figure 2.7: Nonparametric d test power comparision with and without weight

unweighted nonparametric d test(c=0). Thus there is a scope to have better power using higher c values to successfully detect heterogeneity.

2.4   Proof of Concept:

To understand the application of nonparametric d test on heterogeneity detection we first used volcanic eruption duration information from faithful data from R [67]. The duration of eruptions were coded as minutes and contain heterogeneity (Appendix figure 4). After applying the NpD test the heterogeneity was confirmed and the results are depicted in table 2.3.

Table 2.3: NpD test result on eruption distribution

Significance Level	Critical values	NpD test Statistic	Decision
10%	0.051	0.239	Heterogeneity detected
5%	0.061		Heterogeneity detected
1%	0.084		Heterogeneity detected

The NpD test was also applied on real data to detect heterogeneity in the discrepancy among prescribed and practiced fluid removal in the ICU among CRRT (Continuous Renal Replacement Therapy) patients. The data was obtained from University of Kentucky clinic from August 2017- April 2021. Initially 1539 adult patients' electronic health record were collected. The data derivation is depicted in figure 2.8

The fluid removal data has cumulative fluid removal information on 793 patients throughout their CRRT duration. There are multiple factors in ICU that disrupts the fluid removal process and this creates a discrepancy among prescribed and practiced fluid removal rate. Thus we wanted to see if it is possible to identify subgroup of patients based on the discrepancy adjusted percentage for CRRT duration (hours) and weight (Kg) depicted in figure 2.9).

To avoid bias from wrong entry on EHR health record, 1% and 99% exclusions were applied as there were still some extreme and unrealistic values in the fluid data. The analysis data have complete information on 777 patients regarding adjusted discrepancy, whether extra fluid was removed from the body and in hospital mortality.
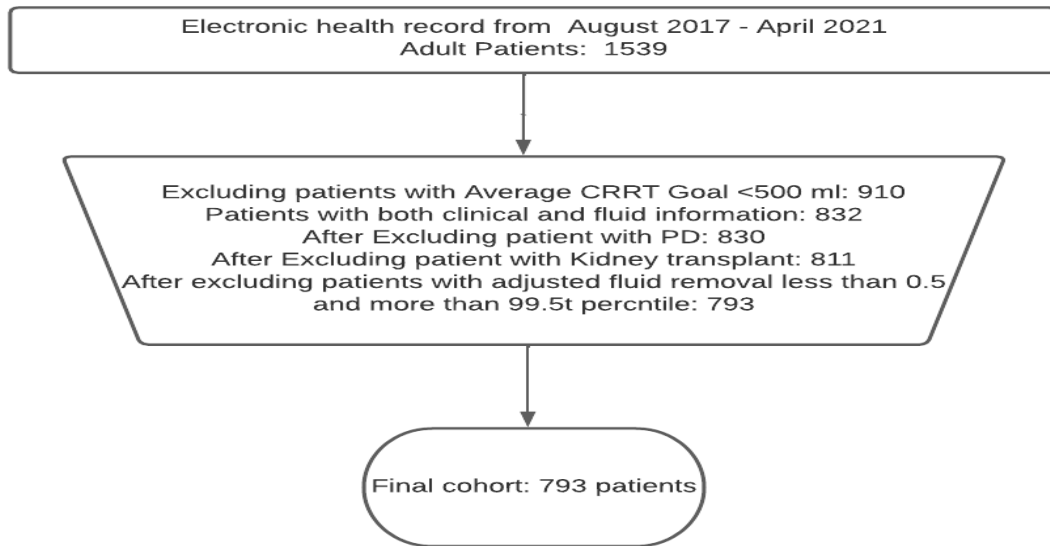
Figure 2.8: Fluid Removal Data Derivation Process
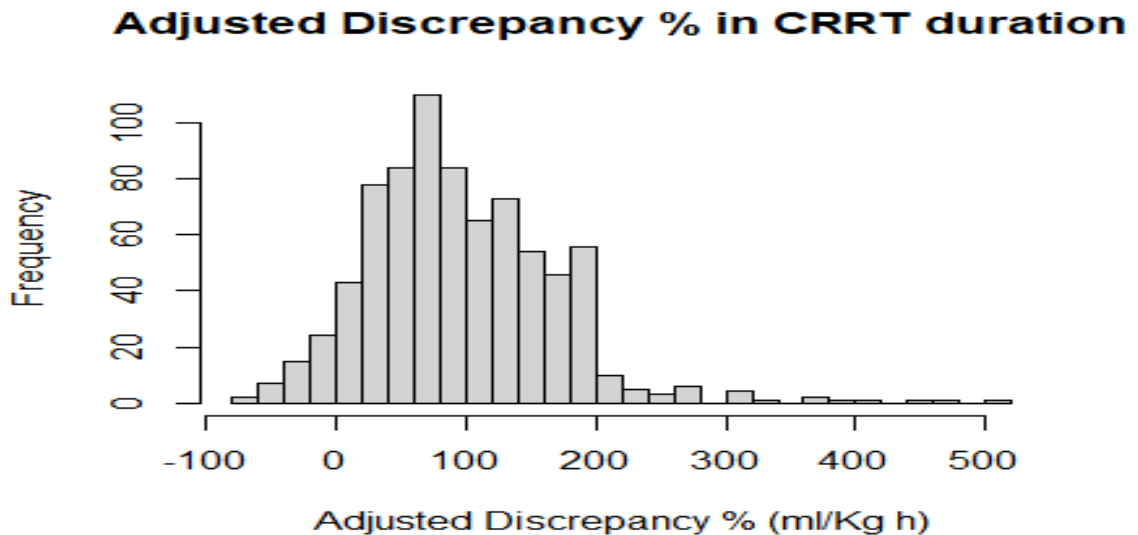
Figure 2.9: Distribution of Adjusted Discrepancy % (ml/ Kg h)

Among 777 patient 464 patients died during their hospitalization period. We applied the NpD test to detect heterogeneity in the adjusted discrepancy among patients.

Table 2.4 shows that there exits heterogeneity in the adjusted discrepancy throughout CRRT duration but failed to obtain evidence at 1% level of significance. First we

Table 2.4: NpD test result on Adjusted discrepancy (%ml/Kg h) during CRRT

Significance Level	Critical values	NpD test Statistic	Decision
10%	0.027	0.041	Heterogeneity detected
5%	0.033		Heterogeneity detected
1%	0.045		Not detected

applied bootstrap likelihood ratio test and estimated there may be 3 or 4 components present in the data. Then using K means clustering, we applied 3 components model and checked how they behave with mortality adjusting with the under or over fluid removal information in figure 2.10 [68].
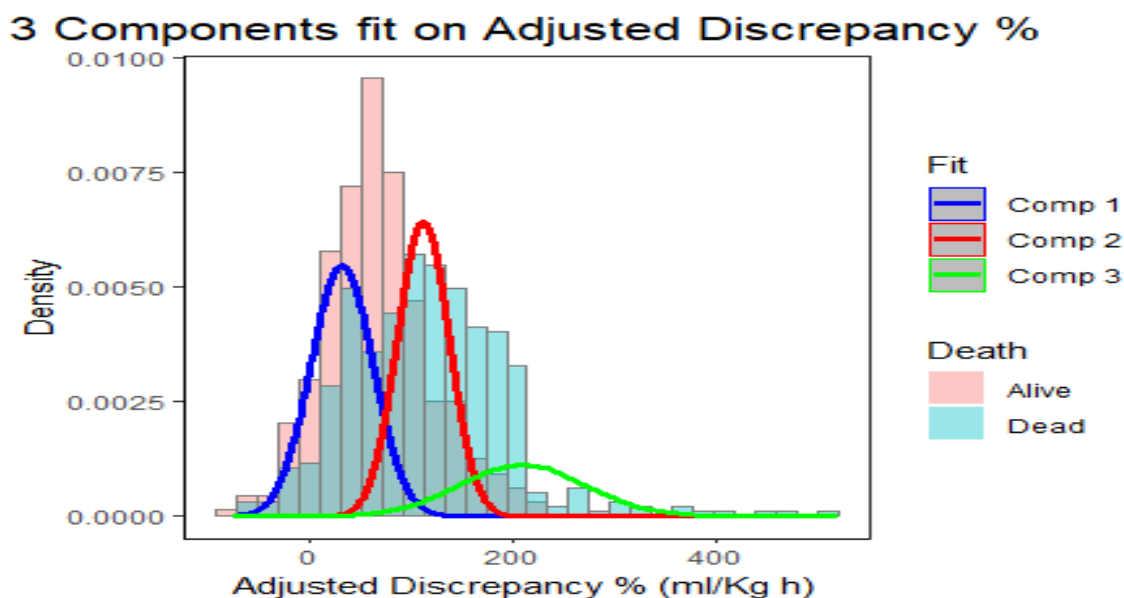
Figure 2.10: 3-Components fit on Adjusted Discrepancy % (ml/ Kg h)

We applied logistic regression to compare the performance of predicting death when using categorical vs continuous adjusted discrepancy. Table 2.5 shows that, using grouped adjusted discrepancy obtained from fitting 3 components performs similarly at predicting hospitalization mortality compared to using numeric values (ROC curve Appendix 5). With lower AIC (Akaike's Information Criteria) and higher Sensitivity value, adjusted discrepancy when analyzed as grouped variable predicts the hospital-

31

ization with the additional advantage of risk grouping.

Table 2.5: Model comparison of Grouped vs Numeric Adjusted Discrepancy

Adjusted discrepancy %ml/ Kg h	Categorical			Continuous
Level	1	2	3	
No. of Patients	325	317	135	777
Actual death	140	205	119	464
Predicted death	0	254	135	367
AIC	948.78			950.43
Specificity	0.703			0.741
Sensitivity	0.640			0.616
Optimal predicted probability cut off based on accuracy	0.452			0.681

The results from table 2.5 shows that the adjusting for under or over fluid achievement the patients in the component 1 are predicted to be healthy or less risk of mortality (predicted death 0%, actual death 43%), component 3 is at high risk of mortality (predicted death 100% actual death 88%) and component 2 (predicted death 80% actual death 65%). Thus using components or categorical version of the information the clinicians will be able to quickly risk stratify the patients and intervene to adjust the treatment. Also, this gives an advantage over not worrying about extreme values when grouped adjusted discrepancy is used.

This is a exploratory investigation to understand how identifying groups will be beneficial compared to just using the continuous information. For future research, we need to investigate how these groups are related to hospital mortality when adjusted for other important clinical factors that affects patients health.

2.5   Discussion:

The nonparametric d test is flexible at detecting heterogeneity compared to other available parametric heterogeneity detection methods. Although the limitation does not detect heterogeneity in data with a small distance in mean components and with an equal mixture, there are still room for improvement using the weighted version of the test. We propose to further investigate appropriate weight function to adapt to the unimodal situation.

We showed how identifying components can improve model prediction. This encourages us to apply mixture modeling while adjusting for other covariates' effects in improving prediction. In the next chapter, we will identify important risk factors by doing a systematic literature search on the fluid removal issues in ICU admitted CRRT patients. After the identifications we will investigate more on the discrepancy between prescribed and practiced fluid removal rate and how it affects in hospital mortality.

Chapter 3 A systematic literature search: Effect of Net Ultrafiltration (NUF) on Mortality among ICU admitted adults

3.1 Introduction

Over the past decade, one of the frequent complications in the intensive care unit(ICU) admitted patients or critically ill patients is acute kidney injury (AKI) [28–30]. About 20-50% of ICU patients suffer from AKI and more than half of them are associated with a greater risk of mortality (50-70%) [69–71]. Overall 5-6% of ICU patients undergo renal replacement therapy or dialysis to reduce mortality [72]. Within a few hours of ICU admission, the kidney suffers from structural damage and loss of function which results in complications in administering treatments among ICU admitted patients [73]. As injured kidney cannot filter out enough waste and toxic substances, the body starts to accumulate excessive fluids, causing fluid overload (FO) [31, 40–42, 74]. Recent studies have shown that FO>10% of body weight is associated with a higher risk of mortality and lower renal recovery [44, 75]

To balance body fluid, continuous renal replacement therapy (CRRT) is a popular treatment, that conducts ultrafiltration 24 hours a day without putting much stress on the heart [76, 77]. Ultrafiltration rate is the suggested volume of fluid removed per hour according to body weight. Even though ultrafiltration rate is used to reduce FO, it did not produce sufficient evidence on how to adjust the rate based on patients' illness severity to mitigate mortality risk [78, 79]. A recent proposed net ultrafiltration (NUF) rate (3.2.3) has become popular to understand the effect on mortality, which is calculated over the treatment duration, but there are some variability in results [80–82]. Some studies suggest that less intensive (slower rate or smaller volume) NUF may be associated with tissue and organ swelling and increased morbidity and mortality

[46] [47]. Some others suggest a faster rate or larger volume causes hemodynamic and cardiovascular stress, which also leads to increased morbidity and mortality [48] [49]. Unfortunately, no optimum rate has yet been found.

With inconsistent conclusions among different studies, we systematically examined the results from randomized controlled trials (RCT) and observational studies conducted on ICU admitted AKI patients undergoing CRRT with the focus on understanding the effect of NUF on mortality and what factors are causing these inconsistencies. If we can identify and control these effective factors, the relationship between NUF rate and mortality will show a p potential path to obtain an optimum NUF rate based on the patient's severity of illness.

3.2 Method

3.2.1 Scope of the Study and Search Strategy

This systematic literature search selected observational and RCT studies that investigated mortality or survival outcomes using NUF among patients who were undergoing CRRT and treated with ultrafiltration. This study was conducted in adherence with the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) with PubMed, CINHAL and Google Scholar employed as the primary search databases [83].

There were no restrictions on language, country, or study design, but publication year. Since the standard definition of FO was proposed in the year 2001, we restricted our publication search starting from the year 2001 [43]. Key search terms for searching all databases are presented in Appendix Table 2.

3.2.2 Eligibility criteria

All studies that investigated the effect of NUF on mortality among ICU admitted adults undergoing CRRT were eligible for review. To ensure homogeneity, we limited our study to which followed certain eligibility criteria. Publications excluded in the study were (1) age ≤ 18 years, (2) review, commentary or editorial article, (3) ESRD patient, (4) investigated ultrafiltration rate, (5) includes COVID-19 cases, (6) Kidney transplant patient, and (7) Unavailability of full text or translated text. Based on the title and abstract, 325 articles were identified, and after removal of duplication and application of eligibility criteria ,37 articles were assessed for full-text review.

3.2.3 Parameters of interest

NUF is defined as the volume of fluid removed per hour adjusted for patient body weight over the treatment duration. That is NUF is the treatment performance measurement of CRRT. Due to many other issues impeding the prescribed fluid removal rate, NUF may account for the gap with the delivered fluid removal rate. The NUF rate or intensity is measured as follows [44]:

$$\text{NUF rate (ml/kg/h)} = \frac{\text{Total NUF volume (ml)}}{\text{weight(kg)} \times \text{treatment duration (h)}}$$

Some researchers used days instead of hours based on their research of interest [80]. In many researches NUF has been categorized in three groups (Low: <1.01 ml/kg/h Moderate: 1.01 to 1.75 ml/kg/h High: >1.75 ml/kg/h) [81, 84, 85]. Apart from categorized NUF, continuous NUF has also been tested in many researches [80, 82].

The primary objective of this study was to evaluate the effect of different NUF rates on mortality. We also want to identify how other factors such as sepsis, hypertension, and illness severity change with NUF. Thus, we compared the clinical characteristics of

the study patients among these publications. We investigated the statistical analysis conducted in these studies and performed an appraisal of their appropriateness based on the presence of indication bias.

3.3 Results

3.3.1 Baseline and clinical characteristics

From 37 full-text assessments, eligibility criteria for study inclusion were applied, and 34 articles were excluded, leaving 3 articles for review. The PRISMA flowchart in Figure 3.1 portrays the study selection procedure.

The baseline and clinical characteristics of patients studied in the 3 selected literatures are represented in Table 3.1. Murugan et.al. 2018 [80] was a secondary analysis of a randomized trial [86] and Murugan et.al. 2019 [81] and Tehranian et.al. 2020 [82] were observational retrospective. The selected studies took place in Australia, New Zealand, and the USA. There were a total of 3907 patients with an approximate mean age of 62 years, and 61.1% of the patients were male. During the studies, 51.2% of the patients died.

All the selected studies observed mortality at different follow-up times and NUF with different treatment duration units (Table 3.1). While Murugan et.al. 2018 were interested in one-year survival, Murugan et.al. 2019 and Tehranian et.al. 2020 were interested in 90 days and 30 days mortality respectively. Tehranian et.al. 2020 also explored 90-day mortality as a secondary outcome. Both Murugan et.al. 2018 and 2019 found nonlinear association of continuous NUF rate with their respective mortality outcome and selected categorized NUF rate (Murugan et.al. 2018: $\leq 20, 20-25, \geq 25$ ml/kg/day, Murugan et.al. 2019: $\leq 1.01, 1.01 - 1.75, \geq 1.75$ ml/kg/hour) based on tertiles with lowest Akaike Information Criterion (AIC) value. Tehranian et.al. 2020 selected a categorized NUF rate ($\leq 35, \geq 35$ ml/kg/day) based on median value after
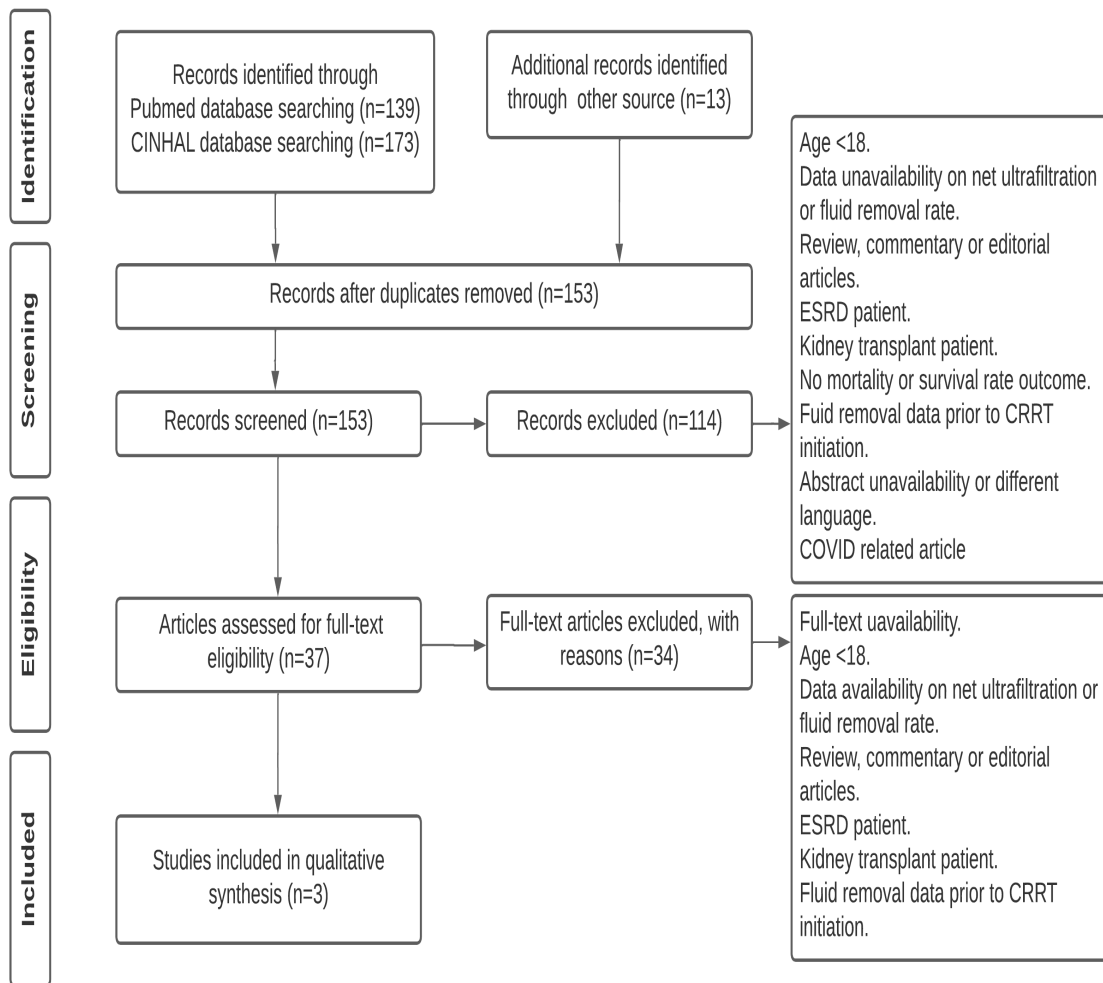
Figure 3.1: PRISMA flowchart of study selection & search strategy

exploring the relationship between continuous NUF and 30-day mortality.

Thus, with different exposure and outcome measurements, we could not apply meta-analysis for quantitative assessment of the studies. But to assess the information, we dove into the clinical and methodological diversity of the studies for further quality assessments.

Table 3.1: Study participant baseline and clinical characteristics

Study	Murugan et.al. 2019			Murugan et.al. 2018			Tehranian et.al. 2020	
Country	Australia, New Zealand			USA			USA	
Study design	Secondary analysis of RCT			Observational retrospective			Observational retrospective	
Total no. of patient	1434			1075			1398	
Outcome	Mortality 90 days from ICU admission			Mortality 1 year from ICU admission			Mortality 30 days from ICU admission	
Exposure measurement duration and unit	CRRT duration ml/kg/hour			CRRT duration ml/kg/day			CRRT duration ml/kg/day	
Exposure level	<1.01	1.01-1.75	>1.75	≤20	20-25	≥25	<35	≥35
No. of patient in exposure group	477	479	478	475	166	434	696	702
Death (%)	214 (45)	188 (39)	232 (49)	331 (70)	100 (60)	258 (59)	420 (60)	335 (48)
Age Median (IQR)/ Mean ±SD	69.3 (61.0-77.4)	68.1 (57.2-76.1)	63.8 (51.4-74.2)	61 (52-69)	59 (51-71)	58 (48-70)	63±15	60±15
Male (%)	311 (65.2)	330 (68.9)	283 (59.2)	301 (63.4)	114 (68.7)	218 (50.2)	445 (63.9)	382 (54.4)
APACHE III Score Median (IQR)/ Mean ±SD	101 (84-118)	100 (83-118)	101 (84-117)	95 (70-118)	91 (71-116)	91 (69-112)	103±29	104±30
Sepsis (%)	221 (46.3)	235 (49.1)	253 (53.0)	128 (26.9)	39 (23.5)	138 (31.8)	NA	NA
Mechanical ventilation (%)	330 (69.2)	356 (74.3)	371 (77.6)	353 (74.3)	129 (77.7)	329 (75.8)	586 (84)	648 (92)

In both Murugan et.al. 2019 and Tehranian et.al. 2020 the median or mean APACHE III score was higher and the median or mean cohort age was also older compared to Murugan et.al. 2018. Unlike the other two studies, Tehranian et.al. 2020 did not have any information regarding sepsis incidence. In Murugan et.al 2019, approximately 50% of patients in each exposure group had sepsis, but Murugan et.al 2018 had a lower sepsis percentage in each exposure group. Mechanical ventilation was observed and present in more than half of the study cohort among all three publications.

All of the studies applied some similar exclusion criteria to provide valid comparison of the results and reduce confounding effects. Murugan et.al. 2019 excluded patients with missing treatment duration or NUF<0.01 ml/kg/hour. Murugan et.al. 2018 excluded patients with missing baseline weight, NUF, and fluid balance (missing or <5% of body weight). They also excluded patients if the ICU duration was ≤48 hours and they died within 72 hours of ICU admission. Tehranian et.al. 2020 also excluded patients with FO<5% of their body weight and patients who died within 24 hours of CRRT initiation.

3.3.2 Methodological diversity

To understand the relationship between NUF and mortality, the studies explored factors that were significant in both unadjusted and adjusted settings. In table 3.2, we summarized the potential risk factors and statistical methods applied to investigate the NUF and mortality relationship.

Table 3.2: Statistical method and adjusted risk factors

Study	Murugan et.al. 2019	Murugan et.al. 2018	Tehranian et.al. 2020
Statistical method	Gray's time-varying survival model	Gray's time-varying survival model	Logistic regression
Adjusted variables	Age, Sex, APACHE III score, Mechanical ventilation, ICU to CRRT duration		
Other adjusted variables in the statistical model	Baseline eGFR, ICU & Hospital type, Region, SOFA score, CRRT duration, Admission source, Region, Cumulative fluid balance.	Race, BMI, Liver disease, Admission source, Baseline eGFR, Sepsis, MAP, Oliguria, ICU type, Cumulative vasopressor dose, Cumulative fluid balance.	BMI, Baseline serum creatinine, Charlson score, Hypotension, Fluid balance.
Missing potential information	Charlson score, Race, Comorbid conditions, Hypotension, Diabetes	SOFA score, Charlson socre	Sepsis, Diabetes, Vasopressor dose

Murugan et.al. 2018 and 2019 used Gray's piecewise time-varying survival model to evaluate one-year and 90-day mortality respectively. The Gray's survival model allows of violation of proportional hazard assumption by using time varying covariates in subdistributional hazard model in the presence of competing risk [87, 88]. Tehranian et.al. 2020 used Logistic regression for 30-day mortality due to assumption violation for Cox's PH model. In Murugan et.al. 2018, high-intensity NUF ($\geq$25 ml/kg/day) was significantly associated with lower mortality risk until 39 days after ICU admission compared to low-intensity NUF ($\leq$20 ml/kg/day). When compared to moderate-intensity NUF (20-25 ml/kg/day) high NUF was associated with lower mortality until 15 days after ICU admission. In Murugan et.al. 2019, high NUF ($>$1.75 ml/kg/h) was significantly associated with higher mortality but only after 6 days of ICU admission. Appendix figure 6 shows the significant results obtained from these two studies using Gray's survival model.

To find associated risk factors, all three publications adjusted age, sex, APACHE III score, mechanical ventilation, cumulative fluid balance, and ICU to CRRT duration for the association between NUF and mortality. Both Murugan et.al. 2018 and Tehranian et.al. 2020 adjusted for BMI, fluid balance, or cumulative fluid balance, and both Murugan et.al. 2018 and 2019 adjusted for baseline eGFR, ICU type, admission source, and cumulative fluid balance. They have also adjusted some other variables and missed some potential risk factors to investigate, as shown in table 3.2.

Table 3.3: Evaluation of bias assessment and sensitivity analysis

Study	Murugan et.al. 2019	Murugan et.al. 2018	Tehranian et.al. 2020
Sensitivity analyses	Propensity score matching. Analysis of NUF restricted to 72h of therapy initaiton. Analysis of NUF with different threshold. Analysis of NUF as continuous value. Analysis of with and without patient with missing treatment hours. Stratified analysis of restricting therapy duration (3 or more days and 5 or more days) Subgroup analysis with and without (organ edema, sepsis, eGFR>60)	Propensity score matching. Analysis of NUF restricted to 72h or therapy initation. Analysis of NUF with different threshold. Quantitative bias sensitivity analysis for unmeasured binary confounder. Subgroup analysis of patient with > 20% FO	Analysis excluding patient with FO < 5% or died within 24h of therapy initiation. Analysis of NUF with different category based on quartiles. Subgroup analysis for different ICU types. Propensity score derived inverse probability weighting accounting for early hypotension.
Biases addressed	Measured confounding bias. Selection bias. Indication bias.	Measured Confounding bias. Selection bias. Indication bias.	Measured Confounding bias. Selection bias. Indication bias
Bias not/ can't be addressed	Unmeasured confounder. Measurement bias	Unmeasured confounder. Measurement bias	Unmeasured confounder
Limitations	Missing information on potential risk factors. Did not check association with prescribed dose.	Not generalizable. Did not test interaction. Did not check if difference in mortality after hospitalization is due to any other issues.	Not generalizable Did not check association with prescribed dose.

To understand the appropriateness of statistical application in the robustness of the results, we evaluated the assessment of bias and sensitivity analysis. Table 3.3 summarizes the bias assessment among the studies. Since all of them are observational studies, selection bias and indication bias are inevitable, along with unmeasured confounding bias. For example, the patients are receiving high fluid removal rate based on their sickness and the high fluid removal may in term cause patients intolerance and harm other organs and increase the mortality risk. The selected patients are very sick and thus there already exists a high mortality risk. Also, it is not possible to measure all clinical factors and treatment interruptions. But the studies used appropriate schemes to reduce the effect of bias in results by doing sensitivity analyses.

The studies still suffer from unmeasured confounding effects due to their observational nature, but regardless of their limitations, all of the studies emphasized the importance of randomized clinical trials and investigation for optimal NUF rate.

The relationship between NUF and mortality from the studies is not consistent. Table 3.4 shows the conclusions and other findings among the studies. Murugan et.al. 2019 showed that higher NUF ($\geq$1.75 ml/kg/hour) is associated with higher mortality, which is contradictory compared to the other two studies. Moreover, longer ICU duration, oliguria, liver disease, and other comorbid conditions are associated with NUF and mortality.

Table 3.4: NUF and mortality association among studies

Study	Key findings	Other Findings
Murugan et.al. 2019	High NUF($\geq$1.75 ml/kg/hour) associated with higher mortality	High NUF patients had longer ICU, hospital and mechanical ventilation duration.
Murugan et.al. 2018	High NUF($\geq$25 ml/kg/day) associated with low mortality	High and moderate NUF patients had higher prevalence of oliguria
Tehranian et. al. 2020	High NUF ($\geq$ 35 ml/kg/day) associated with low mortality	High FO patients benefitted from high NUF and ICU type and NUF had significant interaction effect.

3.4 Discussion

This systematic literature search suggests that NUF is associated with mortality but cannot identify the nature of the relationship due to conflicting conclusions. More information and research is needed to characterize the relationship. The cohort of interest in various studies is different in terms of age, illness severity, exposure measurement, and follow-up time. Even though sensitivity analysis has been performed, the NUF categories are specific to the individual studies. To reduce inter-study variation and increase agreement in future research, it is helpful to identify a general high, medium, and low category for the NUF rates.

However, the choice of NUF categories could also be investigated with other related variables. For example, none of the 37 papers extracted for full-text evaluation explored the relationship of NUF categories with prescribed dose, a limitation acknowledged by, e.g. Murugan et.al. 2018 [80]. Low NUF can be related to low fluid removal prescriptions, which can be due to the patient's being less ill compared to patients who received higher fluid removal prescriptions. Thus, future research should analyze whether the prescribed dose is a potential confounder in the relationship between NUF and mortality.

Our study is not without limitations. With very little research conducted on NUF with patients not with ESRD, it is not possible to reach a clear conclusion at this time. Also, there is a possibility that our study may suffer from positive result bias from publication (only positive results' preference for publication), and some studies were not evaluated due to the unavailability of translated texts. Even though the study populations are quite different in terms of age and illness severity, we emphasize multiple inclusion-exclusion criteria to mitigate inter-study variation. Our study could not identify the actual relationship between NUF and mortality but explored the advantages and shortcomings of the selected studies. For future research, we want

to highlight the importance of including clinical information like sepsis, diabetes and searching for possible interactions between them. Because the effect of the NUF rate can be different based on different group of patients who may have different clinical characteristics. Also, the prescription dose might be informative in explaining the conflicting results.

All the selected studies investigated and confirmed the nonlinear 'J-shaped' association between NUF and mortality under the respective follow-up period [80–82]. This indicates that the effects of NUF can be positive or negative based on a specific group of patients. Identification of factors causing the nonlinear association would be imperative in calculating the optimum NUF rate based on patients' illness severity.

In conclusion, there is a pressing need for a clinical trial to find optimal NUF rates and understand the nature of the effect of NUF on mortality. With the complications in implementing clinical trials, more observational studies can also be helpful by investigating the relationship between prescribed dose and potential interaction effects to portray the relationship nature of NUF on mortality.

Chapter 4 Assessment of Fluid Removal among CRRT Patients in ICU

4.1 Introduction

Continuous renal replacement therapy (CRRT) is a popular treatment among critically ill patients admitted to the ICU [89–91]. It helps patients in the removal of excess fluid from their bodies and the maintenance of fluid balance. When the kidneys fail to work effectively, the body stores excess fluid, resulting in fluid overload (FO). The presence of FO in the ICU raises the risk of adverse outcomes and mortality. [42, 92–94]. During CRRT, patients are given a fluid removal objective (daily CRRT goal) for each day to maintain bodily fluid balance. While it is ideal to achieve the prescribed amount of fluid removal, most of the time it is not possible due to the severity of the disease and other technical challenges in administering treatment. As a result, there is a difference between the fluid removal rate that is practiced and the rate that is prescribed.

Theoretically, the less variation in therapy execution, the more fluid equilibrium in the body, which will assist patients to avoid the adverse effects of FO issues. The discrepancy could be caused by the patient's incapacity to tolerate treatment or by variations in treatment evaluation and management in the ICU [50]. In this study, we investigated whether the rise in treatment discrepancy is linked to characteristics that can be measured during CRRT administration in the ICU, and whether they can be managed or improved to give better healthcare.

4.2 Methods

4.2.1 Study design and population

Using data from the ICU to 48 hours following the start of CRRT, we investigated the relationship between fluid removal discrepancy and hospital mortality in CRRT patients, as well as the characteristics that are associated to changes in the discrepancy. We want to use a prediction model to evaluate the role of disparity in predicting hospital mortality after 48 hours of CRRT initiation. We ran a longitudinal analysis using daily fluid data to see how the patient's sickness and therapy implementation are linked to the daily change in disparity.

We used electronic health records (EHR) collected from a single-center, retrospective cohort study of ICU admitted critically ill patients at the University of Kentucky from August 2014 to April 2021 (from IRB approved study). In the primary analysis of the hospital mortality prediction, all adult patients who received CRRT and remained alive for at least 48 hours after the beginning of CRRT were included. Day 0 is the first day of the CRRT (first 24 hours) determined from midnight, day 1 is the second 24 hours, and so on. To identify the association of daily discrepancy, we focused on a week's worth of data on patients who survived at least a week of their CRRT length. The timeframes and information used for the analysis are depicted in Appendix figure 12 .

Implementation of CRRT does not always begin at day 0 onset for certain patients. As a result, the intended aim of fluid removal on day 0 is frequently not fulfilled due to admission at a later time of the day or the suggestion of not enough fluid to be removed in a short period of time. Since we did not have the whole day's information on day 0, we used the CRRT goal from day 1 to correctly identify the discrepancy. Day 1 fluid information also carries the influence of the treatment at day 0, therefore omitting day 0 in the analysis will not result in a much loss of information due to the

information instability.

4.2.2 Data Collection and Management

To adjust for confounding and selection bias and ensure reproducible results, inclusion and exclusion criteria were applied. Patients with average CRRT goal < 500 ml were excluded from the study. Also only patients who had both clinical and daily fluid information available were included in the study. Due to human error there were patients with incorrect admission date or CRRT start date. Since there is no other way to know the correct date we had to exclude those patients as well. To understand the fluid discrepancy effect and ensure we have patients information for at least 48 hours of CRRT initiation all patients who died within 48 hours of onset of CRRT were excluded from the study. Patients were also excluded if they had received peritoneal dialysis (PD) or had undergone kidney transplant. Since we could not use day 0 information we included patients based on available information on day 1.

After the application of inclusion-exclusion criteria, 690 patients were selected for the study. After evaluating extreme observations, to adjust for wrong entry on EHR, percentile exclusion on patients (< 1% and > 99%) were implemented based on daily CRRT goal. The choice of percentile exclusion was based on clinical rational and descriptive analysis. The percentile exclusion for EHR bias resulted in 680 patients, removing 10 patients from the study (Appendix figure 4.2).

Missing values in the daily fluid data were imputed using successive value imputation after percentile exclusion. In general, the daily CRRT goal should be prescribed quite similarly for two days in a row. Thus we imputed missing values on day 0 with day 1 and day 1 with day 2, while missing value 2 was imputed with day 1 and so on. After manual imputation, on day 1, there were still 8 patients with missing information on their daily CRRT goal. There were several other variables in the data set that had

missing values, depicted in appendix figure 7.

The missing values among the clinical and fluid data provided were not more than 25% (Appendix figure 7). To impute other missing values, we relied on multiple imputation (MI) [95–97]. MI with classification and regression tree is a good solution since it is robust against outliers, multicollinearity, and also well suited for skewed data [98, 99]. As there were no missing outcome values and the missing percentage among the variables was less than 40%, we used MI with CART. To show missing values are missing completely at random, we applied the 'mcar_test' in R 'naniar' package and obtained a P value of < 0.001. This indicated that the missing values are not completely at random [100]. That is the data is either missing at random or non-ignorable. With this limitation we used both the potential variables related to mortality and other auxiliary variable that may or may not be associated with mortality to imeplement CART imputation [101].

CART imputation was applied using 'mice' package in R [102]. The method generates five (default) sets of data with imputed values and then randomly selects one value from each imputed set to replace the missing value in the original data. The distribution comparison of imputed data set and original data is presented in Appendix figure 8 (original data: blue, imputed data: red). The distribution curves for the majority of missing variables are similar. For some of them the shapes are quite different but those variables did not have missing values for more than 5%. Thus, with the limited resources and time constraint, we used CART imputation with the advantage of random imputation selection.

4.2.3 Variables of Interest:

Based on clinical perspective, the discrepancy between the prescription and the practiced fluid removal rate measured in percent milliliters is computed as follows:

$$\text{Discrepancy}(\%\text{ml}) = \frac{\text{Prescribed goal fluid in the body} - \text{Net fluid in the body}}{\text{Prescribed goal fluid in the body}}$$
$$\times\ 100$$

Since the discrepancy also depends on the treatment duration and body weight, we selected adjusted discrepancy as the variable of interest to evaluate the discrepancy. The adjusted discrepancy (%ml/Kgh) is more useful for assessing discrepancy over weight and treatment hours.

$$\text{Adjusted daily discrepancy (\%ml/Kgh)} = \frac{\text{Daily Disrcrepancy}}{\text{Weight} \times \text{Treatement hours}}$$

The adjusted discrepancy will be interpreted as follows: A negative adjusted discrepancy of 16 ml/Kgh would be defined as the fluid being removed but it was 16% more than the prescribed goal per kg per hour. A positive adjusted discrepancy of 20 ml/kgh would be defined as enough fluid was not removed and it was 20% less than the prescribed goal per Kg per hour. There were still some patients with a high ($> \pm 100\% ml/Kgh$) daily discrepancy on day 1, as depicted in figure 4.1. We investigated whether various categorized versions of adjusted discrepancy are better at identifying patient clusters and predicting hospital mortality.

The primary outcome of interest is hospital mortality after 48 hours of CRRT initiation. We are interested in hospital mortality from day 2 (Appendix figure 12). Our secondary outcome of interest is daily discrepancy analysis among patients who survived up to a week from their CRRT initiation. We analyzed patients who had
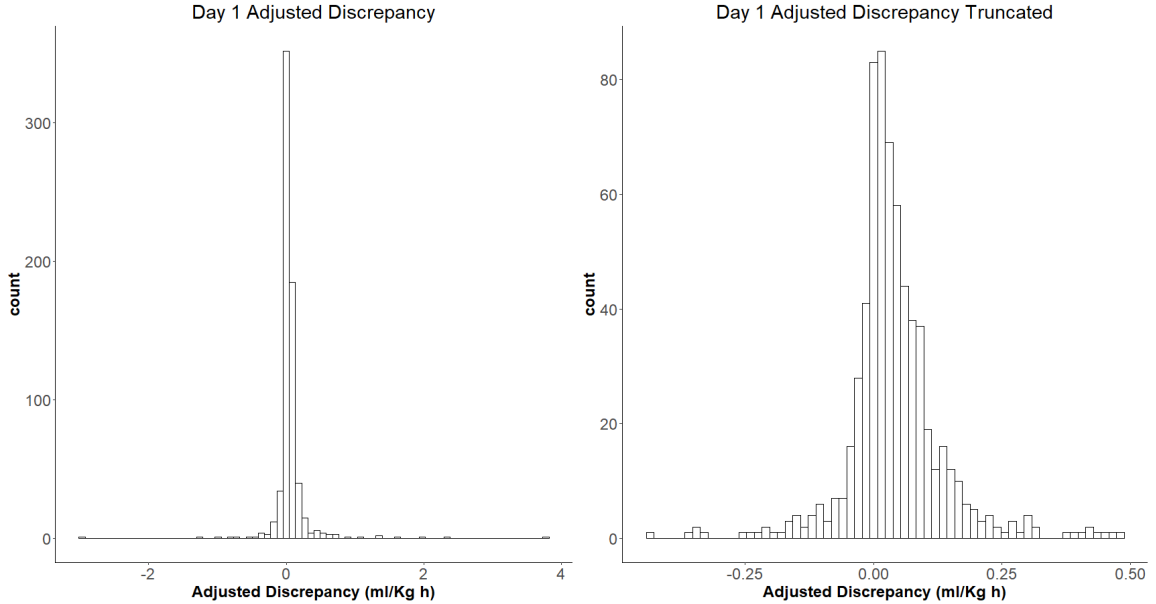
Figure 4.1: Adjusted discrepancy at day 1

daily fluid information from day 1 to day 6 (6 days). We explored how actual fluid re-
moved from the machine, other clinical parameters related to sickness of the patients
(end stage renal disease (ESRD), Sequential Organ Failure Assessment (SOFA) Score,
Charlson score, FO%) and demographic information, are related to daily discrepancy
change.

4.3 Statistical Analysis

4.3.1 Primary Analysis

To avoid overfitting the hospital mortality prediction model, 70% of the study pa-
tients' (476 patients) information was selected at random as the train data for building
the model, while the remaining 30% patients (204 patients) information were used to
test prediction performance. Other splitting percentages (40%-60% and 20%-80%)
were also explored, but the initial variable selection did not show much variation.
Some clinical characteristics (e.g. race, ESRD) were similar in more than 80% of the
study patients. Thus, we proceeded with a 30:70 split based on a rule of thumb in

order to have adequate information on the test data prediction [103, 104].

To check for unadjusted association, Fisher's exact, Chi-square & t-test (based on the variable type and distribution) were used. To investigate the categorical version of the adjusted discrepancy at day 1, manual selection and different statistical clustering algorithms (K means clustering, EM algorithm from R 'stat' and 'mixtools' package) were utilized [64, 105–107]. The K-means and EM algorithm clustering techniques are sensitive to data with high kurtosis (highly peaked data). Thus, we truncated the adjusted discrepancy within ± 50% on day 1 to apply these clustering techniques. Due to the truncation, there were 24 patients omitted from the cluster analysis. These patients were eventually included in the extreme groups based on their signs of change (Appendix figure 4.1).

Important variables were investigated utilizing random forest, single regression tree, logistic regression and gradient boosting for prediction model. R 'vip' package was used to identify first top 10 important variables based on their prediction score for each methods are shown in appendix figure 9 [108, 109]. Based on the all important variables identified from various approaches, we fitted a logistic regression (R 'stat' package) and conducted stepwise selection to determine the factors that contributed the most [64]. Other clinically relevant variables (e.g. ICU to CRRT duration, Charlson Score) were also investigated in the model with the likelihood ratio test (LRT).

Potential interaction effects were investigated based on the feasible solution algorithm and prior literature suggestions [44, 110]. The feasible solution algorithm was applied with both with and without fixed main effects. Single regression tree was also used to identify interaction effects based on their odds of mortality from defined cutoff points [111]. For initial interaction identification we limited the model to have one interaction term. After selecting multiple potential interaction effects, we used LRT

and Akaike information criterion (AIC) values to determine which and how many interaction term best fits the model.

For final model selection was based on AIC and accuracy from 3-fold cross validation (CV). We also performed a 10-fold CV but received warnings due to some characteristics separation (race is same among 89% of the patients). To evaluate the fit of the model and prediction performance, the Hosmer-Lemeshow goodness of fit test, sensitivity, specificity, and ROC curve were explored among test and train data.

4.3.2 Secondary Analysis

To identify associations with daily discrepancy and clinical factors, a linear mixed effect model with subject specific random intercept and slope was used. We used SAS 9.4 version for this longitudinal analysis [112]. Initially we explored daily actual fluid removed, fluid overload at ICU admission, SOFA score at ICU admission and CRRT initiation, Charlson score at ICU admission, ESRD status, demographic information: age, race, sex, duration from ICU to CRRT initiation and ICU type.

Several covariance structures: unstructured, topelitz, gaussian errors and exponential decay were explored. We also explored random group effect based on ESRD, race and sex. Both quadratic and spline models were compared, and the model selection was based on AIC, Bayesian information criterion (BIC) and LRT. We used empirical estimates, as it gives a consistent estimates of covariance structure (robust covariance estimator) whether or not the fitted covariance structure is true structure.

4.4 Results

4.4.1 Study Population & Clinical Characteristics

There were 680 patients selected to analyze hospital mortality. The cohort derivation process is represented in detail in figure 4.2.
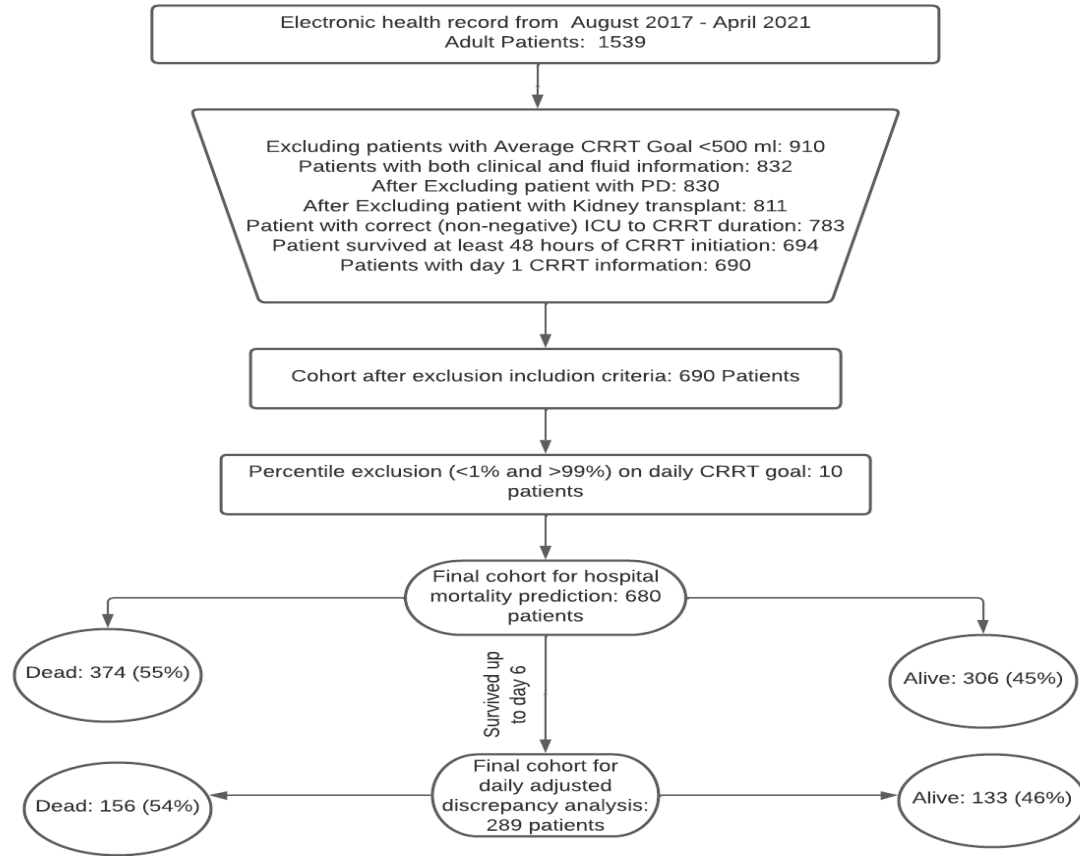
Figure 4.2: Cohort Derivation Based on Inclusion-Exclusion Criteria

During hospitalization, patients' demographic and clinical characteristics from both test and train data are represented in table 4.1. Because the data came from a single center research, there isn't much demographic variation; 90% of the patients are white. Patients from 'Asian' , 'African/American' and other races were combined into "Non-white" group beacause they were very small in proportion. There were slightly more male patients than females. No severe differences among the train and test data were observed. Overall, among 680 patients, 374 (55%) died after 48 hours of CRRT initiation. That is, more than half of the study patients experience the event of interest (hospital mortality) according to the descriptive statistics.

Table 4.1: Patient's Characteristics Across Train and Test Data Throughout Hospitalization

Patient Characteristics	Train Data: N=476	Test Data: N=204
Age	60.5 (48 - 68)	57 (47.75 - 64)
Weight (Kg)	91 (77.2 - 108.03)	95.4 (78.68 - 113.35)
Sex (Male)	289 (61 %)	123 (60 %)
Race (White)	426 (89 %)	184 (90 %)
Time from ICU to CRRT Initiation	2.07 (1.12 - 5.63)	2.46 (1.29 - 4.65)
Baseline Serum Creatinine (mg/dL)	1.73 (1.06 - 3.19)	1.65 (1.02 - 3.27)
Baseline eGFR (ml/min/1.73m^2)	38.11 (18.46 - 71.48)	44.28 (16.45 - 77.51)
ICU FO% (L)	2.19 (0 - 6.31)	3.02 (0.33 - 9.14)
Charlson Score	5 (3 - 7)	5 (2 - 7)
SOFA Score at ICU Start	11 (9 - 14)	11 (9 - 14)
SOFA Score at CRRT Start	13 (11 - 15)	14 (11 - 16)
CRRT Duration (day)	4.81 (2.88 - 8.64)	4.3 (2.65 - 8.48)
Hospital Length of Stay (day)	20 (10.67 - 35.85)	18.95 (11 - 38.42)
ICU Length of Stay (day)	11.8 (6.88 - 22.75)	11.3 (6.47 - 21.4)
ICU Type (Cardio)	153 (32 %)	58 (28 %)
ICU Type (MICU)	267 (56 %)	116 (57 %)
ICU Type (Surgery)	48 (10 %)	27 (13 %)
ESRD	396 (83 %)	167 (82 %)
RRT Dependence	375 (79 %)	142 (70 %)
IABP or VAD	13 (3%)	3 (1 %)
ECMO	72 (15%)	29 (14 %)
Mechanical Ventilation	441 (93%)	190 (93 %)
FO% at CRRT Day 1 (L)	-0.47 (-1.86 - 1.12)	-0.39 (-1.85 - 1.69)
Adjusted Discrepancy at Day 1	0.03 (0 - 0.08)	0.03 (0 - 0.09)
Last SCr	1.58 (0.96 - 3.06)	1.43 (0.9 - 2.55)
Last eGFR	43.94 (19.17 - 79.36)	52.53 (23.69 - 84.14)
Hospital Mortality	268 (56 %)	106 (52 %)

*Categorical variable: N(%), Continuous variable: Median (IQR)

Patients are admitted to the ICU with a higher median FO% compared to the median FO% at CRRT initiation. This was anticipated as patients receive necessary treatments to remove excess fluid from the body, thus lowering FO at the beginning of CRRT. The median duration of CRRT initiation from ICU admission is more than 2 days for both train (2.07 days) and test (2.46 days) data. The median adjusted discrepancy at day 1 is positive in both train (0.03) and test (0.03) data. That is, we have more patients in our data who did not meet the prescribed fluid removal goal than those who attained the goal fluid removal. When comparing the median SOFA score at both ICU admission and CRRT initiation, there appears to be a slight increase, which indicates that the patients got more sick and thus had to start CRRT. In both train and test data, almost 83% have a history of end stage renal disease. Mechanical ventilation was used by 93 percent of patients during their hospital stay, but the majority of them started it after 48 hours of starting CRRT.

Among 680 patients, only 289 patients survived for at least a week (up to day 6) from the onset of CRRT. Thus, the analysis data for the secondary outcome included 289 patients' daily fluid and demographic information for a week (figure 4.2). We did not have complete 7-day information for all patients (Appendix table 4). That is, the data was not balanced. Among 289 patients, 156 (54%) died during hospitalization after receiving CRRT for at least a week. The median patients was 57 years (IQR: 46-65). There were 62.3% male and 90% of the patients were white.

4.4.2 Unadjusted Association with Primary Outcome

During hospitalization, 56% of patients (268) died in the train data. Based on the train data, table 4.2 demonstrates an unadjusted relationship between hospital mortality and clinical and demographic factors. We applied t-test and found no evidence of association between continuous adjusted discrepancy at day 1 and hospital mortality. Similarly, there were no significant relationship with mortality based on baseline

SCr and eGFR, FO% at ICU admission, Charlson score and SOFA score at ICU admission (P-value > 0.05).

Table 4.2: Patient's Characteristics in Train Data by Hospital Mortality

Variable	Alive: N=208	Dead: N=268	P-value
Age	56 (47 - 62)	59 (48 - 66)	0.06
Weight (Kg)	94.35 (80.35 - 110)	96.05 (76.25 - 117.15)	0.64
Sex (Male)	121 (58 %)	168 (63 %)	0.37
Race (White)	193 (93 %)	233 (87 %)	0.06
Baseline Serum Creatinine, SCr (mg/dL)	1.68 (1.09 - 3.47)	1.45 (0.97 - 3.12)	0.67
Baseline eGFR (ml/min/$1.73m^2$)	41.41 (15.82 - 70.57)	53.99 (17.08 - 83.51)	0.27
ICU FO% (L)	2.97 (0.41 - 7.89)	3.08 (0.16 - 9.68)	0.35
Charlson Score	5 (2 - 7)	5 (3 - 8)	0.32
SOFA Score at ICU Start	11.5 (9 - 13)	11 (9 - 14.75)	0.74
SOFA Score at CRRT Start	13 (10 - 15)	14 (12 - 16)	0.01
ICU Type (Cardio)	68 (33 %)	85 (32 %)	
ICU Type (MICU)	107 (51 %)	160 (60 %)	0.07
ICU Type (Surgery)	29 (14 %)	19 (7 %)	
ESRD	164 (79 %)	232 (87 %)	0.04
FO% at CRRT Day 1 (L)	-1.23 (-2.32 - 0.23)	0.42 (-1.15 - 2.1)	<0.01
Adjusted Discrepancy at Day 1	0.01 (-0.01 - 0.05)	0.05 (0.01 - 0.1)	0.24

*Categorical variable: N(%), Continuous variable: Median (IQR)

Even though SOFA score and FO at ICU admission did not show a significant relationship, their values at the beginning of CRRT showed evidence of an unadjusted association with mortality. Compared to ICU admission, fluid overload declines by CRRT initiation. The median FO% at CRRT initiation was -1.23. This indicates less fluid in the body. On the other hand, patients who died during hospitalization, had a positive (0.42) median FO%, which suggests they had excess fluid in the body. ESRD status was also associated with mortality and was present in the majority of the patients who died.

4.4.3 Adjusted Discrepancy and Hospital Mortality

We decided to examine the adjusted discrepancy as a categorical variable since there was no indication of a link between continuous adjusted daily discrepancy and mortality. Also based on figure 4.1 it seems reasonable to explore the adjusted discrepancy as a categorical variable.

Initially, we looked for heterogeneity using data from 656 patients' adjusting discrepancies on day 1 by applying the nonparametric d test proposed in Chapter 2. The non parametric d test calculated a test statistic value of 0.26, which is much higher than the critical values for sample size 600 (5% level of significance: 0.038) and 800 (5% level of significance: 0.033) represented in table 2.1). Thus, we can reject the null hypothesis and conclude that there exists heterogeneity in the adjusted discrepancy at day 1. We also applied bootstrap LRT and obtained evidence for the presence of either 3 or 4 components in the data [13]

At first we used K-means clustering with three components, but the K-means plot prompted us to utilize four components [105]. We implemented the 4 components K-means algorithm, which identified 4-clusters among 656 patients (appendix figure 10). We generated and compared 95% and 68.3% confidence interval for the mean adjusted discrepancy based on each cluster and identified criteria for four groups: $< -8.5\%, > -8.5\% - 5.5\%, > 5.5\% - 20\%$ & $> 20\%$.

The EM algorithm was then used to identify another group of clusters using the k-mean values as initial values [10]. We identified another 4 groups in the adjusted discrepancy at day 1, and the distribution of each group is represented in appendix figure 11. After exploring 95% and 68.3% confidence interval for the mean adjusted discrepancy based on the identified groups, we developed another criteria to categorize adjusted discrepancy on day 1: $< -6\%, > -6\% - 3\%, > 3\% - 40\%$ & $> 40\%$.

Manual identification of groups (3 groups: $< -5\%$, $> -5\% - 5\%$, $> 5\%$; 4 groups: $< -10\%$, $> -10\% - 0\%$, $> 0\% - 10\%$, $> 10\%$) was also explored based on the adjusted discrepancy at day 1 in order to find an association with mortality. After identifying different categorical versions of the adjusted discrepancy, we investigated the relationship with hospital mortality. Figure 4.3 depicts the distribution of dead and alive patients during hospitalization for each selected categorical version of adjusted discrepancy at day 1.
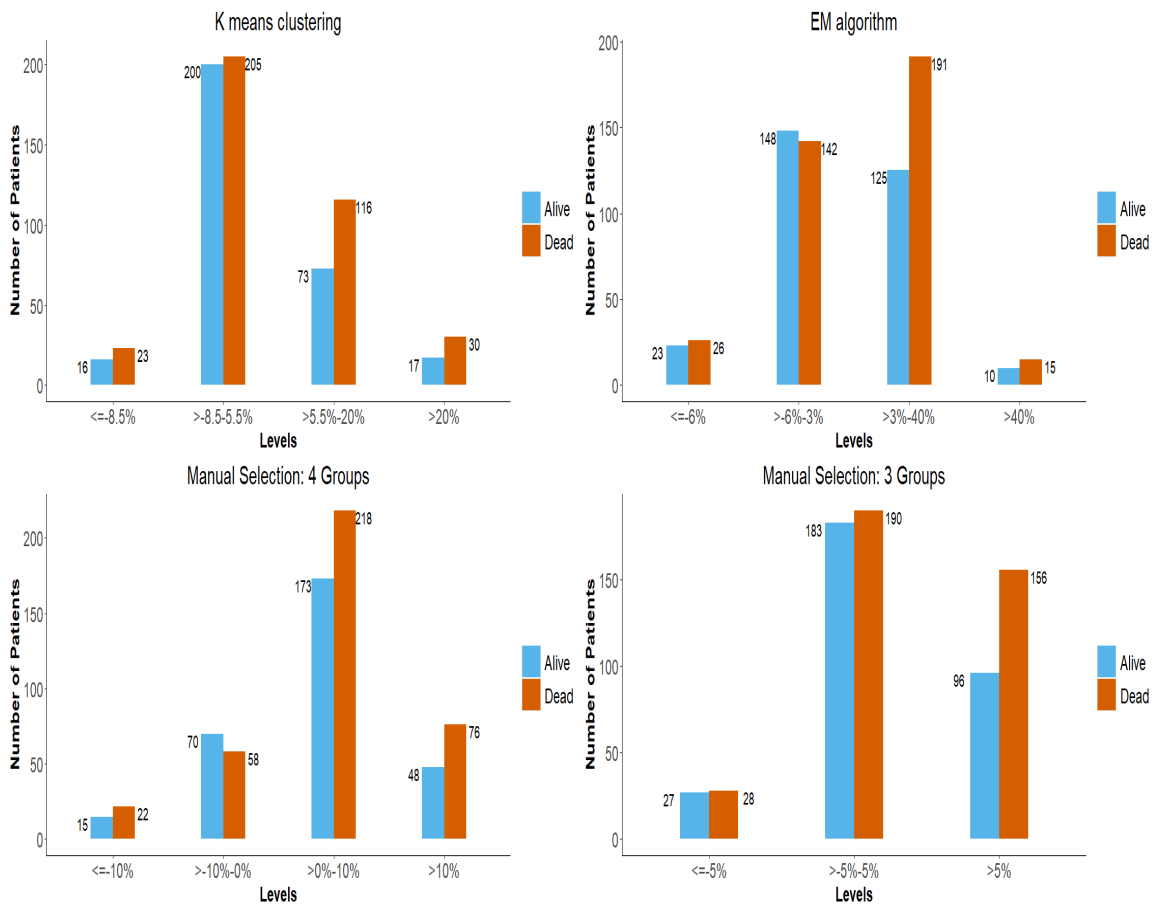

Figure 4.3: Distribution of Dead and Alive Patients for Different Categorical Adjusted Discrepancy at Day 1

Overall, different levels based on different categorical versions of adjusted discrepancy have a higher number of dead patients than alive patients. We then applied Chi-square test based on train data to check for an unadjusted association with hospital

mortality, but no significant results were found for any of the categorical versions of adjusted disrcrepancy. The unadjusted oddsratio are presented in table **??** which does not show any evidence (p-value > 0.05; 95% CI includes 1) of association with mortality. In fact it is noticeable that the 95% CIs are too wide and away from 1 for any version of the adjusted discrepancy variable. This gives us a hint about the uncertainty of the variable information.

Table 4.3: Unadjusted Odds Ratio and 95% Confidence Interval

Adjusted Discrepancy at Day 1	Level	OR (95% CI)	P-value
K means (Ref: >-8.55-5.5%)	<-8.5%	1.25 (0.551-2.97)	0.60
	>5.5%-20%	1.21 (0.797-1.84)	0.38
	>20%	1.18 (0.55-2.62)	0.67
EM (Ref: >-6%-3%)	<-6%	1.03 (0.48-2.23)	0.94
	>3%-40%	1.19 (0.81-1.75)	0.37
	>40%	1.2 (0.45-3.45)	0.71
Manual (Ref: >-5%-5%)	<-5%	0.99 (0.49-2.03)	0.98
	>5%	1.22 (0.83-1.8)	0.31
Manual (Ref: >-10%-0%)	<-10%	1.11 (0.45-2.84)	0.82
	>0%-10%	0.96 (0.59-1.56)	0.88
	>10%	1.28 (0.70-2.38)	0.43

4.4.4   Hospital Mortality Prediction

We used a logistic regression to predict hospital mortality after 48 hours of CRRT initiation with main effect of adjusted discrepancy at day 1, age, race, SOFA score at CRRT start, actual fluid removed at day 1, FO% at day 1, ESRD status and an interaction effect of ICU type and time between ICU admission to CRRT initiation. Adjusted discrepancy was explored both as a continuous and categorical variable. All versions of categorical and continuous adjusted discrepancy yielded similar AIC values (Continuous: 607.21, 3 groups: 608.83, 4 groups:609.03, K-means: 609.02, EM: 609.68)) in the model. Multiple levels of categorical variables contribute to the slight increase in AIC values, so we can assume that the AICs are not much different among the models (95% CIs are overlapping). Since we are interested in hospital mortality

prediction, we performed 3-fold cross validation to compare the accuracy of hospital mortality prediction presented in figure 4.4. Based on the highest prediction accuracy and clinical relevance, we selected the categorical adjusted discrepancy based on K-means clustering to include in the final model.
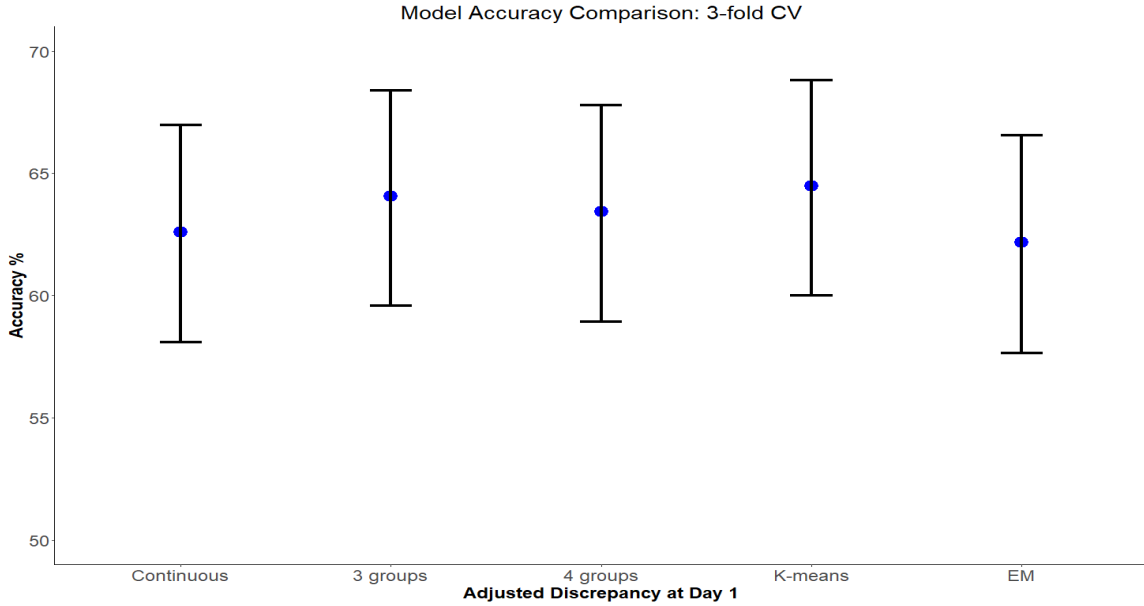
Figure 4.4: Prediction Accuracy Comparison: 3-fold CV

The variance inflation factors (VIFs) were calculated to check for multicollinearity, and none of the main effects had a VIF value more than 3.5. The interaction effect and related main effect's VIF value were close to 17, which is safe to disregard for testing multicollinearity because of its association [113, 114]. Hence, we can say that the final model showed no effect of multicollinearity and the estimated coefficients are reliable (appendix table 5).

The final regression model yielded significant evidence of association among mortality and some important predictors showed in table 4.4. The odds ratio with 95% CI indicates that actual fluid (L) removed by the machine, FO% (L), age, SOFA score at CRRT start, having end stage renal disease (ESRD) and interaction effect between ICU to CRRT duration and MICU showed significant (p-value< 0.05) contribution

in predicting hospital mortality.

Table 4.4: Multivariate Logistic Regression Model

Variables	Level	OR (95% CI)	P-value
Adjusted discrepancy	<-8.5%	1.34 (0.54- 3.46)	0.54
%ml/Kg h	>5.5%-20%	0.81 (0.43-1.53)	0.52
(Ref: -8.55-5.5%)	>20%	0.44 (0.15-1.27)	0.13
Day 1 actual fluid removed (l)		1.19 (1.03-1.39)	0.02
Day 1 FO%(l)		1.26 (1.09-1.46)	<0.01
Age		1.03 (1.02-1.05)	<0.01
Race (Ref:Non-white)	White	0.48 (0.23-0.96)	0.04
SOFA at CRRT initiation		1.14 (1.07-1.22)	<0.01
	MICU	0.77 (0.44-1.33)	0.35
ICU type (Ref: Cardio)	Surgery	0.27 (0.09-0.74)	0.01
	Other	2.83 (0.13-310.4)	0.56
ICU to CRRT duration(day)		1.00 (0.96-1.03)	0.73
ESRD (Ref: No)	Yes	2.11 (1.23-3.63)	<0.01
	MICU	1.15 (1.05-1.28)	<0.01
ICU start to CRRT duration (Ref: Cardio)	Surgery	1.07 (0.97-1.22)	0.27
	Other	0.87 (0.55-1.18)	0.43

Even with the regression model, adjusted discrepancy at day 1 did not appear to make a significant contribution in predicting hospital mortality after 48 hours. Table 4.4 shows that the 95% CIs are too wide and still far away from 1. However, the odds ratio changed for positive discrepancy groups ($> 5.5\% - 20\%$: 0.81, $> 20\%$: 0.44). This indicates that, there might be some variable in the model which has a confounding effect on the adjusted discrepancy. We investigated this by fitting a linear mixed effect model to the daily adjusted discrepancy using daily fluid and demographic information.

Table 4.4 demonstrates that patients with ESRD history has 111%(2.11-1) higher odds of dying during hospitalization compared to patients who do not have ESRD history holding other information fixed at a certain value. Similarly, with an year increase in age, the odds of dying during hospitalization increases by 3%. White patients have lower odds of dying by 52% (1-0.48) compared to non-white patients.

With the one point increase of SOFA score at CRRT onset, the odds of dying increases by 14%. With the increase of 1 L in FO%, the odds of dying increases by 26%. With 1 L change in actual fluid removed odds of dying increases by 19%.

We identified a significant interaction effect between time from ICU admission to CRRT duration (days) and ICU type MICU (appendix table 5). Among MICU patients, with one day increase in ICU to CRRT duration the odds of dying is 12.2% (1-exp(0.14-0.01-0.26)) lower compared to patients who are in Cardio ICU. Similarly with 1 day increase in ICU admission to CRRT initiation, patients in Surgery have 71.6%(1-exp(0.07-0.01-1.32)) lower odds of dying compared to patients in Cardio. But the interaction effect failed to prove significant among surgery ICU patients(appendix table 5). This information indicates patients in the Cardio ICU may be very sick compared to patients in Surgery or MICU, and are at higher risk of hospital mortality or implementation of earlier CRRT is more effective among the other ICU types compared to Cardio.

4.4.5   Prediction Performance on Test Data

After the selection of the final model based on train data, we investigated the prediction performance by applying the model to test data. An optimal predicted probability cutoff of 0.513 was identified based on maximum Youden's index (function of sensitivity and specificity) [115, 116]. Using the optimal predicted probability cutoff, we compared the prediction performance in both train and test data for model validation.

From table 4.5, the model obtained a 68% prediction accuracy on train data and a 62% accuracy on test data. The sensitivity in the test data was 62% and the specificity was 63%. Positive predictive value on the test data was 73%. Although the negative predictive value were somewhat lower in both the test and train data,

Table 4.5: Prediction Performance Comparison

Metrics	Train Data	Test Data
Accuracy (95% CI)	0.68 (0.63-0.72)	0.62 (0.55-0.69)
Kappa	0.34	0.24
Sensitivity	0.70	0.62
Specificity	0.65	0.63
Positive Predictive Value	0.75	0.73
Negative Predictive Value	0.58	0.51

higher positive predictive value indicates a good sign, because we are interested in correctly predicting hospital mortality to provide more attention to this high-risk patients. Hosmer-Lemeshow goodness of fit test was performed to check for the logistic regression fit and did not show any evidence of lack of fit for the train data (P-value 0.883), but it was marginally significant when the model was fitted on test data (P-value 0.0501). That means, for test data, the model may not be a good fit, but we do not have much evidence. This can also arise from the re-estimation of the odds ratio from the test data.
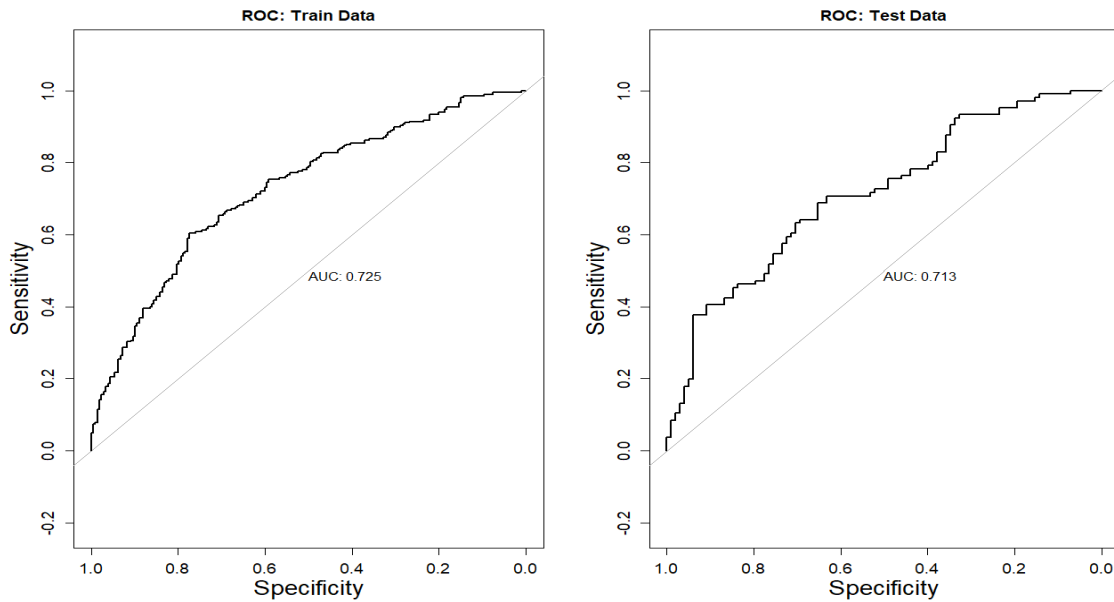

Figure 4.5: ROC Comparison

We generated the receiver operating curve (ROC) and area under the curve (AUC) depicted in figure 4.5 to examine the model further. The AUC values for both train and test data are slightly higher than 0.7, indicating that the model is capable of discriminating the categorical outcome variable [117]. That is, both in test and train data, the model is capable of differentiating between dead and alive prediction to an acceptable degree.

4.4.6 Secondary Outcome

Daily adjusted discrepancy (%ml/kg h) from day 1 to day 6 were used to fit a linear mixed effect model. In figure 13, we initially investigated the relationship between time and adjusted discrepancy by fitting a lowess smooth curve using R package 'ggplot2' [118, 119]. Based on the initial investigation, we fitted a linear mixed effect model with day as quadratic effect. We also compared the initial model with piecewise linear at day (4) but the quadratic effect model produced lower AIC, BIC and higher log-likelihood value. Thus, we fitted a linear mixed effect model with both fixed and random effect for quadratic time effect and random slope and intercept for each patient to see if the adjusted discrepancy was confounded with any other predictors obtained in the prediction model and to account for subject-specific variations. We also explored random effect for the quadratic term in day.

For main effects, we explored ESRD, SOFA score at CRRT, FO%(L) at ICU admission to account for patients' sickness, actual removed fluid(L) by machine and time from ICU to CRRT initiation for treatment effect, and age, sex, race, ICU type to quantify for demographic information. Model selection was based on AIC, BIC and LRT. We also checked for interaction over time with actual fluid removed but did not find any significant evidence. For the final model, we selected the main effects of day, actual fluid removed, sex, race, esrd and FO% at ICU admission and random group effect of ESRD. That is different unstructured covariance structure will be fitted based

on whether the patients have ESRD or not. Table 4.6 shows that only actual fluid removed from the machine is significantly associated with adjusted fluid discrepancy over time.

Several covariance stuctures like toeplitz, gaussian error, exponential decay were explored, but based on AIC, BIC and loglikehilood value, unstructred covariance structure with group effect of ESRD produce optimal result within given information. We also used empirical standard error. Because empirical standard errors are robust and will produce consistent results even thought the covariance structures are not correctly identified.

Table 4.6: Daily Adjusted Discrepancy Association

Effect	Level	Estimate(95% CI)	Standard Error	P-value
Intercept		15.97 (11.83-20.12)	2.06	<0.001
Day		-0.72 (-1.99-0.55)	0.65	0.26
Day^2		0.08 (-0.10-0.26)	0.10	0.37
Actual fluid removed (L)		-2.58 (-3.02–2.15)	0.22	<0.001
Sex (Ref: Male)	Female	0.52 (-0.69-1.73)	0.62	0.40
Race (Ref: Non-White)	White	-1.70 (-4.09-0.70)	1.22	0.16
ESRD (Ref: No)	Yes	-0.41 (-2.23-1.40)	0.92	0.65
FO% at ICU admission (L)		-0.03 (-0.09-0.03)	0.03	0.32

Table4.6 shows that, daily discrpancy is significantly related to actual fluid removed adjusted for FO% at ICU admission, ESRD, race and sex. The estimated coefficents say that, with 1L increase in actual fluid removed the change in daily adjusted discrepancy over one day increase will be -2.58%ml/kgh, fixing other factors at a certain value. Even though we did not find evidence of the association of other factors with adjusted discrepancy, we used this model to understand the effect of treatment implementation adjusting for patients' sickness. Also, while looking back at the hospital mortality prediction model results in table 4.4, the odds ratio for adjusted discrepancy changed for the positive discrepancy group compared to the odds ratio in unadjusted

association with mortality depicted in table 4.3. Thus, it is safe to conclude that actual fluid removed has a confounding effect on the adjusted discrepancy in association with hospital mortality.

4.5   Discussion

In this study, we were unable to obtain evidence of a significant contribution of the discrepancy in treatment towards increased hospital mortality risk among patients requiring CRRT. The wide 95% confidence interval indicates that the variability in the discrepancy is very high. We had a lot of patients in both dead and alive group at any discrepancy level. The exploratory linear mixed effect model was fitted to understand whether any factor in the mortality prediction model is confounded with the adjusted discrepancy in predicting mortality. We found that the actual fluid removed from the machine is significantly associated with the daily discrepancy changes. We did not have enough patients on the 'Other' ICU type, thus the the 95% CI of the odds ratio represented in table 4.4 is not meaningful.

Our study is not without limitations. Because the data came from observational research, there might be other confounding factors that weren't taken into account. We were not able to use the first 24 hours of CRRT information due to the CRRT goal not being informative. Future research can overcome this by using an adjusted daily CRRT goal and calculating the discrepancy accordingly. This study was also a single center study with very low geographical diversity. Mechanical ventilation, ecmo, vad, and iabp, which have been relevant in other studies, were not able to be used since for majority of the patients the period of interest did not overlap. Future research can focus on the CRRT duration and predict mortality after the completion of CRRT. Also, because the discrepancy variable is a modified variable of fluid overload, we couldn't utilize it in our secondary analysis. We used ICU FO% as a fixed value but did not find any evidence of association with daily discrepancy change during CRRT.

69

In this study it is evident that using adjusted discrepancy on day 1 is not enough to predict mortality, and future work should focus on adjusted discrepancy during the whole period of CRRT to understand how they are associated with mortality. Along with whole CRRT duration information, it will be more informative to obtain the clinical factors that caused the discrepancy. This also presses the need for a multicenter study. We did not have enough geographical diversity in our data. Thus, we cannot generalize the results to a broader population.

It is evident from this single-center study that we need to use more days of data to account for all factors influencing their treatment in order to have a clearer picture of the connection to hospital mortality. We did obtain the importance of fluid overload in relation to hospital mortality. We have also found that fluid overload is correlated with discrepancy (correlation coefficient: 0.49). It is clear that, more expertise is needed to understand whether fluid overload and discrepancy should be studied together or separately.

In the future, we'll concentrate on data collected throughout the CRRT duration and how it relates to post-CRRT mortality. We will also explore how daily prescribed goal is related to daily fluid overload, which might provide a clearer picture of the treatment's impact on patients' hospital mortality risk.

Fluid discrepancy is a novel variable in measuring treatment performance and is definitely related to fluid removal. The fact that patients in the ICU are frequently unable to remove fluid as directed is alarming. It may not be solely due to treatment intolerance. Other factors like the different drug use, other therapies may contribute also. Thus, more exploration is needed to determine how this disparity relates to mortality risk in ICU-admitted CRRT patients.

Chapter 5 Final Thoughts & Conclusion

5.1   Heterogeneity Detection & Nonparametric D test

While parametric tests for heterogeneity has been developed based on mixture models, choice of the null hypothesis has always been a recent issue. The nonparametric D test overcomes this limitation by using an empirical null distribution based on the data [58]. Even though the theoretical evidence for the asymptotic properties are still under construction, by using bootstrap and power calculation, the test has proven to be successful in detecting heterogeneity in a lot of cases.

Although in some circumstances, the parametric test will have greater power than the nonparametric test, specially when the null and alternative hypothesis are defined accurately. But the unique feature of detecting different types of heterogeneity is sometimes preferable. Also the choice of wrong null hypothesis will lead to wrong decision making. In large scale hypothesis testing this will increase the type II error. With the original d-test we risk the possibility of choosing a wrong null hypothesis but not with non parametric d test [16]. While the nonparametric version has less power hypothetically, it has a greater probability of avoiding the consequences of a false null hypothesis.

With the limitation of failing to detect heterogeneity in a small sample size, we overcome the problem by proposing a weighted version of the test. We have also provided the option of choosing one's own bandwidth without worrying about the change. Other weight functions will be investigated in future studies to test for power improvement.

We compared our test findings to bootstrap LRT regarding time to detect hetero-

geneity and power analysis for two component contaminated mixture only and were pleased with the results. While the bootstrap LRT takes longer to compute than the nonparametric d-test, it gives an idea of how many components to test. We will try to include the notion of determining the number of components in data with heterogeneity as another future path of nonparamteric d test improvement.

5.2 Fluid Assessment among CRRT ICU Patients

Identifying clinical parameters associated with an increased mortality risk has been a common issue among CRRT patients admitted to the ICU [120–122]. The systematic literature search concluded that more information is needed in regards to understanding how the prescribed dose is associated with mortality risk. This dissertation provided more knowledge not explored in the previous literature by examining if treatment implementation discrepancy is also a predictor of increased mortality risk.

To understand the discrepancy, we limited our analysis to 48 hours of CRRT data in order to predict hospital mortality after 48 hours. We kept the time window similar to ongoing AKI research [123–125]. Using only 48 hours of data, our research failed to find evidence of a link between disparity in therapy application and mortality. This suggests that we should concentrate our efforts on a broader time frame in order to fully comprehend the relationship for future research keeping in ming that the findings might also be affected by the sample size and power of the tests. We also identified other risk factors that has not been explored before such as, the actual fluid removed by the machine. Although failing to establish a rationale for prescription disagreement with an ambitious time-frame window, we uncovered that fluid removal within 48 hours does contribute in hospital mortality after 48 hours.

Fluid overload has been studied as one of the most important characteristics among

patients at the risk of mortality undergoing CRRT [44,126]. Our research also showed similar findings. We also found fluid overload, ICU type age race to have significant association with mortality as mentioned from the previous literature reviews in chapter 3. The relationship between the prescribed goal and fluid overload should be a future direction in understanding how they are related to mortality risk using more than 48 hours of data. We were unable to obtain evidence for some other valuable pieces of information like mechanical ventilation, use of ECMO, VAD or IABP, because vast majority of patients recorded these information after our study period of interest. For 28 patients, these information was available after the end of the CRRT duration. Thus, increasing the timeline window will allow the researchers to incorporate more relevant information in their mortality prediction.

Due to the severity of their illness, it is desirable to have a prediction model that will assist ICU patients with CRRT in making necessary modifications in a short period of time. However, due to the nature of their illness, utilizing a limited time window forecast is ineffective. Rather, we should focus on discovering major risk factors that are associated with increased mortality risk and altering the patient's treatment accordingly.

While conducting this research, we also faced a time constraint due to data availability, which limited our capacity to explore additional choices to have a better knowledge of the fluid removal assessment. We did not have data on patients' critical information regarding sepsis, diabetes, hypertension, and so on. However, with limited information, we were able to gather several critical findings and future directions that will help us better understand fluid evaluation and contribute to patient healthcare improvement.

5.3 Collaborative Approach

In order to gain relevant insights and progress in science, collaborative effort is a crucial instrument in research [127–129]. Chapters 2, 3, and 4 of this dissertation rely on the multi-domain knowledge of clinical nephrologists, epidemiologists, and statisticians. Clinical and epidemiological experts offered domain-specific expertise about the heterogeneity detection method's application in Chapter 2. The inclusion and exclusion criteria for publication selection in Chapter 3 were based on clinical competence in the field of nephrology and librarian search strategies. Clinical expertise was used to determine disagreement in chapter 4, which was then examined using epidemiological and statistical insights.

This research would not have been possible without the collaboration of domain experts. In the world of medical science, better healthcare cannot be provided just on the basis of clinician skill without proof of therapy. A statistician cannot provide reasoning without first having clinical and epidemiological understanding of the disease. As a result, the researchers can only deliver better and more precise healthcare solutions if they use a team scientific strategy.

Appendices

Appendix A: Theoretical details

1. Illustration of nonparametric d-statistic

D test statistic proof:

$$d = \int \left[\hat{f}_h(z) - f(z|\hat{\mu}_0, \hat{\sigma}_0) \right]^2 dz$$

$$= \int \left[\frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-z_i}{h}\right)^2} - \frac{1}{\sqrt{2\pi}\hat{\sigma}_0^2} e^{-\frac{1}{2\hat{\sigma}_0^2}(z-\hat{\mu}_0)^2} \right]^2 dz$$

Let,

$$m_i = \frac{1}{nh}; \quad m_0 = -\frac{1}{\hat{\sigma}_0}; \quad h_i = h$$

$$h_0 = \hat{\sigma}_0; \quad z_i = z_i \quad and \quad z_0 = \hat{\mu}_0$$

Then,

$$d = \int \Big[\sum_{i=1}^{n} m_i \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2h_i^2}(z-z_i)^2} \quad + \frac{m_0}{\sqrt{2\pi}} e^{-\frac{1}{2h_0^2}(z-z_0)^2} \Big]^2 dz$$

$$= \int \Big[\sum_{i=0}^{n} \frac{m_i}{\sqrt{2\pi}} e^{-\frac{1}{2h_i^2}(z-z_i)^2} \Big]^2 dz$$

$$= \int \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} e^{-\frac{1}{2h_i^2}(z-z_i)^2 - \frac{1}{2h_j^2}(z-z_j)^2} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} \int e^{-\frac{1}{2h_i^2}(z-z_i)^2 - \frac{1}{2h_j^2}(z-z_j)^2} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} \int e^{-\frac{1}{2h_i^2 h_j^2}\left[h_j^2(z-z_i)^2 + h_i^2(z-z_j)^2 \right]} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} \int e^{-\frac{1}{2h_i^2 h_j^2}\left[z^2(h_j^2+h_i^2) - 2z(h_j^2 z_i + h_i^2 z_j) + h_j^2 z_i^2 + h_i^2 z_j^2 \right]} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} \int e^{-\frac{(h_j^2+h_i^2)}{2h_i^2 h_j^2}\left[z^2 - 2z\frac{(h_j^2 z_i + h_i^2 z_j)}{(h_j^2+h_i^2)} + \frac{h_j^2 z_i^2 + h_i^2 z_j^2}{(h_j^2+h_i^2)} \right]} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} \int e^{-\frac{(h_j^2+h_i^2)}{2h_i^2 h_j^2}\left[z^2 - 2z\frac{(h_j^2 z_i + h_i^2 z_j)}{(h_j^2+h_i^2)} + \frac{h_j^2 z_i^2 + h_i^2 z_j^2}{(h_j^2+h_i^2)} + \frac{(h_j^2 z_i + h_i^2 z_j)^2}{(h_j^2+h_i^2)^2} - \frac{(h_j^2 z_i + h_i^2 z_j)^2}{(h_j^2+h_i^2)^2} \right]} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} \frac{\sqrt{2\pi \frac{h_i^2 h_j^2}{h_i^2+h_j^2}}}{\sqrt{2\pi \frac{h_i^2 h_j^2}{h_i^2+h_j^2}}}$$

$$\times e^{-\frac{(h_i^2+h_j^2)}{2h_i^2 h_j^2}\frac{1}{(h_i^2+h_j^2)}\left[(h_j^2 z_i^2 + h_i^2 z_j^2) - \frac{(h_j^2 z_i + h_i^2 z_j)^2}{(h_i^2+h_j^2)} \right]}$$

$$\times \int e^{-\frac{1}{2\frac{h_i^2 h_j^2}{h_i^2+h_j^2}}(z - \frac{h_j^2 z_i + h_i^2 z_j}{h_j^2+h_i^2})^2} dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{\sqrt{2\pi}} \sqrt{\frac{h_i^2 h_j^2}{h_i^2 + h_j^2}} e^{-\frac{1}{2h_i^2 h_j^2}\left[(h_j^2 z_i^2 + h_i^2 z_j^2) - \frac{(h_i^2 z_i + h_i^2 z_j)^2}{h_i^2+h_j^2} \right]} \times 1$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{\sqrt{2\pi}} \sqrt{\frac{h_i^2 h_j^2}{h_i^2 + h_j^2}} e^{-\frac{1}{2h_i^2 h_j^2}\left[(h_j^2 z_i^2 + h_i^2 z_j^2) - \frac{(h_i^2 z_i + h_i^2 z_j)^2}{h_i^2+h_j^2} \right]}$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{\sqrt{2\pi}} \sqrt{\frac{h_i^2 h_j^2}{h_i^2 + h_j^2}} e^{-\frac{1}{2h_i^2 h_j^2}\left[(h_j^2 z_i^2 + h_i^2 z_j^2) - \frac{(h_i^2 z_i + h_i^2 z_j)^2}{h_i^2+h_j^2} \right]}$$

$$+ \sum_{i=1, j=0} m_i m_0 \sqrt{\frac{h_i^2 h_0^2}{h_i^2 + h_0^2}} e^{-\frac{1}{2h_i^2 h_0^2}\left[(h_0^2 z_i^2 + h_i^2 z_0^2) - \frac{(h_0^2 z_i + h_i^2 z_0)^2}{h_i^2+h_0^2} \right]}$$

$$+ \sum_{j=1, i=0} m_0 m_j \sqrt{\frac{h_0^2 h_j^2}{h_0^2 + h_j^2}} e^{-\frac{1}{2h_0^2 h_j^2}\left[(h_j^2 z_0^2 + h_0^2 z_j^2) - \frac{(h_j^2 z_0 + h_0^2 z_j)^2}{h_0^2+h_j^2} \right]}$$

76

Substituting the values,

$$A = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\frac{1}{n^2h^2}}{\sqrt{2\pi}} \sqrt{\frac{h^4}{2h^2}} e^{-\frac{1}{2h^4}\left[h^2(z_i^2+z_j^2)-\frac{h^4(z_i+z_j)^2}{2h^2}\right]}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2n^2h\sqrt{\pi}} e^{-\frac{1}{2h^2}\left[(z_i^2+z_j^2)-\frac{z_i^2+z_j^2+2z_iz_j}{2}\right]}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2n^2h\sqrt{\pi}} e^{-\frac{1}{4h^2}(z_i-z_j)^2}$$

$$B = \sum_{i=1,j=0} \frac{\left(\frac{1}{nh}\right)\left(\frac{-1}{\hat{\sigma}_0}\right)}{\sqrt{2\pi}} \sqrt{\frac{h^2\hat{\sigma}_0^4}{h^2+\hat{\sigma}_0^2}} e^{-\frac{1}{2h^2\hat{\sigma}_0^2}\left[(\hat{\sigma}_0^2z_i^2+h^2\hat{\mu}_0^2)-\frac{(\hat{\sigma}_0^2z_i+h^2\hat{\mu}_0)^2}{h^2+\hat{\sigma}_0^2}\right]}$$

$$= -\sum_{i=1,j=0} \frac{1}{n\sqrt{2\pi(h^2+\hat{\sigma}_0^2)}} e^{-\frac{h^2\hat{\sigma}_0^2}{2h^2\hat{\sigma}_0^2}\left[\frac{z_i^2+\hat{\mu}_0^2-2z_i\hat{\mu}_0}{h^2+\hat{\sigma}_0^2}\right]}$$

$$= -\sum_{i=1,j=0} \frac{1}{n\sqrt{2\pi(h^2+\hat{\sigma}_0^2)}} e^{-\frac{1}{2(h^2+\hat{\sigma}_0^2)}(z_i-\hat{\mu}_0)^2}$$

$$= C \quad \text{[By the nature of symmetricity]}$$

$$D = \frac{1}{\sqrt{2\pi}\hat{\sigma}_0^2} \sqrt{\frac{\hat{\sigma}_0^2}{2}} e^{-\frac{0}{2\hat{\sigma}_0^4}}$$

$$= \frac{1}{2\hat{\sigma}_0\sqrt{\pi}} \times 1 \quad = \frac{1}{2\hat{\sigma}_0\sqrt{\pi}}$$

Thus,

$$d = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2n^2h\sqrt{\pi}} e^{-\frac{1}{4h^2}(z_i-z_j)^2}$$

$$-2\sum_{i=1} \frac{1}{n\sqrt{2\pi(h^2+\hat{\sigma}_0^2)}} e^{-\frac{1}{2(h^2+\hat{\sigma}_0^2)}(z_i-\hat{\mu}_0)^2}$$

$$+\frac{1}{2\hat{\sigma}_0\sqrt{\pi}}$$

Illustration of nonparametric weighted d statistic

Weighted nonparametric D test statistic proof:

$$d_w = \int \left[\hat{f}_h(z) - f(z|\hat{\mu}_0, \hat{\sigma}_0) \right]^2 w(z) dz$$

$$= \int \left[\frac{1}{nh} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-z_i}{h}\right)^2} - \frac{1}{\sqrt{2\pi}\hat{\sigma}_0^2} e^{-\frac{1}{2\hat{\sigma}_0^2}(z-\hat{\mu}_0)^2} \right]^2 w(z) dz$$

Let,

$$m_i = \frac{1}{nh}; \quad m_0 = -\frac{1}{\hat{\sigma}_0}; \quad h_i = h$$

$$h_0 = \hat{\sigma}_0; \quad z_i = z_i \quad and \quad z_0 = \hat{\mu}_0$$

Then,

$$d = \int \left[\sum_{i=1}^{n} m_i \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2h_i^2}(z-z_i)^2} + \frac{m_0}{\sqrt{2\pi}} e^{-\frac{1}{2h_0^2}(z-z_0)^2} \right]^2 w(z) dz$$

$$= \int \left[\sum_{i=0}^{n} \frac{m_i}{\sqrt{2\pi}} e^{-\frac{1}{2h_i^2}(z-z_i)^2} \right]^2 w(z) dz$$

Let,

$$w(z) = e^{-\frac{c(z-z_0)^2}{h_0^2}}$$

Then,

$$d_w = \int \sum_{i=0}^{n} \sum_{j=0}^{n} m_i m_j \frac{1}{2\pi} e^{-\frac{1}{2h_i^2}(z-z_i)^2 - \frac{1}{2h_j^2}(z-z_j)^2 - \frac{c(z-z_0)^2}{h_0^2}} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} \int e^{-\frac{1}{2h_i^2 h_j^2 h_0^2}[h_j^2 h_0^2(z-z_i)^2 + h_i^2 h_0^2(z-z_j)^2 + 2ch_i^2 h_j^2(z-z_0)^2]} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} \int e^{-\frac{1}{2h_i^2 h_j^2 h_0^2}[z^2(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2) - 2z(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0) + (h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)]} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} \int e^{-\frac{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}{2h_i^2 h_j^2 h_0^2}\left[z^2 - 2z\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)}{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)} + \frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}\right]} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} e^{-\frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{2h_i^2 h_j^2 h_0^2}} \times$$

$$\int e^{-\frac{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}{2h_i^2 h_j^2 h_0^2}\left[z^2 - 2z\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)}{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)} \pm \left(\frac{h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0}{h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2}\right)^2\right]} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} e^{-\frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{2h_i^2 h_j^2 h_0^2}} e^{\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)^2}{2h_i^2 h_j^2 h_0^2(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}} \times$$

$$\int e^{-\frac{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}{2h_i^2 h_j^2 h_0^2}\left[z - \frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)}{(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}\right]^2} \, dz$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j}{2\pi} e^{-\frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{2h_i^2 h_j^2 h_0^2}} e^{\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)^2}{2h_i^2 h_j^2 h_0^2(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}} \times$$

$$\sqrt{2\pi \frac{h_i^2 h_j^2 h_0^2}{h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2}}$$

$$= \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{m_i m_j h_i h_j h_0}{\sqrt{2\pi(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}} e^{\left[\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)^2}{2h_i^2 h_j^2 h_0^2(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)} - \frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{2h_i^2 h_j^2 h_0^2}\right]}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{m_i m_j h_i h_j h_0}{\sqrt{2\pi(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)}} e^{\left[\frac{(h_j^2 h_0^2 z_i + h_i^2 h_0^2 z_j + 2ch_i^2 h_j^2 z_0)^2}{2h_i^2 h_j^2 h_0^2(h_j^2 h_0^2 + h_i^2 h_0^2 + 2ch_i^2 h_j^2)} - \frac{(h_j^2 h_0^2 z_i^2 + h_i^2 h_0^2 z_j^2 + 2ch_i^2 h_j^2 z_0^2)}{2h_i^2 h_j^2 h_0^2}\right]}$$

$$+ 2 \sum_{i=1}^{n} \frac{m_i m_0 h_i h_0^2}{\sqrt{2\pi(h_0^4 + h_i^2 h_0^2 + 2ch_i^2 h_0^2)}} e^{\left[\frac{(h_0^4 z_i + h_i^2 h_0^2 z_0 + 2ch_i^2 h_0^2 z_0)^2}{2h_i^2 h_0^4(h_0^4 + h_i^2 h_0^2 + 2ch_i^2 h_0^2)} - \frac{(h_0^4 z_i^2 + h_i^2 h_0^2 z_0^2 + 2ch_i^2 h_0^2 z_0^2)}{2h_i^2 h_0^4}\right]}$$

$$+ \frac{m_0^2 h_0^3}{sqrt2\pi(2h_0^4 + 2ch_0^4)} e^{\left[\frac{(2h_0^4 z_0 + 2ch_0^4 z_0)^2}{2h_0^6(2h_0^4 + 2ch_0^4)} - \frac{(2h_0^4 z_0^2 + 2ch_0^4 z_0^2)}{2h_0^6}\right]}$$

Substituting the values,

$$d_n w = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\hat{\sigma_0}}{2hn^2\sqrt{\pi(\hat{\sigma_0}^2 + ch^2)}} e^{\left[\frac{(\hat{\sigma_0}^2 z_i + \hat{\sigma_0}^2 z_j + 2ch^2\hat{\mu_0})^2}{4h^2\hat{\sigma_0}^2(\hat{\sigma_0}^2 + ch^2)} - \frac{\hat{\sigma_0}^2 z_i^2 + \hat{\sigma_0}^2 z_j^2 + 2ch^2\hat{\mu_0}^2}{2h^2\hat{\sigma_0}^2}\right]}$$

$$- 2\sum_{i=1}^{n} \frac{1}{n\sqrt{2\pi(\hat{\sigma_0}^2 + h^2 + 2ch^2)}} e^{\left[\frac{(\hat{\sigma_0}^2 z_i + h^2\hat{\mu_0} + 2ch^2\hat{\mu_0})^2}{2h^2\hat{\sigma_0}^2(\hat{\sigma_0}^2 + h^2 + 2ch^2)} - \frac{\hat{\sigma_0}^2 z_i^2 + h^2\hat{\mu_0}^2 + 2ch^2\hat{\mu_0}^2}{2h^2\hat{\sigma_0}^2}\right]}$$

$$+ \frac{1}{\hat{\sigma_0}^2\sqrt{\pi(1 + c)}}$$

Appendix B: Tables

Table 1: Modelling Critical values vs Sample size

Model type	Model	AIC values
Linear model	Log cv= a + Log N	-77.43
Linear model	Log cv = a+ N	-44.52
Exponential decay model	Log (y-θ)=a +bx	3.79

Table 2: Key terms for database search

Population	Exposure	Outcome
Acute kidney failure, Acute renal failure, Acute kidney insufficienc, Acute renal insufficienc, Acute kidney injur, Continuous Renal Replacement Therapy, Hemofi ltration, Hemodiafi ltration, Renal Replacement Therapy, Critical Illness, Critically Ill, Critical Care, ICU, Intensive Care Units, Kidney replacement therapy, Continuous	Ultra filtration, NUF, Fluid Overload, FO, Fluid Management, Fluid Balance, Fluid Removal	Mortality, Hospital Mortality, Survival rate

Table 3: Adjusted Discrepancy Continuous vs Categorical

Adjusted Discrepancy Comparison	AIC	LR Test P-value
With Continuous Adjusted Discrepancy	607.21	<0.01
With Categorical Adjusted Discrepancy (over or under)	606.36	

Table 4: No. Patient's Information Available for Each Day

Day	No. of Patients
1	277
2	274
3	269
4	262
5	254
6	258

Table 5: Logistic Regression Estimates for Hospital Mortality Prediction

| Variables	Levels	Estimate	Std. Error	P-value
Intercept		-3.45	0.82	<0.001
Adjusted discrepancy	<-8.5%	0.29	0.47	0.54
at day 1	>5.5%-20%	-0.21	0.32	0.52
(%ml/kg h)	>20%	-0.83	0.54	0.13
Actual fluid removed(L)		0.18	0.08	0.02
FO% at day 1 (L)		0.23	0.08	<0.001
Age		0.03	0.01	<0.001
Race (Ref: Non-white)	White	-0.73	0.36	0.04
SOFA at CRRT start		0.13	0.03	<0.001
ESRD status (Red: No)	Yes	0.74	0.28	0.01
	MICU	-0.26	0.28	0.35
ICU type (Ref: Cardio)	Other	1.04	1.77	0.56
	Surgery	-1.32	0.53	0.01
Time from ICU admission to CRRT initiation		-0.01	0.02	0.73
ICU type \times Time	MICU	0.14	0.05	<0.001
from ICU admission to	Other	-0.14	0.18	0.43
CRRT initiation (Ref: Cardio)	Surgery	0.07	0.06	0.27

Appendix C: Figures
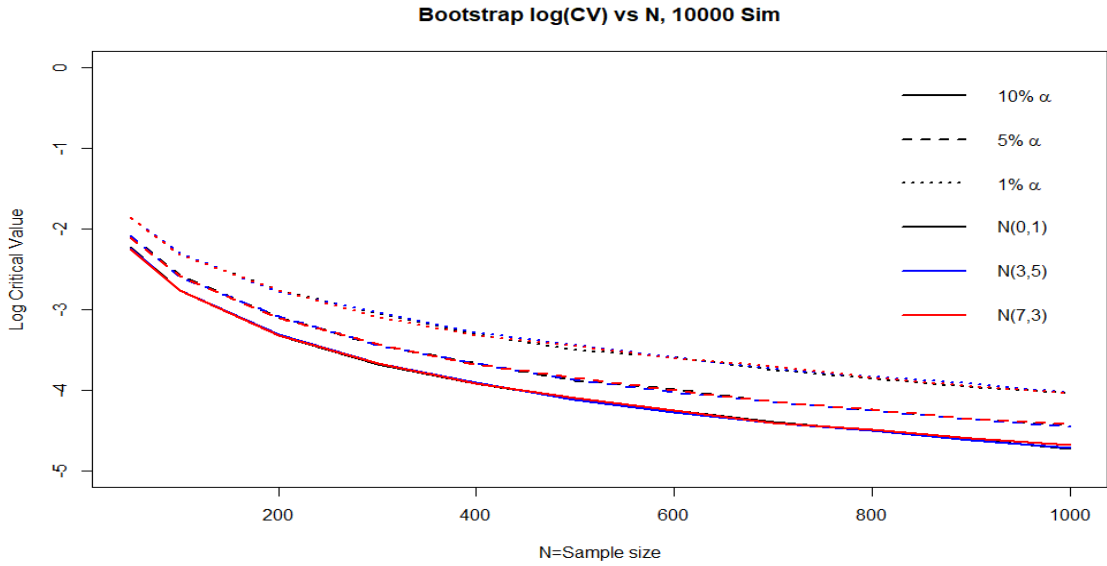
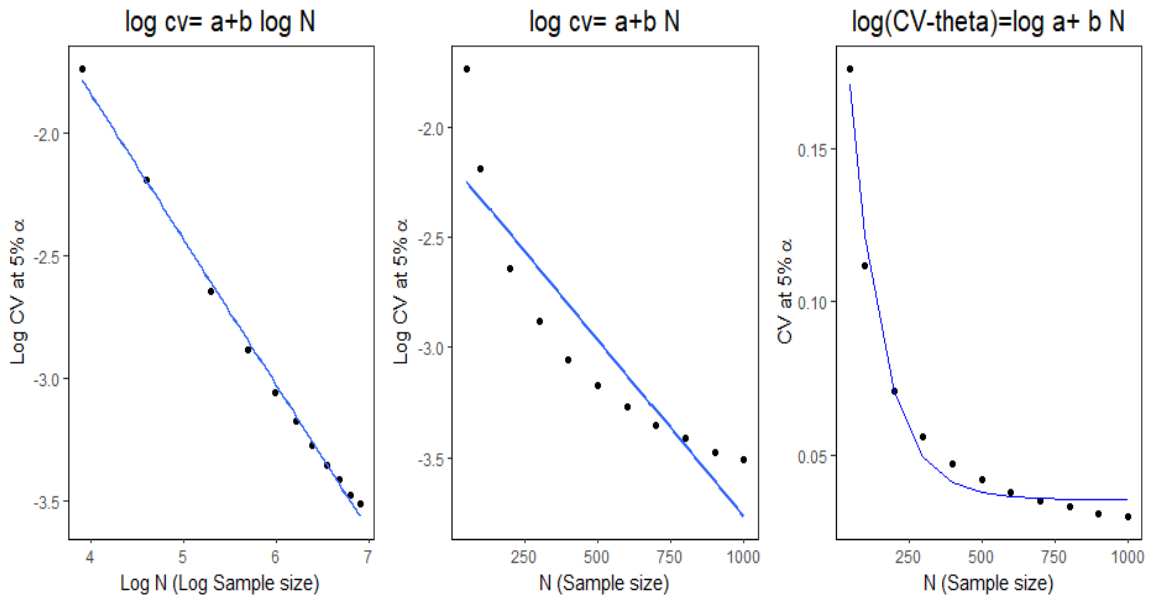Figure 1: Critical Values Bootstrap vs. Fixed

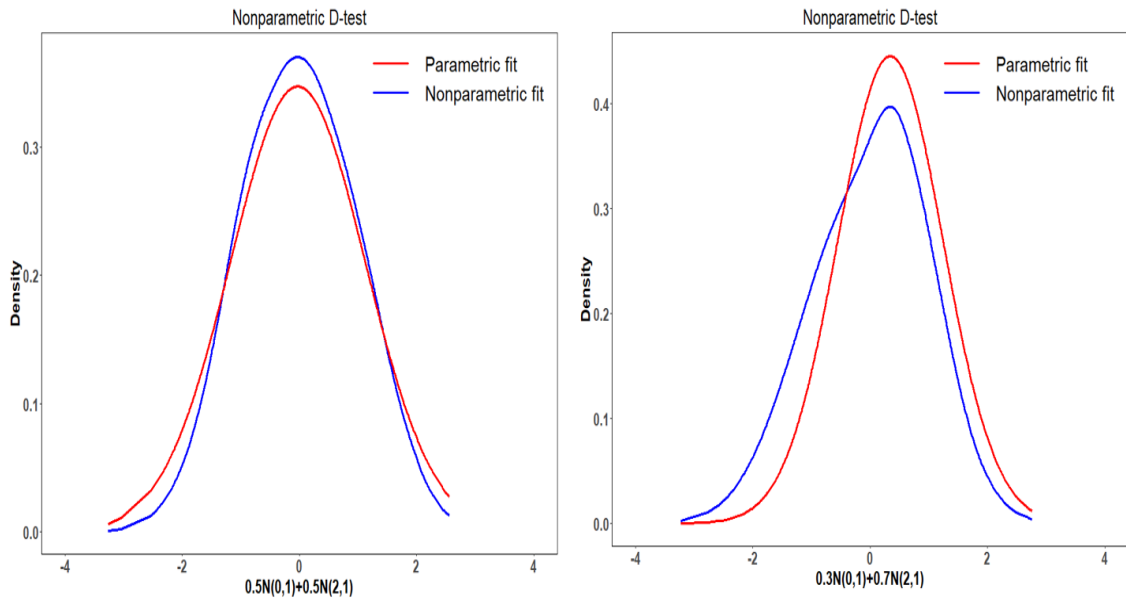Figure 2: Model fit on Critical values vs Sample size
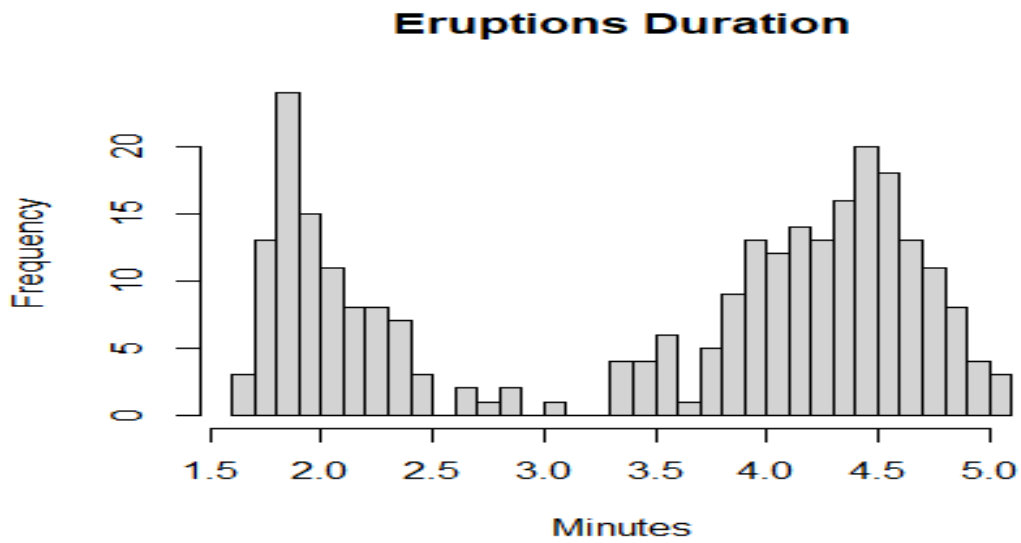
Figure 3: Skewness effect on NpD test

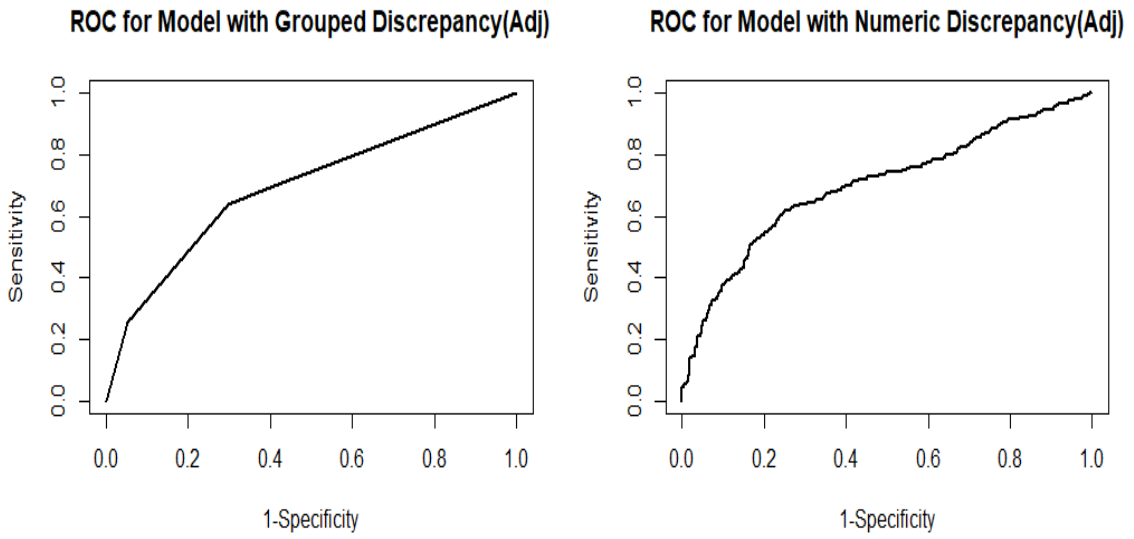Figure 4: Distribution of eruption duration in minutes

Figure 5: ROC comparison for Adjusted Discrepancy
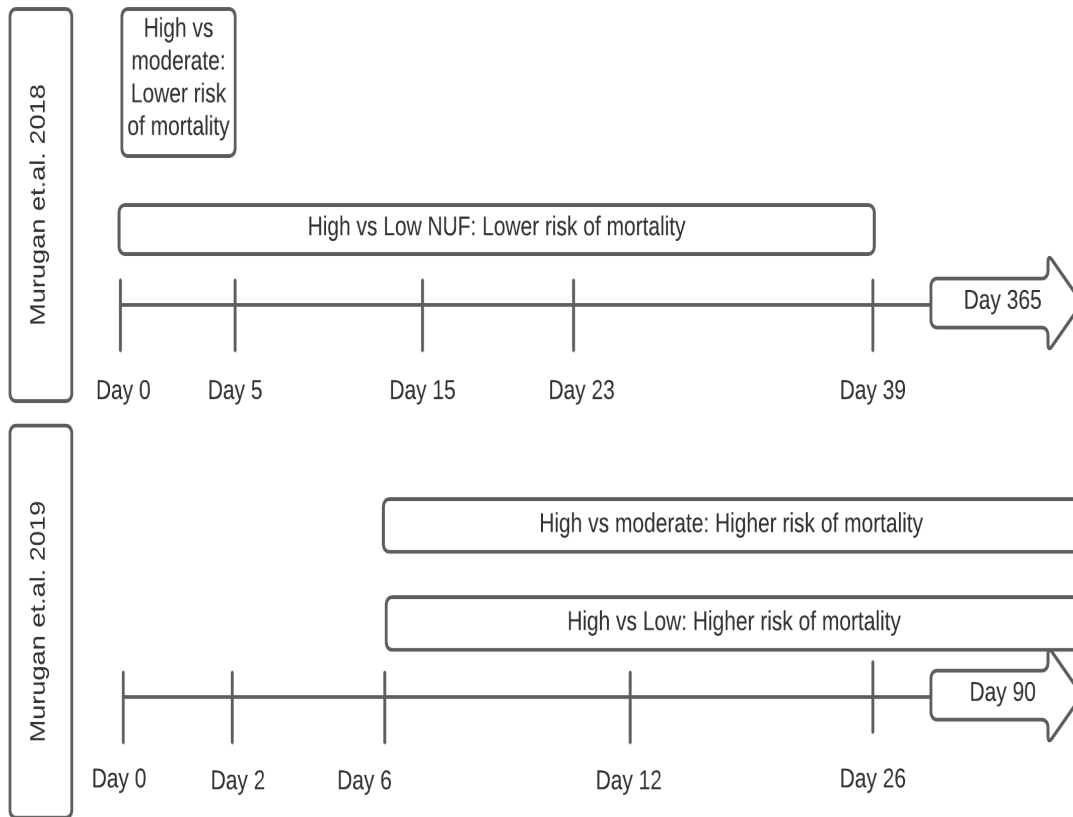
Figure 6: Summary of Gray's survival model result in Murugan et.al. 2018 and 2019

Figure 7: Missing value percentage

Figure 8: MI performance with CART

Figure 9: Variable Importance Based on Different Methods

Figure 10: K-means Clusters on Adjusted Discrepancy at Day 1

87

Figure 11: Distribution of clusters based on EM algorithm for Adjusted Discrepancy at Day 1
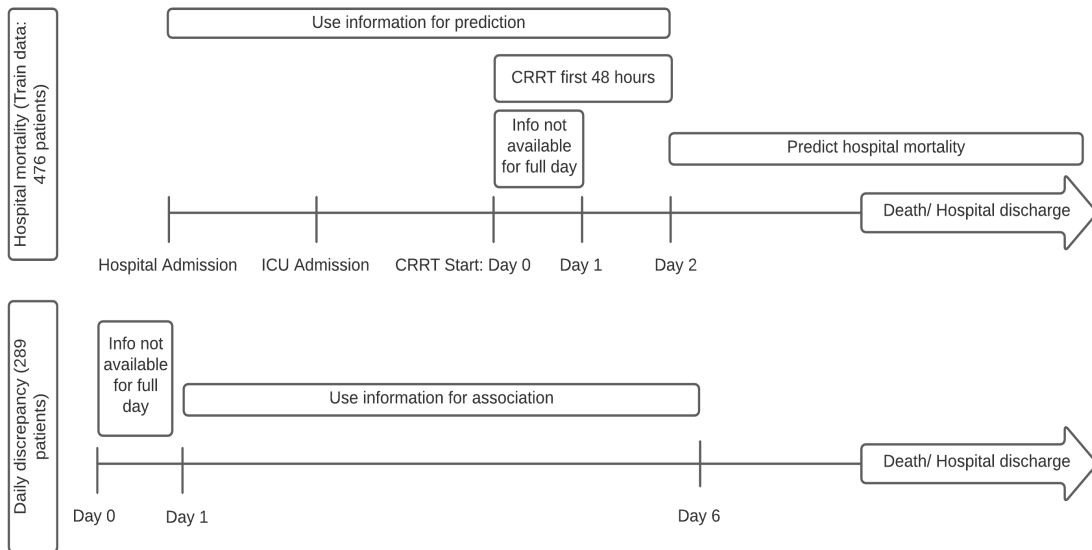
Figure 12: Timeline and Information of Interest
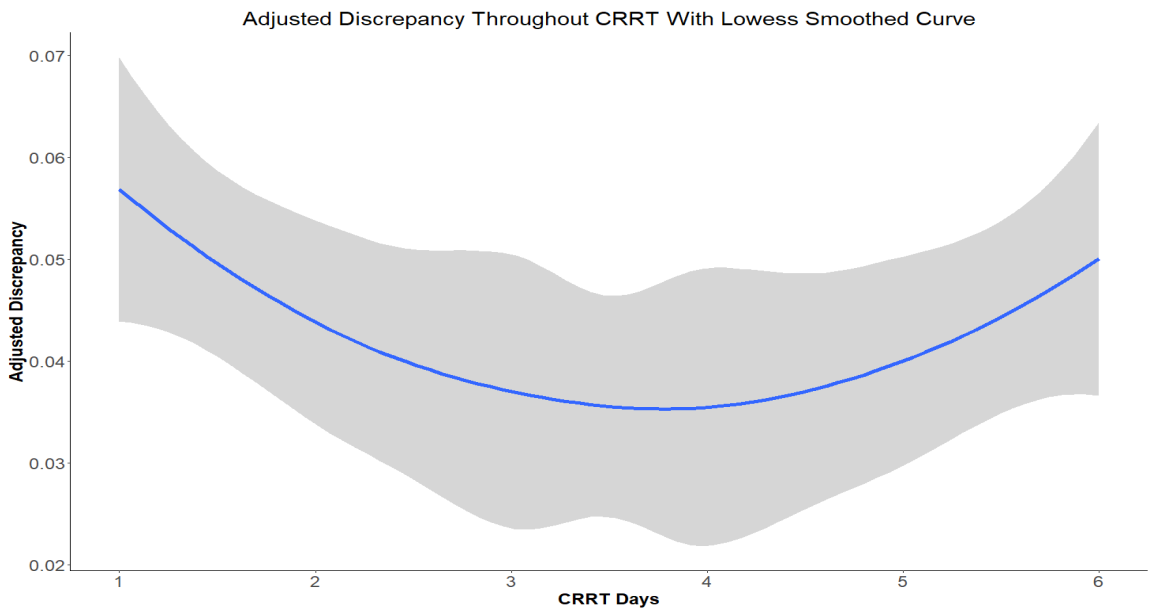
Figure 13: Daily Adjusted Discrepancy Throughout CRRT

Bibliography

[1] Wikipedia contributors, "Mixture model — Wikipedia, the free encyclopedia," 2020. [Online; accessed 12-November-2020].
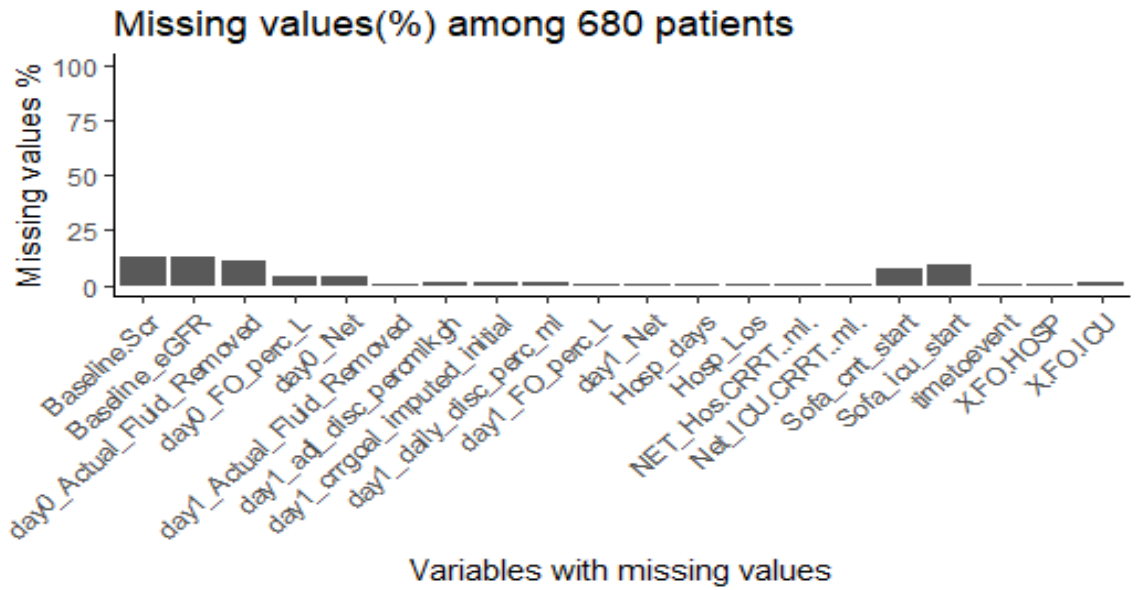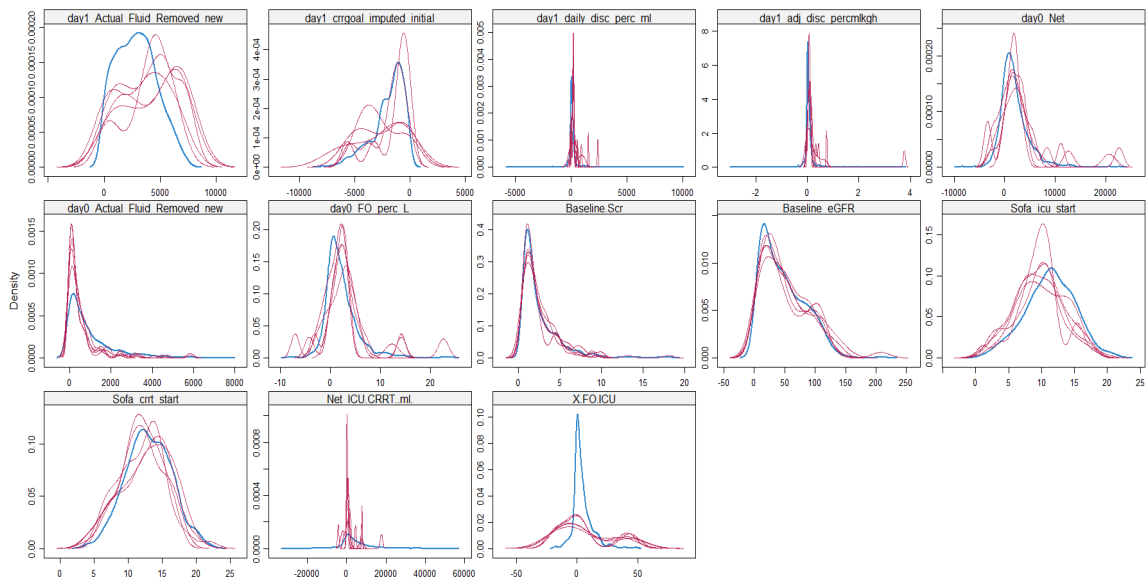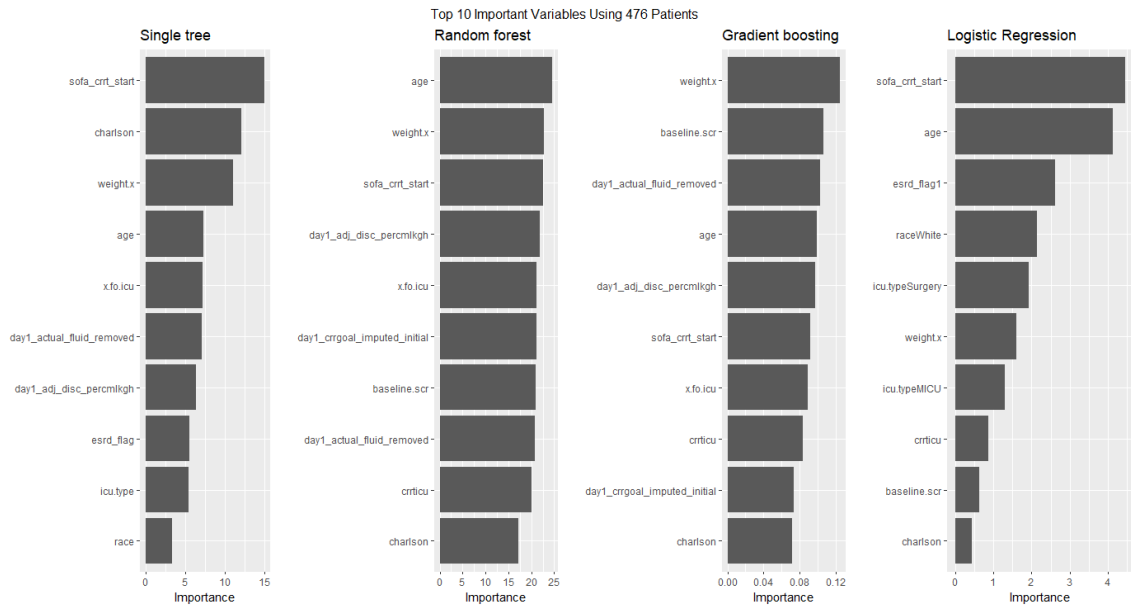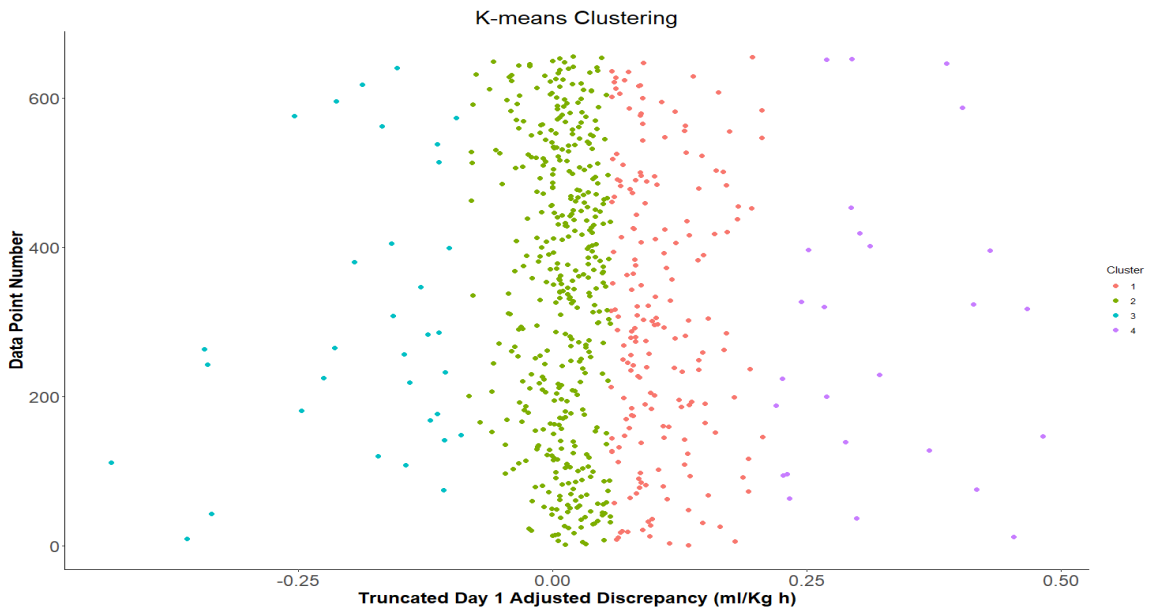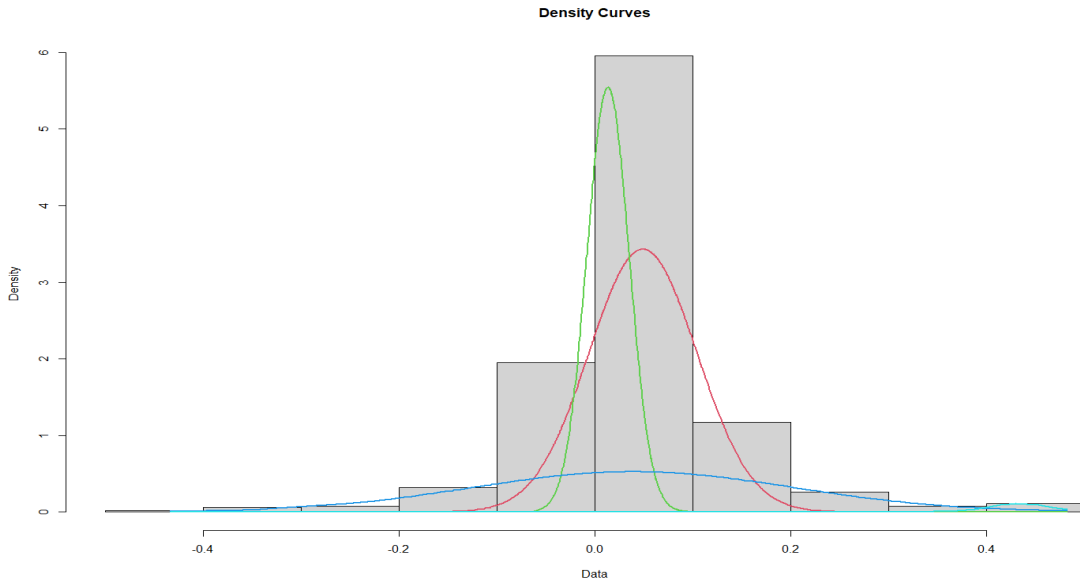
[2] H. Dai and R. Charnigo, "Omnibus testing and gene filtration in microarray data analysis," *Journal of Applied Statistics*, vol. 35, no. 1, pp. 31–47, 2008.

[3] P. Broët, A. Lewin, S. Richardson, C. Dalmasso, and H. Magdelenat, "A mixture model-based strategy for selecting sets of genes in multiclass response microarray experiments," *Bioinformatics*, vol. 20, no. 16, pp. 2562–2571, 2004.

[4] J. Tohka, E. Krestyannikov, I. D. Dinov, A. M. Graham, D. W. Shattuck, U. Ruotsalainen, and A. W. Toga, "Genetic algorithms for finite mixture model based voxel classification in neuroimaging," *IEEE transactions on medical imaging*, vol. 26, no. 5, pp. 696–711, 2007.

[5] J. Q. Trojanowski, H. Vandeerstichele, M. Korecka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, *et al.*, "Update on the biomarker core of the alzheimer's disease neuroimaging initiative subjects," *Alzheimer's & Dementia*, vol. 6, no. 3, pp. 230–238, 2010.

[6] A. Militino, M. Ugarte, and C. Dean, "The use of mixture models for identifying high risks in disease mapping," *Statistics in Medicine*, vol. 20, no. 13, pp. 2035–2049, 2001.

[7] P. Schlattmann and D. Böhning, "Mixture models and disease mapping," *Statistics in medicine*, vol. 12, no. 19-20, pp. 1943–1950, 1993.

[8] R. Charnigo, L. W. Chesnut, T. LoBianco, and R. S. Kirby, "Thinking outside the curve, part i: modeling birthweight distribution," *BMC pregnancy and childbirth*, vol. 10, no. 1, p. 37, 2010.

[9] R. Pilla and C. Charnigo, "Consistent estimation and model selection in semi-parametric mixtures," tech. rep., Technical Report, 2005.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[11] H. Chen and J. Chen, "The likelihood ratio test for homogeneity in finite mixture models," *Canadian Journal of Statistics*, vol. 29, no. 2, pp. 201–215, 2001.

[12] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the em algorithm," *SIAM review*, vol. 26, no. 2, pp. 195–239, 1984.

[13] G. J. McLachlan, "On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 36, no. 3, pp. 318–324, 1987.

[14] P. Bickel, "Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem," *Statistics and Probability: A Raghu Raj Bahadur Gestschrift*, 1993.

[15] H. Chen, J. Chen, and J. D. Kalbfleisch, "A modified likelihood ratio test for homogeneity in finite mixture models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 19–29, 2001.

[16] R. Charnigo and J. Sun, "Testing homogeneity in a mixture distribution via the l 2 distance between competing models," *Journal of the American Statistical Association*, vol. 99, no. 466, pp. 488–498, 2004.

[17] D. W. Scott, "Parametric modeling by minimum l2 error," *Technical Report 98-3*, 1998.

[18] R. A. Johnson, D. W. Wichern, *et al.*, *Applied multivariate statistical analysis*, vol. 5. Prentice hall Upper Saddle River, NJ, 2002.

[19] R. Charnigo and J. Sun, "Asymptotic relationships between the d-test and likelihood ratio-type tests for homogeneity," *Statistica Sinica*, pp. 497–512, 2010.

[20] P. Li, J. Chen, and P. Marriott, "Non-finite fisher information and homogeneity: an em approach," *Biometrika*, vol. 96, no. 2, pp. 411–426, 2009.

[21] K. Roeder, "A graphical technique for determining the number of components in a mixture of normals," *Journal of the American Statistical Association*, vol. 89, no. 426, pp. 487–495, 1994.

[22] H. Dai and R. Charnigo, "Inferences in contaminated regression and density models," *Sankhyā: The Indian Journal of Statistics*, pp. 842–869, 2007.

[23] D. B. Allison, G. L. Gadbury, M. Heo, J. R. Fernández, C.-K. Lee, T. A. Prolla, and R. Weindruch, "A mixture model approach for the analysis of microarray gene expression data," *Computational Statistics & Data Analysis*, vol. 39, no. 1, pp. 1–20, 2002.

[24] D. Moore, R. D'Agostino, and M. Stevens, "Goodness-of-fit techniques," 1986.

[25] R. Reghunathan, M. Jayapal, L.-Y. Hsu, H.-H. Chng, D. Tai, B. P. Leung, and A. J. Melendez, "Expression profile of immune response genes in patients with severe acute respiratory syndrome," *BMC immunology*, vol. 6, no. 1, pp. 1–11, 2005.

[26] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, ch. 6, pp. 191–218. Springer series in statistics New York, 2001.

[27] C. Loader, *Local regression and likelihood*. Springer Science & Business Media, 1999.

[28] N. H. Lameire, A. Bagga, D. Cruz, J. De Maeseneer, Z. Endre, J. A. Kellum, K. D. Liu, R. L. Mehta, N. Pannu, W. Van Biesen, *et al.*, "Acute kidney injury: an increasing global concern," *The Lancet*, vol. 382, no. 9887, pp. 170–179, 2013.

[29] E. D. Siew and A. Davenport, "The growth of acute kidney injury: a rising tide or just closer attention to detail?," *Kidney international*, vol. 87, no. 1, pp. 46–61, 2015.

[30] E. A. Hoste and M. Schurgers, "Epidemiology of acute kidney injury: how big is the problem?," *Critical care medicine*, vol. 36, no. 4, pp. S146–S151, 2008.

[31] R. Bellomo, J. A. Kellum, and C. Ronco, "Acute kidney injury," *The Lancet*, vol. 380, no. 9843, pp. 756–766, 2012.

[32] E. P. Paganini, M. Tapolyai, M. Goormastic, W. Halstenberg, L. Kozlowski, M. Leblanc, J. C. Lee, L. Moreno, and K. Sakai, "Establishing a dialysis therapy/patient outcome link in intensive care unit acute dialysis for patients with acute renal failure," *American journal of kidney diseases*, vol. 28, no. 5, pp. S81–S89, 1996.

[33] A. G. Schneider, R. Bellomo, S. M. Bagshaw, N. J. Glassford, S. Lo, M. Jun, A. Cass, and M. Gallagher, "Choice of renal replacement therapy modality and dialysis dependence after acute kidney injury: a systematic review and meta-analysis," *Intensive care medicine*, vol. 39, no. 6, pp. 987–997, 2013.

[34] J. Cerda, A. Tolwani, and P. Palevsky, "Challenges of performing renal replacement therapy in the intensive care unit-the nephrologist perspective," *Clinical Nephrology*, vol. 90, no. 1, p. 11, 2018.

[35] S. L. Goldstein, "Continuous renal replacement therapy: mechanism of clearance, fluid removal, indications and outcomes," *Current opinion in pediatrics*, vol. 23, no. 2, pp. 181–185, 2011.

[36] A. Berbece and R. Richardson, "Sustained low-efficiency dialysis in the icu: cost, anticoagulation, and solute removal," *Kidney international*, vol. 70, no. 5, pp. 963–968, 2006.

[37] Wikipedia contributors, "Hypervolemia — Wikipedia, the free encyclopedia." `https://en.wikipedia.org/w/index.php?title=Hypervolemia&oldid=976397884`, 2020. [Online; accessed 20-October-2020].

[38] C. Ronco, M. R. Costanzo, R. Bellomo, and A. S. Maisel, *Fluid overload*. Karger Medical and Scientific Publishers, 2010.

[39] R. Claure-Del Granado and R. L. Mehta, "Fluid overload in the icu: evaluation and management," *BMC nephrology*, vol. 17, no. 1, pp. 1–9, 2016.

[40] M. Heung, D. F. Wolfgram, M. Kommareddi, Y. Hu, P. X. Song, and A. O. Ojo, "Fluid overload at initiation of renal replacement therapy is associated with lack of renal recovery in patients with acute kidney injury," *Nephrology Dialysis Transplantation*, vol. 27, no. 3, pp. 956–961, 2012.

[41] S. T. Vaara, A.-M. Korhonen, K.-M. Kaukonen, S. Nisula, O. Inkinen, S. Hoppu, J. J. Laurila, L. Mildh, M. Reinikainen, V. Lund, *et al.*, "Fluid overload is associated with an increased risk for 90-day mortality in critically ill patients with renal replacement therapy: data from the prospective finnaki study," *Critical care*, vol. 16, no. 5, p. R197, 2012.

[42] Y.-C. Tsai, J.-C. Tsai, S.-C. Chen, Y.-W. Chiu, S.-J. Hwang, C.-C. Hung, T.-H. Chen, M.-C. Kuo, and H.-C. Chen, "Association of fluid overload with kidney

disease progression in advanced ckd: a prospective cohort study," *American Journal of Kidney Diseases*, vol. 63, no. 1, pp. 68–75, 2014.

[43] S. L. Goldstein, H. Currier, J. M. Graf, C. C. Cosio, E. D. Brewer, and R. Sachdeva, "Outcome in children receiving continuous venovenous hemofiltration," *Pediatrics*, vol. 107, no. 6, pp. 1309–1312, 2001.

[44] C. W. Woodward, J. Lambert, V. Ortiz-Soriano, Y. Li, M. Ruiz-Conejo, B. D. Bissell, A. Kelly, P. Adams, L. Yessayan, P. E. Morris, *et al.*, "Fluid overload associates with major adverse kidney events in critically ill patients with acute kidney injury requiring continuous renal replacement therapy," *Critical care medicine*, vol. 47, no. 9, pp. e753–e760, 2019.

[45] N. Alwall, "On the artificial kidney. i: Apparatus for dialysis of the blood in vivo," *Acta Medica Scandinavica*, vol. 128, no. 4, pp. 317–325, 1947.

[46] J. E. Flythe, G. C. Curhan, and S. M. Brunelli, "Disentangling the ultrafiltration rate–mortality association: The respective roles of session length and weight gain," *Clinical Journal of the American Society of Nephrology*, vol. 8, no. 7, pp. 1151–1161, 2013.

[47] S. J. Davies, E. A. Brown, W. Reigel, E. Clutterbuck, O. Heimbürger, N. V. Diaz, G. J. Mellote, J. Perez-Contreras, R. Scanziani, C. D'Auzac, *et al.*, "What is the link between poor ultrafiltration and increased mortality in anuric patients on automated peritoneal dialysis? analysis of data from eapos," *Peritoneal dialysis international*, vol. 26, no. 4, pp. 458–465, 2006.

[48] J. O. Burton, H. J. Jefferies, N. M. Selby, and C. W. McIntyre, "Hemodialysis-induced repetitive myocardial injury results in global and segmental reduction in systolic cardiac function," *Clinical Journal of the American Society of Nephrology*, vol. 4, no. 12, pp. 1925–1931, 2009.

[49] J. A. Silversides, R. Pinto, R. Kuint, R. Wald, M. A. Hladunewich, S. E. Lapinsky, and N. K. Adhikari, "Fluid balance, intradialytic hypotension, and outcomes in critically ill patients undergoing renal replacement therapy: a cohort study," *Critical care*, vol. 18, no. 6, p. 624, 2014.

[50] R. Murugan, M. Ostermann, Z. Peng, K. Kitamura, S. Fujitani, S. Romagnoli, L. Di Lullo, N. Srisawat, S. Todi, N. Ramakrishnan, *et al.*, "Net ultrafiltration prescription and practice among critically ill patients receiving renal replacement therapy: a multinational survey of critical care practitioners," *Critical care medicine*, vol. 48, no. 2, pp. e87–e97, 2020.

[51] H. Choi, D. Ghosh, and A. I. Nesvizhskii, "Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling," *Journal of proteome research*, vol. 7, no. 01, pp. 286–292, 2008.

[52] J. Schäfer and K. Strimmer, "An empirical bayes approach to inferring large-scale gene association networks," *Bioinformatics*, vol. 21, no. 6, pp. 754–764, 2005.

[53] M. Pagel and A. Meade, "A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data," *Systematic biology*, vol. 53, no. 4, pp. 571–581, 2004.

[54] T. M. Nguyen and Q. J. Wu, "Dirichlet gaussian mixture model: Application to image segmentation," *Image and Vision Computing*, vol. 29, no. 12, pp. 818–828, 2011.

[55] S. Rattanasiri, D. Bohning, P. Rojanavipart, and S. Athipanyakom, "A mixture model application in disease mapping of malaria," *Southeast Asian journal of tropical medicine and public health*, vol. 35, pp. 38–47, 2004.

[56] J. Detilleux and P. Leroy, "Application of a mixed normal mixture model for the estimation of mastitis-related parameters," *Journal of dairy science*, vol. 83, no. 10, pp. 2341–2349, 2000.

[57] R. C. Jansen, D. L. Johnson, and J. A. Van Arendonk, "A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families," *Genetics*, vol. 148, no. 1, pp. 391–399, 1998.

[58] B. Efron, "Large-scale simultaneous hypothesis testing: the choice of a null hypothesis," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 96–104, 2004.

[59] D. Dacunha-Castelle and E. Gassiat, "Testing the order of a model using locally conic parametrization: population mixtures and stationary arma processes," *The Annals of Statistics*, vol. 27, no. 4, pp. 1178–1209, 1999.

[60] M. Wand, *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*, 2020. R package version 2.23-17.

[61] P. Hall, S. J. Sheather, M. Jones, and J. S. Marron, "On optimal data-based bandwidth selection in kernel density estimation," *Biometrika*, vol. 78, no. 2, pp. 263–269, 1991.

[62] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 53, no. 3, pp. 683–690, 1991.

[63] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[64] R. C. Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[65] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization.* John Wiley & Sons, 2015.

[66] S. J. Sheather, "Density estimation," *Statistical science*, pp. 588–597, 2004.

[67] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S.* New York: Springer, fourth ed., 2002. ISBN 0-387-95457-0.

[68] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the royal statistical society. series c (applied statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[69] E. A. Hoste, S. M. Bagshaw, R. Bellomo, C. M. Cely, R. Colman, D. N. Cruz, K. Edipidis, L. G. Forni, C. D. Gomersall, D. Govil, *et al.*, "Epidemiology of acute kidney injury in critically ill patients: the multinational aki-epi study," *Intensive care medicine*, vol. 41, no. 8, pp. 1411–1423, 2015.

[70] W. Druml, F. Lax, G. Grimm, B. Schneeweiss, K. Lenz, and A. Laggner, "Acute renal failure in the elderly 1975-1990.," *Clinical nephrology*, vol. 41, no. 6, pp. 342–349, 1994.

[71] P. Susantitaphong, D. N. Cruz, J. Cerda, M. Abulfaraj, F. Alqahtani, I. Koulouridis, and B. L. Jaber, "World incidence of aki: a meta-analysis," *Clinical Journal of the American Society of Nephrology*, vol. 8, no. 9, pp. 1482–1493, 2013.

[72] M. Heung, S. M. Bagshaw, A. A. House, L. A. Juncos, R. Piazza, and S. L. Goldstein, "Crrtnet: a prospective, multi-national, observational study of con-

tinuous renal replacement therapy practices," *BMC nephrology*, vol. 18, no. 1, pp. 1–7, 2017.

[73] K. Makris and L. Spanou, "Acute kidney injury: definition, pathophysiology and clinical phenotypes," *The clinical biochemist reviews*, vol. 37, no. 2, p. 85, 2016.

[74] J. Bouchard, S. B. Soroko, G. M. Chertow, J. Himmelfarb, T. A. Ikizler, E. P. Paganini, and R. L. Mehta, "Fluid accumulation, survival and recovery of kidney function in critically ill patients with acute kidney injury," *Kidney international*, vol. 76, no. 4, pp. 422–427, 2009.

[75] A. S. Messmer, C. Zingg, M. Müller, J. L. Gerber, J. C. Schefold, and C. A. Pfortmueller, "Fluid overload and mortality in adult critical care patients—a systematic review and meta-analysis of observational studies," *Critical care medicine*, vol. 48, no. 12, pp. 1862–1870, 2020.

[76] A. Tolwani, "Continuous renal-replacement therapy for acute kidney injury," *New England Journal of Medicine*, vol. 367, no. 26, pp. 2505–2514, 2012.

[77] S. Tandukar and P. M. Palevsky, "Continuous renal replacement therapy: who, when, why, and how," *Chest*, vol. 155, no. 3, pp. 626–638, 2019.

[78] O. Joannes-Boyau, P. M. Honoré, P. Perez, S. M. Bagshaw, H. Grand, J.-L. Canivet, A. Dewitte, C. Flamens, W. Pujol, A.-S. Grandoulier, *et al.*, "High-volume versus standard-volume haemofiltration for septic shock patients with acute kidney injury (ivoire study): a multicentre randomized controlled trial," *Intensive care medicine*, vol. 39, no. 9, pp. 1535–1546, 2013.

[79] A. Paterson, A. Johnston, D. Kingston, and R. Mahroof, "Clinical and economic impact of a switch from high-to low-volume renal replacement therapy

in patients with acute kidney injury," *Anaesthesia*, vol. 69, no. 9, pp. 977–982, 2014.

[80] R. Murugan, V. Balakumar, S. J. Kerti, P. Priyanka, C.-C. H. Chang, G. Clermont, R. Bellomo, P. M. Palevsky, and J. A. Kellum, "Net ultrafiltration intensity and mortality in critically ill patients with fluid overload," *Critical Care*, vol. 22, no. 1, p. 223, 2018.

[81] R. Murugan, S. J. Kerti, C.-C. H. Chang, M. Gallagher, G. Clermont, P. M. Palevsky, J. A. Kellum, and R. Bellomo, "Association of net ultrafiltration rate with mortality among critically ill adults with acute kidney injury receiving continuous venovenous hemodiafiltration: a secondary analysis of the randomized evaluation of normal vs augmented level (renal) of renal replacement therapy trial," *JAMA network open*, vol. 2, no. 6, pp. e195418–e195418, 2019.

[82] S. Tehranian, K. Shawwa, and K. B. Kashani, "Net ultrafiltration rate and its impact on mortality in patients with acute kidney injury receiving continuous renal replacement therapy," *Clinical kidney journal*, vol. 14, no. 2, pp. 564–569, 2021.

[83] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, "The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *Journal of clinical epidemiology*, vol. 62, no. 10, pp. e1–e34, 2009.

[84] T. Naorungroj, A. S. Neto, L. Zwakman-Hessels, Y. Fumitaka, G. Eastwood, R. Murugan, J. A. Kellum, and R. Bellomo, "Mediators of the impact of hourly net ultrafiltration rate on mortality in critically ill patients receiving continuous

renal replacement therapy," *Critical care medicine*, vol. 48, no. 10, pp. e934–e942, 2020.

[85] T. Naorungroj, A. Serpa Neto, R. Murugan, J. A. Kellum, and R. Bellomo, "Continuous renal replacement therapy: the interaction between fluid balance and net ultrafiltration," *American journal of respiratory and critical care medicine*, vol. 203, no. 9, pp. 1199–1201, 2021.

[86] R. R. T. S. Investigators, "Intensity of continuous renal-replacement therapy in critically ill patients," *New England Journal of Medicine*, vol. 361, no. 17, pp. 1627–1638, 2009.

[87] Z. Valenta and L. Weissfeld, "Estimation of the survival function for gray's piecewise-constant time-varying coefficients model," *Statistics in medicine*, vol. 21, no. 5, pp. 717–727, 2002.

[88] P. C. Austin, A. Latouche, and J. P. Fine, "A review of the use of time-varying covariates in the fine-gray subdistribution hazard competing risk regression model," *Statistics in medicine*, vol. 39, no. 2, pp. 103–113, 2020.

[89] M. Legrand, M. Darmon, M. Joannidis, and D. Payen, "Management of renal replacement therapy in icu patients: an international survey," *Intensive care medicine*, vol. 39, no. 1, pp. 101–108, 2013.

[90] M. Bell, F. Granath, S. Schön, A. Ekbom, and C.-R. Martling, "Continuous renal replacement therapy is associated with less chronic renal failure than intermittent haemodialysis after acute renal failure," *Intensive care medicine*, vol. 33, no. 5, pp. 773–780, 2007.

[91] A. Y. Wang and R. Bellomo, "Renal replacement therapy in the icu: intermittent hemodialysis, sustained low-efficiency dialysis or continuous renal replace-

ment therapy?," *Current opinion in critical care*, vol. 24, no. 6, pp. 437–442, 2018.

[92] L. Zhang, Z. Chen, Y. Diao, Y. Yang, and P. Fu, "Associations of fluid overload with mortality and kidney recovery in patients with acute kidney injury: a systematic review and meta-analysis," *Journal of critical care*, vol. 30, no. 4, pp. 860–e7, 2015.

[93] H. E. Wang, P. Muntner, G. M. Chertow, and D. G. Warnock, "Acute kidney injury and mortality in hospitalized patients," *American journal of nephrology*, vol. 35, no. 4, pp. 349–355, 2012.

[94] S. G. Coca, B. Yusuf, M. G. Shlipak, A. X. Garg, and C. R. Parikh, "Long-term risk of mortality and other adverse outcomes after acute kidney injury: a systematic review and meta-analysis," *American journal of kidney diseases*, vol. 53, no. 6, pp. 961–973, 2009.

[95] P. Hayati Rezvan, K. J. Lee, and J. A. Simpson, "The rise of multiple imputation: a review of the reporting and implementation of the method in medical research," *BMC medical research methodology*, vol. 15, no. 1, pp. 1–14, 2015.

[96] P. D. Faris, W. A. Ghali, R. Brant, C. M. Norris, P. D. Galbraith, M. L. Knudtson, A. Investigators, *et al.*, "Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses," *Journal of clinical epidemiology*, vol. 55, no. 2, pp. 184–191, 2002.

[97] M. G. Kenward and J. Carpenter, "Multiple imputation: current perspectives," *Statistical methods in medical research*, vol. 16, no. 3, pp. 199–218, 2007.

[98] T. Hayes, S. Usami, R. Jacobucci, and J. J. McArdle, "Using classification and regression trees (cart) and random forests to analyze attrition: Results from two simulations.," *Psychology and aging*, vol. 30, no. 4, p. 911, 2015.

[99] A. B. Pedersen, E. M. Mikkelsen, D. Cronin-Fenton, N. R. Kristensen, T. M. Pham, L. Pedersen, and I. Petersen, "Missing data and multiple imputation in clinical epidemiological research," *Clinical epidemiology*, vol. 9, p. 157, 2017.

[100] R. J. Little, "A test of missing completely at random for multivariate data with missing values," *Journal of the American statistical Association*, vol. 83, no. 404, pp. 1198–1202, 1988.

[101] J. A. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter, "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls," *Bmj*, vol. 338, 2009.

[102] Z. Zhang, "Multiple imputation with multivariate imputation by chained equation (mice) package," *Annals of translational medicine*, vol. 4, no. 2, 2016.

[103] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," 2018.

[104] Y. Xu and R. Goodacre, "On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," *Journal of analysis and testing*, vol. 2, no. 3, pp. 249–262, 2018.

[105] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[106] C. Fraley, A. E. Raftery, T. B. Murphy, and L. Scrucca, "mclust version 4 for r: normal mixture modeling for model-based clustering, classification, and density estimation," tech. rep., Technical report, 2012.

[107] B. Tatiana, C. Didier, R. David, and Y. Derek, "mixtools: An r package for analyzing finite mixture models," *Journal of Statistical Software*, vol. 32, no. 6, pp. 1–29, 2009.

[108] B. M. Greenwell, B. C. Boehmke, and A. J. McCarthy, "A simple and effective model-based variable importance measure," *arXiv preprint arXiv:1805.04755*, 2018.

[109] B. M. Greenwell and B. C. Boehmke, "Variable importance plots—an introduction to the vip package," *The R Journal*, vol. 12, no. 1, pp. 343–366, 2020.

[110] J. Lambert, L. Gong, C. F. Elliott, K. Thompson, and A. Stromberg, "rfsa: An r package for finding best subsets and interactions.," *R Journal*, vol. 10, no. 2, 2018.

[111] F. Schiltz, C. Masci, T. Agasisti, and D. Horn, "Using regression tree ensembles to model interaction effects: a graphical approach," *Applied Economics*, vol. 50, no. 58, pp. 6341–6354, 2018.

[112] S. Institute, *SAS 9.4 language reference: Concepts*. SAS Institute, 2018.

[113] R. M. O'brien, "A caution regarding rules of thumb for variance inflation factors," *Quality & quantity*, vol. 41, no. 5, pp. 673–690, 2007.

[114] C. Robinson and R. E. Schumacker, "Interaction effects: centering, variance inflation factor, and interpretation issues," *Multiple linear regression viewpoints*, vol. 35, no. 1, pp. 6–11, 2009.

[115] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the youden index and its associated cutoff point," *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 47, no. 4, pp. 458–472, 2005.

[116] M. D. Ruopp, N. J. Perkins, B. W. Whitcomb, and E. F. Schisterman, "Youden index and optimal cut-point estimated from observations affected by a lower limit of detection," *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, vol. 50, no. 3, pp. 419–430, 2008.

[117] J. Myerson, L. Green, and M. Warusawitharana, "Area under the curve as a measure of discounting," *Journal of the experimental analysis of behavior*, vol. 76, no. 2, pp. 235–243, 2001.

[118] I. Hen, A. Sakov, N. Kafkafi, I. Golani, and Y. Benjamini, "The dynamics of spatial behavior: how can robust smoothing techniques help?," *Journal of neuroscience methods*, vol. 133, no. 1-2, pp. 161–172, 2004.

[119] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

[120] S. Vesconi, D. N. Cruz, R. Fumagalli, D. Kindgen-Milles, G. Monti, A. Marinho, F. Mariano, M. Formica, M. Marchesi, R. Robert, *et al.*, "Delivered dose of renal replacement therapy and mortality in critically ill patients with acute kidney injury," *Critical care*, vol. 13, no. 2, pp. 1–14, 2009.

[121] Y. K. Kee, D. Kim, S.-J. Kim, D.-H. Kang, K. B. Choi, H. J. Oh, and D.-R. Ryu, "Factors associated with early mortality in critically ill patients following the initiation of continuous renal replacement therapy," *Journal of clinical medicine*, vol. 7, no. 10, p. 334, 2018.

[122] H. Brar, J. Olivier, C. Lebrun, W. Gabbard, T. Fulop, and D. Schmidt, "Predictors of mortality in a cohort of intensive care unit patients with acute renal failure receiving continuous renal replacement therapy," *The American journal of the medical sciences*, vol. 335, no. 5, pp. 342–347, 2008.

[123] V. Ortiz-Soriano, S. Kabir, C.-D. Granado, A. Stromberg, R. D. Toto, O. W. Moe, S. L. Goldstein, J. A. Neyra, *et al.*, "Assessment of a modified renal angina index for aki prediction in critically ill adults," *Nephrology Dialysis Transplantation*, 2021.

[124] Y.-C. Chen, F.-C. Tsai, C.-H. Chang, C.-Y. Lin, C.-C. Jenq, K.-C. Juan, H.-H. Hsu, M.-Y. Chang, Y.-C. Tian, C.-C. Hung, *et al.*, "Prognosis of patients on extracorporeal membrane oxygenation: the impact of acute kidney injury on mortality," *The Annals of thoracic surgery*, vol. 91, no. 1, pp. 137–142, 2011.

[125] W. Ribitsch, J. H. Horina, F. Quehenberger, A. R. Rosenkranz, and G. Schilcher, "Contrast induced acute kidney injury and its impact on midterm kidney function, cardiovascular events and mortality," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019.

[126] H. Chen, B. Wu, D. Gong, and Z. Liu, "Fluid overload at start of continuous renal replacement therapy is associated with poorer clinical condition and outcome: a prospective observational study on the combined use of bioimpedance vector analysis and serum n-terminal pro-b-type natriuretic peptide measurement," *Critical Care*, vol. 19, no. 1, pp. 1–8, 2015.

[127] R. Charon, M. G. Greene, and R. D. Adelman, "Multi-dimensional interaction analysis: a collaborative approach to the study of medical discourse," *Social Science & Medicine*, vol. 39, no. 7, pp. 955–965, 1994.

[128] T. J. Beckman and M. C. Lee, "Proposal for a collaborative approach to clinical teaching," in *Mayo Clinic Proceedings*, vol. 84, pp. 339–344, Elsevier, 2009.

[129] A. P. Moore, "Research, the collaborative approach: clinicians and academics," *Physiotherapy*, vol. 83, no. 5, pp. 229–234, 1997.

Vita

<p align="center">Shaowli Kabir</p>

Place of Birth:

- Dhaka, Bangladesh

Education:

- University of Kentucky, Lexington, KY

  MS in Statistics, 2018

- University of Dhaka, Dhaka, Bangladesh

  MS in Statistics, Biostatistics & Informatics, 2016

- University of Dhaka, Dhaka, Bangladesh

  BS(Hons) in Statistics, Biostatistics & Informatics, 2014

Professional Positions:

- Graduate Research Assistant, Department of Biostatistics, University of Kentucky August 2020 - May 2022

- Graduate Teaching Assistant, Department of Biostatistics, University of Kentucky August 2020 - May 2022

- Graduate Research Assistant, Applied Statistics Lab, University of Kentucky May 2018 - July 2020

- Graduate Teaching Assistant, Department of Statistics, University of Kentucky August 2016 - May 2018

Publications & Preprints:

- Assessment of a modified renal angina index for AKI prediction in critically ill adults Ortiz-Soriano, Victor, **Shaowli Kabir**, Claure-Del Granado, Arnold Stromberg, Robert D. Toto, Orson W. Moe, Stuart L. Goldstein, and Javier A. Neyra. Nephrology Dialysis Transplantation

- Predictive value of platelet reactivity unit (PRU) value for thrombotic and hemorrhagic events during flow diversion procedures: a meta-analysis E Ajadi, **S Kabir**, A Cook, S Grupke, A Alhajeri, J Fraser - Journal of neurointerventional surgery, 2019,Volume 11, Issue 11, Page-1123-1128.

- The epidemiology of adult patients receiving Continuous Renal Replacement Therapy: A multicenter CRRTnet study O Rewa, V Ortiz-Soriano, J Lambert, S Kabir, J Neyra...-Journal of the American Society of Nephrology (Submitted)

- Abstract WMP26: Aneurysm Location and Its Relationship With Thrombotic and Hemorrhagic Complications During Flow Diversion Procedures: A Meta-Analysis E Ajadi(Presenter), M Nisiewicz, **S Kabir**, J Fraser - International Stroke Conference, February 2020, Volume 51, Issue Suppl_1.

- Impact of acute exposure to Angiotensin Converting Enzyme Inhibitors and Angiotensin Receptor Blockers on Acute Kidney Recovery: A multicenter prospective cohort study Litteral(Presenter), Sherif Saleh, Weihang Ren, **Shaowli Kabir**, Victor Ortiz-Soriano, Marice Ruiz-Conejo, Ron Wald, Samuel A. Silver, and A. Javier. "NKF 2019 Spring Clinical Meetings Abstracts."