Spring 4-15-2019

Assessing the efficacy of Home-Based Renal Care using Propensity Scores

Eunice Choi

Eun Ho Eunice Choi
Candidate

Mathematics and Statistics
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

James Degnan, Ph.D., Chairperson

Shane Pankratz, Ph.D.

Fares Qeadan, Ph.D.

Fletcher Christensen, Ph.D.

Assessing the efficacy of Home-Based Renal Care using Propensity Scores

BY

Eunice Choi

B.S. Physiology, B.A. Psychology, University of Washington, 2013

Master's Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Statistics

The University of New Mexico

Albuquerque, New Mexico

May, 2019

Dedication

I would like to dedicate this thesis to my parents for their unconditional love and support, to my brother, Sung Ho, for always accepting the depths of me exactly as they are without a hint of judgement, to my grandparents who would have been so proud of me, to Sarah, Shanna and Daniella for many years of companionship and their kind spirited ambitions, to Joe and Diana for always believing in me and to Cody Fritts for his quiet, persistent presence through the uncertainty and for all our conversations that were unequivocally foundational to my ideas.

Acknowledgement

I would like to acknowledge my committee chair and advisor, Dr. James Degnan. He taught me probability, inference, and SAS programming. Most importantly, he has always allowed me the time and space to critically think about my ideas and new concepts independently. It really enhanced my learning because I was able to solidify complicated concepts in my own terms without feeling overwhelmed. Thank you for never rushing me and never saying no to any of my requests.

Dr. Shane Pankratz for helping me analyze this study. He guided me through the bulk of the analysis. I learned so much from handling real world data. I had so much fun analyzing the data and struggling through the literature to learn about generalized estimating equations and propensity score. With the specific skillsets I built through writing my thesis, I feel excited to struggle through even more real world data. Thank you.

Dr. Fares Qeadan for the structure he has provided and the time and energy he has devoted to my learning. His consistent encouragement and devotion were absolutely crucial in completing my thesis. Thank you for always being patient with me even through stressful moments.

Dr. Fletcher Christensen for the thoughtful feedback. It definitely tested my knowledge and gave me an opportunity to consider a different perspective.

Last but not least, I would like to acknowledge Dr. Raj Shah for providing the data. It gave me an opportunity to think critically about health disparities in New Mexico. It made the data analysis so much more meaningful. Also, I would like to express my gratitude towards the Zuni tribe in New Mexico who participated in the study.

Assessing the efficacy of Home-Based Renal Care
using Propensity Scores

by

Eunice Choi

B.S. Physiology, B.A. Psychology, University of Washington, 2013
M.S. Statistics, University of New Mexico, 2019

Abstract

This study investigated the efficacy of Home-Based Renal Care (HBRC) in diabetic Zuni Indians with Chronic Kidney Disease (CKD) in New Mexico using propensity scores. Home based intervention as opposed to standard clinical care is a pragmatic treatment approach that incorporates the preference of population in hopes of addressing a cultural barrier to healthcare in this high risk population. This study uses a logistic regression model and a linear regression model to estimate the average effect of HBRC on increasing the likelihood of participants taking a more active role in the management of their chronic condition compared to the control group. We used generalized estimating equations (GEE) to account for household clustering and stabilized inverse probability of treatment weighting (SIPTW) to reduce any estimate bias that may have been introduced.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 2013, Americans spent an estimated 101.4 billion US dollars on diabetes mellitus management making it the largest spending on a health condition in the United States.[1] Though alarming, it does not come as a surprise that the prevalence of diabetes has been on the rise in the last couple of decades. The estimated percentage of Americans diagnosed with diabetes has doubled from 4% to 8.4% between 1999 and 2018.[2] Diabetes impacts the quality of life of many Americans.

Furthermore, diabetes is a risk factor for many chronic health conditions. For example, diabetes has been previously described as the leading cause of advanced kidney disease worldwide.[3] Specifically, chronic kidney disease (CKD) affects 13% of the general adult American population diagnosed with type 2 diabetes mellitus (T2DM).[4] In high-risk populations, the high prevalence rate of diabetes leads to a faster progression of CKD. Particularly, studies have demonstrated that CKD is prevalent among ethnic and racial minorities.[5] It is important to recognize that diabetes has been identified as a modifiable risk factor for CKD progression that disproportionately affect socially disadvantaged groups.

The burden of CKD is greater in ethnic and racial minorities, and rural communities where access to healthcare is limited.[6] For example, the majority of Zuni Indians in New Mexico live in remote parts of the state and subsequently have limited access to health care. The combination of their lack of access to health services and high rates of chronic diseases among them masks the true nature of their sickness as seen with the cardiovascular disease in the Zuni tribe.[7] This health disparity is worsened by the barriers of healthcare. The focus of our work will be on Zuni Indians in New Mexico.

Cultural barriers are another dimension that goes beyond limited access to health care. However, many interventions fail to address such barriers specifically by not accounting for the cultural awareness and the preference of the population of interest. The Zuni Indians, for instance, have experienced considerable historical and cultural trauma that resulted in fear of participating in health screening and healthcare.[8] In addition to the fear, the mistrust and difficulty in building a trusting relationship with healthcare providers, due to the high turnover rate of medical staff at the Indian Health Services, were identified as barriers within the operation of health care systems.[7]

Home based intervention as opposed to standard clinical care addresses another barrier to healthcare—the preference of population. For instance, a recent study found that the Zuni Indians in New Mexico culturally preferred to receive care where confidentiality is easily attained.[7] In this community, going to a clinic, where patients could be seen by others in waiting rooms, could sometimes lead to feeling embarrassed which in turns might prevent the patients from receiving care at the health services on a regular basis.

In this study, the proposed intervention was designed to address these barriers by using community health workers (CHW) who are members of the Zuni tribe but were trained specifically for the delivery of the intervention. In contrast to standard clinical care, where patients visit clinics, CHWs in this intervention provided care at patients' homes which takes into consideration that most patients feel more comfortable when receiving care from people who look like them. In this 12-month randomized controlled trial, the intervention was specifically designed to improve patients' inclination to take a more active role in the management of their chronic health condition. Hence, the primary outcome of interest of this study was patient's activation score which is a measure of their involvement in the health care management. The objective of this trial is to examine the effectiveness of home-based renal care in comparison to standard care with respect to patients' activation scores. We hypothesized that patients were more likely to engage in their health care management if they received care at home.

Chapter 2

Methods

The aim of this thesis was to analyze the effectiveness of home-based renal care (HBRC) on increasing the likelihood of the diabetic Zuni Indians to be more active in the management of CKD. The primary outcomes of interest were (1) whether patient activation improved where activation is considered level 3 or greater (binary) and (2) the change in the patient activation score collected through a self-reported questionnaire (continuous). A logistic regression model was used to estimate the odds of improvement in patient activation level for the intervention group relative to the control group. A linear regression model was used to estimate the mean patient activation score for patients in the intervention group relative to controls. Because we enrolled family units into the study, Generalized Estimated Equations (GEE) were used to account for household clustering.

Participants

Potential participants (n = 1,436) from a previously established cohort were screened for relevant clinical factors.[9] Exactly 315 individuals were screened for eligibility where 127 met the criteria for inclusion in the study. The criteria included being between 21 and 80 years of age and having urine albumin : creatinine ratio ≥ 30 kg/m^2, hemoglobin A1c $\geq 7\%$, or a family history of diabetes and kidney disease. Two individuals declined to participate in the study. The remaining 125 were enrolled in the 12 month randomized controlled trial. The data from 72 individuals who were on diabetes medication (DM) at baseline were used for the purpose of this thesis.

Randomization

The randomization sequence was generated using the PROC PLAN procedure in SAS to assign participants to either the standard clinical care group or the home-based renal care group. The

randomization procedure permuted the two levels of treatment randomly and without replacement within blocks containing two, four, or six households. More than one person in a household could participate in the study. Because of this, we randomized households in a 1:1 allocation to the standard clinical care or the intervention to ensure that members of the same household were allocated to the same treatment group. While 96 households were enrolled in the study, only 40 households were enrolled in the diabetic subset that were analyzed for this thesis; 14 participants enrolled in the standard clinical care and 17 participants enrolled in the intervention group were from single-participant households. Neither the investigator nor the participants were blinded at randomization because it was clear which participants were receiving the intervention at home or the standard care at a local Indian Health Service clinic.

Data Flowchart

The data flow from enrollment to randomization to final assignment into the intervention and control groups is shown in Figure 2.1.

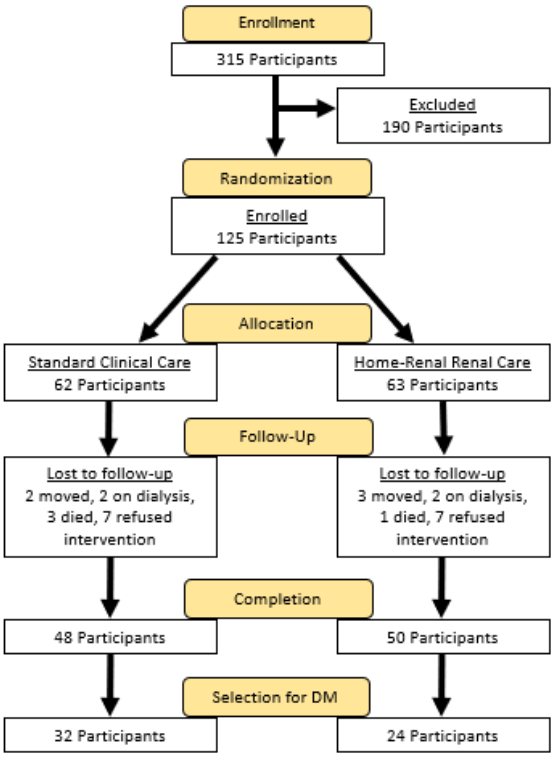


Figure 2.1 Data Flowchart

Explanatory Variables

Intervention/Treatment

The intervention group received Home-Based Renal Care (HBRC). The Community Health Workers (CHWs) were Zuni Indians who were trained to provide the HBRC to the intervention group. The CHWs received 40 hours of training that consisted of education about CKD and its self-management education, the theoretical framework of the intervention and its implementation. They visited the participants on a biweekly basis to educate them on various topics including but not limited to healthy eating, exercise, medication management and risk factor management.

The intervention focused on providing a pragmatic treatment plan that addressed cultural awareness. The participants chose to receive care in their native language if preferred by the participant. The diet and exercise plan incorporated traditional food items and culturally popular activities, and promoted group cohesion. Receiving care at home reflected the preference of the population that aimed to reduce any anxiety and discomfort that commonly result from going to the clinic.[7] They provided point-of care testing. This meant that the lab results were taken and interpreted to the participants immediately during the home visit. They were also provided with cellphones and text message plans, and received motivational text messages a few times per day.

The control group received standard clinical care from a local Indian Health Services clinic. They received publicly-available up-to-date information about diabetes prevention, weight loss, diet and exercise. They were only contacted by the study staff for the purpose of data collection.

Response Variables

Primary Outcomes of Interest

Patient Activation Score

The patient activation score (PAM score) has been previously reported as a validated tool that assesses a patient's ability to effectively participate in his or her care.[10] This instrument was collected through a short form questionnaire consisting 13 questions with 4 response options: (1) strongly disagree, (2) disagree, (3) agree, and (4) strongly agree. The raw score ranged from 39 to 53. If there were any missing response, the total score was divided by the number of answered questions and multiplied by 13 to yield a normalized raw score. A nomogram, provided under a licensing agreement with Insignia Health, converted raw scores to an activation score that ranged from 0 to 100. We note that in this work we will be using the converted activation score and not the raw score.

Patient Activation Level

The patient activation score was further categorized into 4 levels. They were:

Level 1: Believing the patient's role is important but not taking action;

Level 2: Having the confidence and knowledge necessary to take action;

Level 3: Taking action to maintain and improve one's health; and

Level 4: Staying the course even under stress.

Scoring level 3 and higher was grouped into one category representing patient activation. Scoring level 2 and lower was grouped into the other category representing the lack of patient activation. Figure 2.2 shows the identification of each patient activation levels from calibrating the patient activation score from the 13 item questionnaire.[11] The raw score of patient activation scores ranged from 38.6 to 53. A raw score of 39 to 41, 42 to 47, 50 to 51 and 52 to 53 were calibrated into level 1, level 2, level 3 and level 4 respectively.

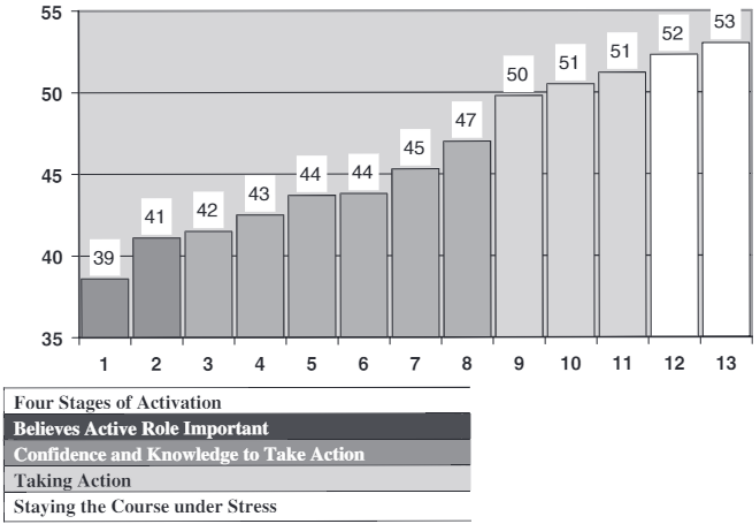


Figure 2.2 Calibration of the patient activation score into patient activation levels

Secondary Outcomes of Interest

Body Mass Index

Among clinicians, Body Mass Index (BMI) has been accepted as the better estimate of total body fat compared to body weight alone.[12] Obesity has been reported to be associated with multiple conditions that are known to cause compromised renal function such as hypertension and diabetes.[13] Also, it was found that it may be independently associated with the risk of developing CKD.[14] Hence, it was measured to evaluate the degree of excess weight and as a risk factor for CKD in this study. BMI was calculated by dividing body weight (kg) by height squared (m^2).

Blood Pressure

Hypertension is present in estimated 80-85% of patients with CKD.[14] The increased prevalence of hypertension has been primarily caused by sodium retention among other factors and been identified as a contributory factor in the development of kidney disease.[15] For this study, blood pressure (systolic and diastolic) was measured 3 times about 5 minutes apart with the participants resting in a seated position, then averaged.

Hemoglobin A1c

It has been found that hemoglobin A1c (HbA1c) level in blood reflects the mean blood glucose over 120-day lifespan of the red blood cell.[16] Hence, it is the most widely used clinical test to estimate the long-term mean blood glucose.[16] It is recommended that the target A1c value should be 7% or lower for most diabetic patients.[17] Hemoglobin A1c level (≥ 6.5 %) is a diagnostic criterion for diabetes.[18]

Serum Glucose

Fasting plasma glucose (≥ 126 mg/dL) is another diagnostic criterion for diabetes.[18] It is important to note that plasma glucose concentration fluctuates within the same day depending on food intake and other factors. A measure of plasma glucose should be supplemented by a measure of the hemoglobin A1c.

Serum Total Protein

Adaptive hyperfiltration induces proteinuria and progressive renal failure.[19] Hyperfiltration and proteinuria could lead to changes in the total protein concentration in plasma. Protein malnutrition is a common finding in chronic renal failure and is associated with poor outcome.[20]

Serum Cholesterol

Hyperlipidemia refers to high levels of lipids in blood including cholesterol and triglycerides. This does not directly cause symptoms but has been identified as a risk factor for cardiovascular disease, diabetes mellitus, and chronic kidney disease.[21] Cholesterol is essential for normal function of all animal cells and is a precursor of various critical substances such as steroid hormones and bile acids.[22] A total cholesterol level of less than 200mg/dL is considered normal.[23] A total cholesterol level greater than or equal to 240 mg/dL is considered high.[23]

Serum Triglycerides

Triglycerides is one type of lipid in blood. Triglycerides is different than cholesterol in that triglycerides represents the main lipid component of dietary fat.[22] High triglyceride level is associated with high risk of cardiovascular disease, diabetes and chronic kidney disease.[21] Less than 150 mg/dL is considered to be normal level and greater than 886 mg/dL is considered very high level of triglycerides.[23]

Serum HDL Cholesterol

High-density lipoproteins (HDL) is commonly known as the "good" cholesterol because it plays the role of reverse transporting cholesterol from different tissues to liver where cholesterol is eventually removed from the body.[22] Hence, HDL prevents excess cholesterol build up in the body.[22] Greater than or equal to 60 mg/dL is considered to be normal level whereas less than 40 mg/dL is considered to be lower than desirable.[23]

Serum LDL Cholesterol

Low-density lipoproteins (LDL) is rich in cholesterol. LDL delivers cholesterol to cells where it can be used for normal cell functions.[22] High levels of LDL is associated with reduced synthesis of LDL receptors which then can lead to excess cholesterol accumulation in blood.[22] LDL cholesterol in blood is calculated by subtracting the HDL cholesterol level and the VLDL cholesterol level from the total cholesterol level.[23] For a high risk individual, the recommended LDL level is less than 130 mg/dL.[23]

Estimated GFR

Serum and urine creatinine were measured by an enzymatic method and estimated glomerular filtration rate (eGFR) was computed using the Chronic Kidney Disease Epidemiology Collaborations (CKD-EPI) equation.[24] GFR is generally used as the best index of overall kidney function.[25] Decline in GFR is a hallmark of progressive kidney disease.[25] Less than 60 mL/min per 1.73 m^2 is considered as decreased GFR.[26] Less than 15 mL/min per 1.73 m^2 is defined as kidney failure.[26]

Urine ACR

Urine albumin-to-creatinine ratio (UACR) is used to estimate 24-hour proteinuria.[27] Proteinuria describes protein excretion in urine. Less than 150 mg/day is considered as normal level of total urinary protein excretion.[28] As UACR is a ratio of albumin to creatinine, it is measured in mg of protein per g of creatinine.[27] Greater than 30 mg/g is considered as an abnormally elevated ACR.[28] Individuals with UACR above this threshold is considered to be at high risk for chronic kidney disease.[28]

High Sensitivity CRP

High Sensitivity C-reactive protein (hsCRP) is an acute phase protein that is produced by hepatocytes and is a biomarker of inflammation.[29] Its pathogenic role in a specific cause for inflammation in the development of chronic kidney disease is currently unknown.[30] There is no standardized hsCRP value that is associated with abnormalities.[31] However, a study reported elevated hsCRP levels (>8 mg/dL) in 46 percent of its cohort (n=280) who were on chronic hemodialysis.[32] We note to take caution in interpreting the results regarding hsCRP given the lack of clinical understanding of its pathogenic role in chronic kidney disease.

Morisky Score

The Morisky score was assessed as a measure of adherence with prescribed medicines.[33] Higher score on the scale correspond to improved quality of life.

KDQDL Measures

Health related quality of life was assessed by the Kidney Disease Quality of Life survey (KDQOL-36).[34] Higher scores on the scales corresponded to improved quality of life.

Optimal Cutoffs for Continuous Measures

There were variables that displayed a ceiling effect where the response variable clustered toward the upper limit of the measurement. There were three such variables—Symptoms/Problems List, Effect of

Kidney Disease, and Burden of Kidney Disease. To dichotomize these variables, we first regressed the *Treatment* group against each variable via logistic regression modeling and secondly we used the area under the ROC curve (AUC) to identify the Euclidean distance that maximizes the difference between the ROC curve and the point (0, 1).[35] In turn, this allowed for minimizing the false positive rate while maximizing the true positive rate. We then applied this optimal cut-off threshold to categorize the three continuous variables into binary variables—1 indicating improvement of health-related quality of life and 0 indicating lack of improvement of health-related quality of life.

Models

Linear regression

Most of the dependent variables (DVs) that we were interested in were continuous with normal distribution. For example, the primary outcome of interest in this study was patient activation score which is a continuous variable. A linear regression model, using the SAS procedure PROC GENMOD, was used to estimate the treatment effect for the primary exposure on these DVs. In order to account for the household clustering, using Generalized Estimated Equations in PROC GENMOD rather than PROC REG was more appropriate.

Logistic regression

There were four variables of interest that were categorized into binary outcomes. For example, the primary outcome of interest, the patient activation level, was categorized into "activation" vs. "lack of activation". A logistic regression model, using the SAS procedure PROC GENMOD, was used to estimate the treatment effect for the primary exposure on these binary DVs. In order to account for the household clustering, using Generalized Estimated Equations in PROC GENMOD rather than PROC REG was more appropriate.

A cross-validation technique was used to examine the generalizability of our findings. We estimated cross-validation error using a Monte-Carlo approach. A training set (80% of the entire dataset)

was randomly selected from the dataset to construct the model. The remaining observations were collectively named the testing set (20% of the entire dataset). The fitted model was tested using the testing set. The matric used to evaluate the results from cross-validation were Root Mean Square Error (RMSE), the mean absolute error (MAE), and R^2 in the test set for the continuous DVs and Mean Percentage of Correct Classifications (MPCC) and its corresponding Standard Error (SE) for binary DVs.

Propensity Score

Propensity score is the probability of receiving treatment given a set of observed covariates. Traditionally, the propensity score was intended to be used specifically as a method of treatment selection bias reduction in non-randomized studies. In this thesis, the purpose of using the propensity score was to adjust for the imbalanced covariates at baseline to obtain more precise estimate for the treatment effect. A logistic regression model, for which the primary exposure was used as the primary outcome, was utilized to estimate the propensity scores. The estimated propensity scores were then used to generate Inverse Probability of Treatment Weights (IPTWs) that eventually were used to account for any imbalance in the covariates between groups at baseline.

Chapter 3

Analysis

The primary outcome in this study was patient activation score (PAM score) which was a continuous measure that ranges from 0 to 100. This was further dichotomized such that patients could be either in an activation level or in a lack of activation level (PAM level). Two models were fitted to analyze the data from the 12-month randomized controlled trial: (a) a logistic regression model was used to analyze the binary response variable "activation/lack of activation" in determining whether patients have higher odds of being activated in the management of their care, and (b) a linear regression model was used to analyze patient activation score that ranged from 0 to 100.

Data Visualization

Before any modeling, we visualized the collected data. We generated bar charts to visualize the changes in the patient activation levels from baseline to the 12-month measurements between groups. The bar charts in Figure 3.1 gave the frequency of participants in each patient activation levels in the control group and in the intervention group. The bar chart on the left gave the baseline measures and the one on the right gave the 12-month measures. The yellow bars represented control group and the blue bar represented intervention group.

At baseline, about 22% of the participants in the control group were in the "lack of activation" level and about 29% of the participants in the intervention group were in the "lack of activation" level. In other words, about 78% of the participants in the control group and about 71% of the participants in the intervention group were in the "activation" level at baseline. At the end of the 12-month randomized controlled trial (RCT), about 59% of the participants in the control group and about 92% of the participants in the intervention group were in the "activation" level.

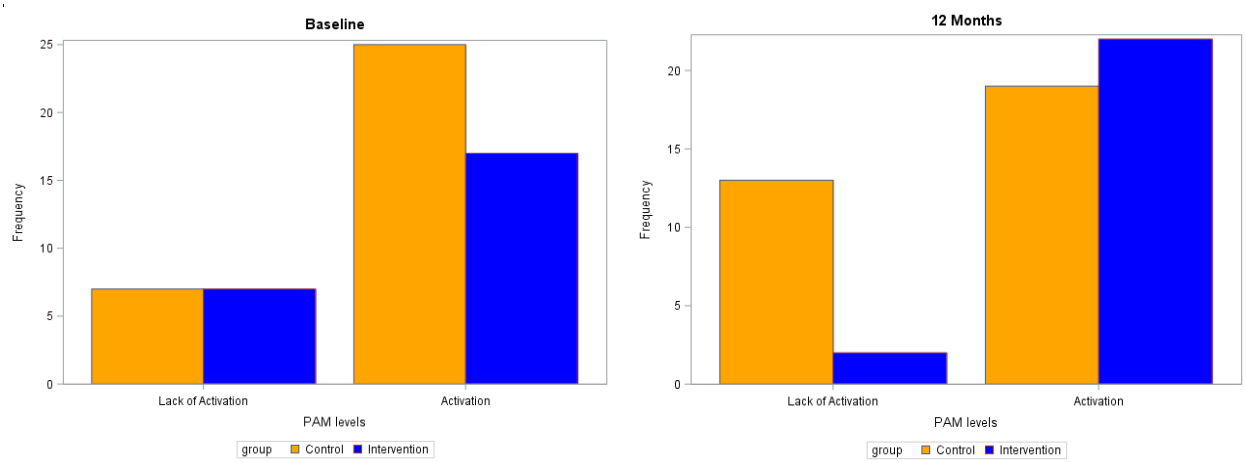


Figure 3.1 Percent of participants (control and intervention groups) in each PAM levels at baseline and 12 months

We generated histograms of the patient activation score (PAM score) differences between baseline measurement and the 12 month measurements showing the measured differences between the control group and the intervention group (Figure 3.2). The yellow represented the control group and the blue represented the intervention group. The positive differences were dominated by the blue whereas the negative differences were dominated by the yellow indicating higher improvement in the participants in the intervention group based on the PAM scores.

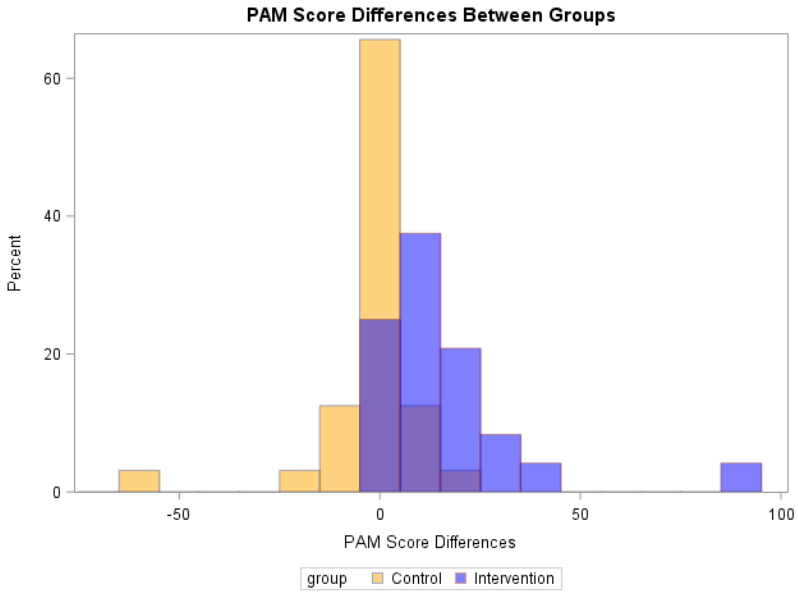


Figure 3.2 PAM score differences between groups.

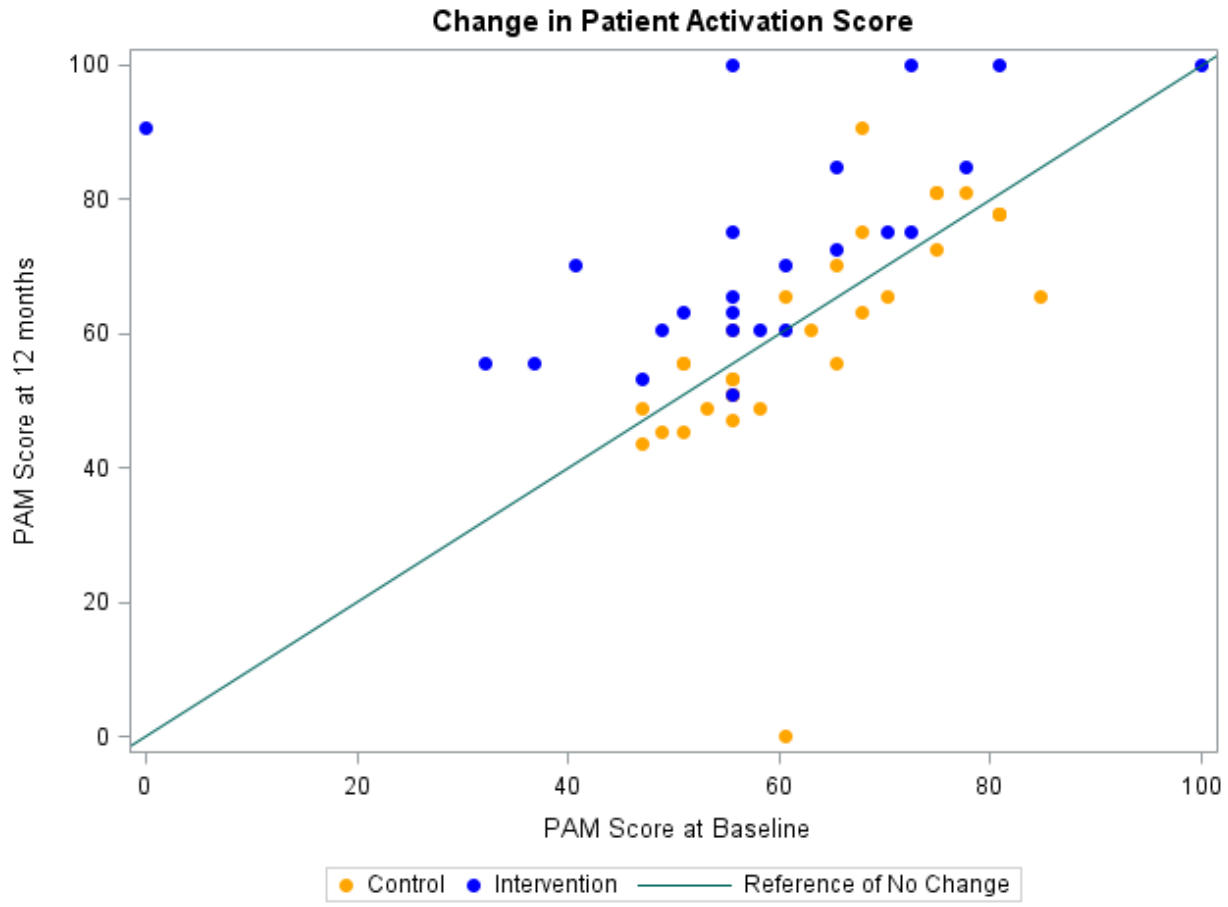


Figure 3.3 Change in PAM score between groups

A scatterplot of the PAM score at 12 months against the PAM score at baseline was generated to visualize the relationship between the baseline and the 12-month patient activation score between groups (Figure 3.3). The yellow filled dots represented the control group. The blue filled dots represented the intervention group. The diagonal line represented no change between the PAM scores from baseline to 12 months. If the PAM score increased, then it would be plotted above the line whereas if the PAM score decreased, then it would be plotted below the line. Mostly, the blue dots representing the intervention group were plotted above the line whereas the yellow dots representing the control group clustered around the reference line of no change. There might be a couple of outliers in the data; see further characterization of the outliers and influential points in chapter 3.

Baseline Comparison

To make a meaningful interpretation of the treatment effect, it is important to ascertain the participants in each group had similar baseline characteristics. The sample mean for numerical demographical variables, primary and secondary outcome measurements was reported as a measure of the central tendency along with the standard deviation as a measure of variability. As an informal test, we generated histograms to visualize the distribution of all of the variables by *Treatment* group. Some of the variables such as glucose, hemoglobin A1c, cholesterol, triglycerides, urine ACR, high sensitive CRP, the KDQDL measures displayed skewedness suggesting that the sample data may not be normally distributed.

Formally, we used the χ^2-tests to examine the association between any of the categorical variables with the *Treatment* groups. Further, to compare population means between the *Treatment* groups, we conducted two samples independent T-tests for continuous variables. To examine the assumptions of the latter test, we used the Folded F-test to determine the constant variance. If constant variance was established, then we reported the p-values from the pooled method. If constant variance was violated, then we reported the p-values from the Satterthwaite approximation method as it does not assume that variances of the two samples are equal. (Table 3.1)

Because of the skewedness of some of the variables in the data, we also reported the sample median and the interquartile range for the continuous variables (Table 3.2). We have conducted Wilcoxon Rank-Signed Tests (WRST) for median comparisons between the *Treatment* groups. The p-values from the WRST tests were reported. The sample median and the interquartile range are less sensitive to extreme observations compared to the sample mean and the standard deviation and hence Table 3.2 might provide better estimates for central tendency and variability of the data compared to Table 3.1. All comparisons were made based on the 5% significance level.

	Usual Care, n=32	Home-Based Renal Care, n=24	
Characteristic	**Mean (±SD) or n (%[a])**	**Mean (±SD) or n (%)**	**p-value[b]**
Demographics			
Age, yr	50.9 (±13.0)	47.4 (±11.2)	0.2871
Women	16 (50%)	16 (66.7%)	0.2785
High school education	22 (68.8%)	17 (70.8%)	1.0000
Primary outcome measures			
Patient activation total score	63.0 (±11.2)	57.3 (±19.1)	0.1999[c]
Patient activation level ≥ 3	25 (78.1%)	17 (70.8%)	0.5507
Secondary outcome measures			
Body mass index, kg/m^2	31.4 (±7.0)	32.8 (±7.8)	0.4877
BP, mm HG			
Systolic	126.6 (±15.5)	128 (±12.2)	0.6556
Diastolic	81.8 (±10.5)	80.3 (±12.9)	0.6465
HbA1c, %	8.5 (±2.3)	9.1 (±2.6)	0.3697
Glucose, mg/dl	167.3 (±81.2)	182.9 (±92.4)	0.5035
Serum total protein, mg/dl	7.5 (±0.5)	7.6 (±0.5)	0.6275
Serum cholesterol, mg/dl	175.3 (±35.7)	220.9 (±69.9)	0.0064[c]
Serum triglycerides, mg/dl	180.9 (±92.6)	411.1 (±656.3)	0.1011[c]
Serum HDL cholesterol, mg/dl	45.8 (±1.4)	46.6 (±1.5)	0.6505[c]
Serum LDL cholesterol, mg/dl	105.0 (±30.5)	125.0 (±42.5)	0.0444
eGFR, ml/min per 1.73 m^2	115.1 (±63.7)	148.9 (±63.2)	0.0540
Urine ACR, mg/g	1082.0 (±1631.6)	598.3 (±1372.1)	0.2458
hsCRP, mg/L	3.6 (±3.6)	10.7 (±13.2)	0.0191[c]
Morisky score[d]	3.9 (±2.2)	4.3 (±1.8)	0.4109
KDQDL measures			
Symptom/ problem list	84.0 (±13.1)	82.2 (±13.8)	0.6232
Effects of kidney disease	90.0 (±11.8)	92.1 (±7.2)	0.4215[c]
Burden of kidney disease	71.2 (±24.8)	67.4 (±19.1)	0.5460
SF-12 physical score	44.9 (±8.3)	45.2 (±9.2)	0.9331
SF-12 mental score	49.9 (±10.1)	45.4 (±11.7)	0.1243

Table 3.1 Baseline characteristics of the participants by treatment group given by the mean

[a] %=column percentage

[b] P-value corresponds to the two sample independent T-test for continuous variables and the χ^2 test for categorical variables.

[c] an indication of using the Satterthwaite p-value for having the constant variance assumption violated.

[d] Use of the Morisky Medication Adherence Scale is protected by United States copyright laws. Permission for use is required. A license agreement is available from Donald E. Morisky, Department of Community Health Sciences, University of California, Los Angeles School of Public Health, 650 Charles E. Young Drive South, Los Angeles, CA 90095-1772.

Characteristic	Usual Care, n=32 Median (Q1,Q3) or n (%[a])	Home-Based Renal Care, n=24 Median (Q1,Q3) or n (%)	p-value[b]
Demographics			
Age, yr	49.5 (43.5, 56.0)	48.5 (41.5, 54.5)	0.5397
Women	16 (50%)	16 (66.7%)	0.2785
High school education, %	22 (68.8%)	17 (70.8%)	1.0000
Primary outcome measures			
Patient activation total score	60.6 (55.6, 72.6)	57.3 (50.0, 67.9)	0.2603
Patient activation level ≥ 3	25 (78.1%)	17 (70.8%)	0.5507
Secondary outcome measures			
Body mass index, kg/m^2	31.0 (26.7, 34.4)	33.6 (26.3, 36.5)	0.3892
BP, mm HG			
Systolic	122.3 (118.7, 139.7)	130.7 (119.5, 137.7)	0.5291
Diastolic	80.3 (77.0, 86.0)	79.7 (70.3, 92.0)	0.6309
HbA1c, %	8.1 (6.1, 10.2)	9.4 (6.8, 11.3)	0.4265
Glucose, mg/dl	134.5 (118.7, 139.7)	149.5 (106.4, 239.2)	0.4816
Serum total protein, mg/dl	7.5 (7.3 ,7.8)	7.6 (7.2, 8.0)	0.8117
Serum cholesterol, mg/dl	174.0 (144.5, 198.5)	197.0 (172.0, 247.5)	0.0113
Serum triglycerides, mg/dl	169.4 (92.0, 235.5)	162.6 (122.4, 270.7)	0.5131
Serum HDL cholesterol, mg/dl	45.5 (36.5, 57.0)	45.5 (38.0, 55.0)	0.8554
Serum LDL cholesterol, mg/dl	99.0 (86.0, 127.0)	117 (107.5, 140.5)	0.0228
eGFR, ml/min per 1.73 m^2	111.4 (68.3, 168.8)	139.7 (114.9, 179.3)	0.0698
Urine ACR, mg/g	348.8 (75.9, 992.1)	135.5 (80.2, 263.9)	0.1382
hsCRP, mg/L	2.1 (1.3, 5.4)	6.3 (0.7, 11.1)	0.1133
Morisky score[c]	3.6 (2.1, 5.8)	4.5 (2.8, 4.8)	0.3779
KDQDL measures			
Symptom/ problem list	86.4 (76.1, 92.0)	86.4 (77.3, 93.2)	0.6749
Effects of kidney disease	95.3 (81.3, 100.0)	93.8 (87.5, 96.9)	0.9517
Burden of kidney disease	68.8 (50.0, 100.0)	62.5 (50.0, 87.5)	0.4914
SF-12 physical score	45.2 (40.1, 50.9)	48.6 (40.9, 51.8)	0.6511
SF-12 mental score	52.7 (43.2, 57.9)	48.4 (35.7, 55.1)	0.1069

Table 3.2 Baseline characteristics of the participants by treatment group given by the median

[a] %=column percentage

[b] P-value corresponds to the Wilcoxon Rank Sum Test for continuous variables and the χ^2 test for categorical variables.

[c] Use of the Morisky Medication Adherence Scale is protected by United States copyright laws. Permission for use is required. A license agreement is available from Donald E. Morisky, Department of Community Health Sciences, University of California, Los Angeles School of Public Health, 650 Charles E. Young Drive South, Los Angeles, CA 90095-1772.

Generalized Estimating Equations

Generalized Estimating Equations (GEE) were introduced by Liang and Zeger (1985).[36] GEE is not a likelihood-based method. It uses pseudo maximum likelihood (PML) method rather than maximum likelihood (ML) to estimate the model parameters. The PML method is based on the exponential family. When estimating parameters using the ML method, one must have the correct specification of the likelihood function. If misspecified, ML estimation could result in invalid conclusions. On the contrary, PML method allows for consistent estimation of the mean structure even if the covariance structure is misspecified. However, the efficiency is lowered when the covariance structure is misspecified.[37]

GEE arises from normality-based log-likelihood without assuming the response is normally distributed.[38] GEE were appropriate to use for the home-based renal care (HBRC) data analysis because we collected repeated measurements of variables of interest over time. GEE allows for modelling potentially correlated data accounting for the household clustering. It was appropriate to model both categorical and continuous response variables using GEE.

Pseudo Maximum Likelihood (PML) Method

The PML method is a generalization of the ML method such that it allows for a partial model misspecification of a density from the linear exponential family.[27] When using PML estimation, only the mean structure must be correctly specified.[27] In GEE, one can choose the mean structure by using a link function from the generalized linear model.[27] Further, the existence of a variance matrix is assumed but the correct specification of the covariance matrix is not required.[27] Also, we assume that the pseudo distribution is in the linear exponential family with fixed nuisance parameter.[27] A nuisance parameter is any parameter that is not of direct interest but must be accounted for in the analysis to estimate the parameters of direct interest. PML is computed by first replacing the nuisance parameters in the model by estimates which in turn reduces the system of likelihood equations (hence called partial ML) and secondly solving the reduced system of equations for the parameters of interest (i.e. the non-nuisance parameters).[39]

In essence, GEE does not specify a complete joint distribution—rather, it uses a marginalized joint distribution in estimating the model parameters.[27] It is important to note that if covariance structure is misspecified, the standard errors will be underestimated.[27] When the covariance structure is correctly specified, the PML solution will agree with the ML estimation.

Covariance Structure

GEE allows for heterogeneity of variance. To specify the covariance structure, firstly we obtained the Quasi-likelihood under the Independence model Criterion (QIC)[40] using the saturated model under each covariance structure type including exchangeable (EXCH), compound symmetry (CS), unstructured (UN) and autoregressive (AR) covariance structure patterns. Secondly, we compared the obtained QICs from all conducted saturated models, covering all covariance structure types, such that a covariance structure model with the smallest QIC was deemed to be the most adequate.[40] In the HBRC study, we used EXCH covariance structure because it gave the smallest QIC score (Table 3.3). The QIC score is analogous to AIC score used for fitting likelihood-based methods. Hence, a small QIC indicates a good fit of the model.

Covariance Pattern Model	QIC
EXCH	52.3621
CS	52.3621
UN	52.3789
AR	52.6008

Table 3.3 Covariance Structure given by the smallest QIC score

Logistic Regression Model

We had four binary outcomes in this study. The primary outcome of interest was PAM level where 1 indicated "activation" and 0 indicated "lack of activation" of patient-involvement in the management of their care. The other three secondary binary outcomes were the following KDQDL

measures, where 1 indicated improvement in quality of health: (i) Symptom/ Problem List, (ii) Effects of Kidney Disease, and (iii) Burden of Kidney Disease. Multiple logistic regression models were employed to examine the effect of the primary exposure (i.e. home-based renal care) on the odds of event of interest in these outcomes while adjusting for (A) outcome levels at baseline only, and (B) outcome levels and imbalanced covariates at baseline including Cholesterol, LDL, and hsCRP. The functional form for Model-A and Model-B were presented below and illustrated only for the primary outcome noting that the same functional form was applied for the secondary outcomes.

Model-A: Adjusting for outcome levels at baseline only:

When modeling PAM level, the outcome response was PAM level at 12 months, the primary exposure was *Treatment* group, and the adjusted-for IV was PAM level at baseline. The results from these logistic models were presented in the bottom of Table 3.6.

Let Z be an indicator of the binary *Treatment* with 1 for home-based renal care (HBRC) and 0 for the standard clinical care (SC), X_1 be the baseline level of the primary outcome of interest, and Y be the primary outcome of interest. Then, mathematically, the logistic regression model for estimating the treatment effect was:

$$[1]\ \text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \gamma Z + \beta_1 X_1 + \varepsilon$$

$$\Leftrightarrow \pi = \frac{\exp(\beta_0 + \gamma Z + \beta_1 X_1 + \varepsilon)}{1 + \exp(\beta_0 + \gamma Z + \beta_1 X_1 + \varepsilon)},$$

where γ was the estimated parameter for the treatment effect and $\pi = \Pr(Y=1)$.

Model-B: Adjusting for outcome levels and imbalanced covariates at baseline:

When modeling PAM level, the outcome response was PAM level at 12 months, the primary exposure was *Treatment* group, and the IVs that were adjusted for were PAM level at baseline, Cholesterol level (continuous measure), LDL (continuous measure), and hsCRP (continuous measure)—

the imbalanced baseline covariates. The results from these logistic models were presented in bottom of Table 3.7.

Let Z be an indicator of the binary *Treatment* with 1 for home-based renal care (HBRC) and 0 for the standard clinical care (SC), X_1 be the baseline level of the primary outcome of interest, X_2 be the baseline cholesterol level, X_3 be the baseline LDL level, X_4 be the baseline hsCRP level, and Y be the primary outcome of interest. Then, mathematically, the logistic regression model for estimating the treatment effect was:

$$[2]\ \text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \gamma Z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$\Leftrightarrow \pi = \frac{\exp(\beta_0 + \gamma Z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon)}{1 + \exp(\beta_0 + \gamma Z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon)},$$

where γ was the estimated parameter for the treatment effect and $\pi = \Pr(Y=1)$.

Checking Assumptions

For our logistic regression analysis, we used GEE that do not assume that the responses are independent. Though it was not necessary to check independence, normality and constant variance for GEE models, we still examined the QQ-plots for normality because GEE is more efficient when estimating parameters for normally distributed data.

Outliers/Influential Points

We checked the standardized Pearson chi-square residuals to identify outlying points on the Y direction, leverage to determine outlying points on the X direction outliers, and Cook's distance to determine influential points. Standardized Pearson's chi-square residuals in the absolute value that were larger than 3.84 (i.e. critical value of the Chi-square distribution with one degrees of freedom) were deemed outliers in the Y direction while leverage values that exceeded $\frac{2p}{n}$, where p was the number of parameters, were deemed outliers in the X direction.

For the HBRC data, we defined X outliers to be any observations with leverage greater than 0.0714 for Model-A and 0.1786 for Model-B. We further defined influential points to be any observations with Cook's D greater than 1.[41] Table 3.4 showed the values of standardized residuals, leverage, and Cook's D for the potential outliers or influential points identified for both models when modeling PAM level. We note that Standardized Pearson's chi-square residuals were calculated with respect to the data without clustering. Table 3.4 revealed one outlying point in the Y direction for Model-A and 14 outlying points on the X direction for Model-A and 7 for Model-B. No influential points were found in both models according to Cook's D.

		ID	PAM level baseline	PAM level 12 months	Group	Standardized residuals	Leverage	Cook's D
Model-A	1	2441-1	1	0	Intervention	**4.640**	0.034	0.252
	2	1008-1	0	0	Control	0.724	**0.118**	0.029
	3	1008-2	0	0	Control	0.724	**0.118**	0.029
	4	1071-4	0	1	Intervention	-0.484	**0.096**	0.009
	5	1123-3	0	0	Control	0.724	**0.113**	0.017
	6	1341-1	0	1	Control	-1.558	**0.111**	0.107
	7	1581-1	0	0	Control	0.724	**0.111**	0.022
	8	1654-2	0	1	Intervention	-0.484	**0.102**	0.011
	9	179-1	0	1	Intervention	-0.484	**0.096**	0.009
	10	1922-2	0	1	Intervention	-0.484	**0.096**	0.009
	11	2064-2	0	0	Intervention	2.292	**0.101**	0.209
	12	2165-6	0	1	Intervention	-0.484	**0.096**	0.009
	13	332-1	0	1	Intervention	-0.484	**0.096**	0.009
	14	383-5	0	0	Control	0.724	**0.111**	0.022
	15	961-6	0	1	Control	-1.558	**0.111**	0.090
Model-B	1	1105-1	1	1	Intervention	-0.644	**0.523**	0.080
	2	1341-1	0	1	Control	-1.499	**0.207**	0.105
	3	1654-2	0	1	Intervention	-0.463	**0.512**	0.041
	4	2064-2	0	0	Intervention	3.137	**0.196**	0.410
	5	2165-6	0	1	Intervention	-0.511	**0.194**	0.012
	6	2441-1	1	0	Intervention	0.516	**0.554**	0.055
	7	961-6	0	1	Control	-1.212	**0.233**	0.058

Table 3.4: Possible outlying points in the X and Y direction and influential points when modeling PAM level for Model-A and Model-B.

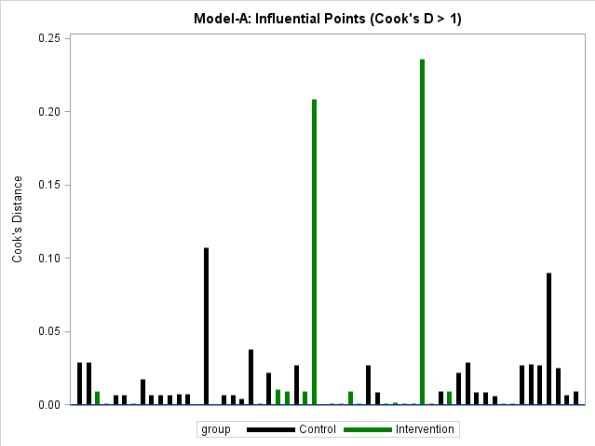


Figure 3.4 Identifying influential points in Model-A when modeling PAM level

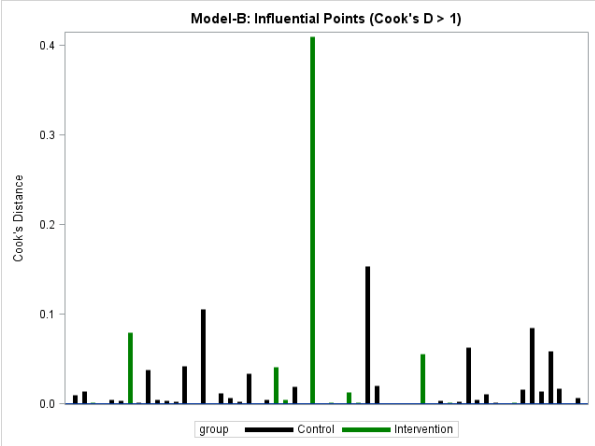


Figure 3.5 Identifying influential points in Model-B when modeling PAM level

Transformation: Categorization

Three variables (secondary outcomes) in the kidney disease quality of life survey (KDQOL) measures displayed ceiling effect where the primary exposure no longer had an effect on them. For ease of interpretation and to arrive at a more meaningful conclusion, these three continuous variables were transformed into binary variables. Symptom/Problems List, Effects of Kidney Disease, and Burden of Kidney Disease were each categorized into binary variables “improvement/lack of improvement”.

There were not clinically relevant thresholds for determining improvements in each of these variables. Instead, we first identified optimal thresholds using QQ-plots as an informal technique by identifying points at which the curvature changes or inflection points occur. Secondly, we confirmed these thresholds formally by using the area under the ROC curve (AUC). This formal technique relies on either minimizing the Euclidean distance of the ROC curve from the point (0, 1) or maximizing the Youden Index.[42] These two distances that either minimized *D* or maximized *J* were illustrated in Figure 3.6.

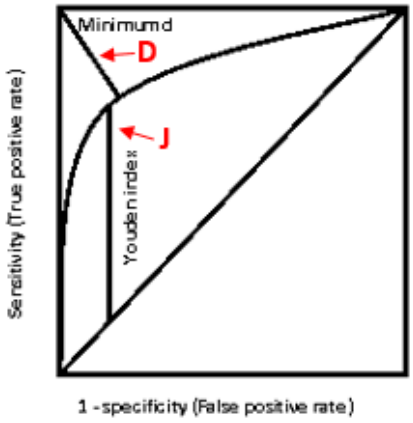


Figure 3.6 Using the ROC curve to find the optimal threshold point

Using the ROC curve to find the optimal "cut-off" threshold

1. *Maximizing the Youden Index*

One way to find the optimal cut-off point is to maximize the Youden Index denoted by *J*. Let S_n be sensitivity and S_p be specificity, then $J = \max [S_n + S_p]$.[42] The idea is to find a point where the distance between the Y = X line and the ROC curve is maximized. In turn, this maximizes the difference between the true positive rate and the false positive rate.

2. *Minimizing the Euclidean distance*

Another way to find the optimal cut-off point is to minimize the Euclidean distance denoted by *D*. *D* is the distance between (0, 1) and the ROC curve and is given by $D = \sqrt{(1 - \mathrm{Sn})^2 + (1 - \mathrm{Sp})^2}$.[42] This would also maximize the difference between the true positive rate and the false positive rate.

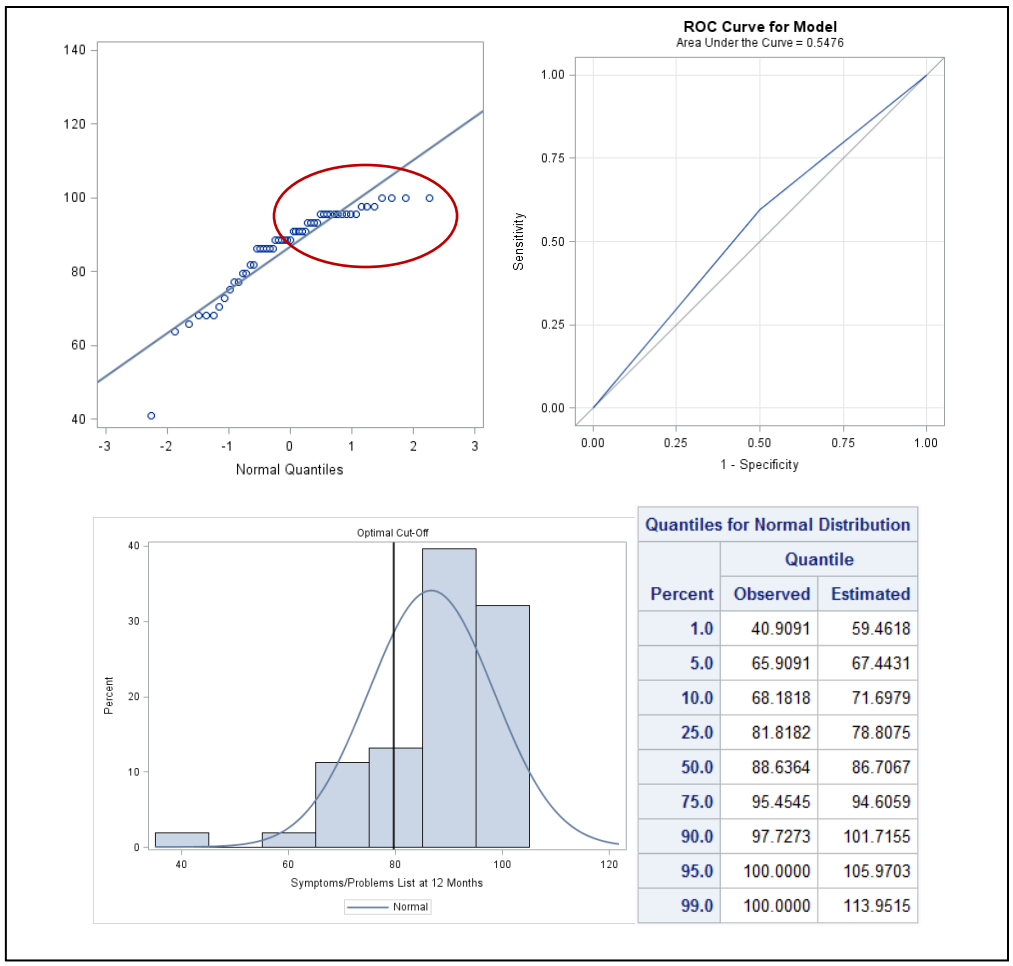
We used the 12 month measurements to generate the ROC curve then calculated both *J* and *D* to determine the optimal cut-off points for the three variables. If they were in disagreement, we used the minimized *D* as the optimal cut-off point. For Symptoms/Problems List, Effect of Kidney Disease, and Burden of Kidney Disease, probability of 0.438 corresponded to a score of 79.545, probability of 0.439

corresponded to a score of 96.875 and probability of 0.462 corresponded to a score of 87.5 respectively. This was presented in Table 3.5.

	Probability at which *J* maximized	**Probability at which *D* is minimized**	**Optimal Cut-off**
Symptoms/Problems List	0.379	0.438	79.545
Effect of Kidney Disease	0.439	0.439	96.875
Burden of Kidney Disease	0.462	0.462	87.5

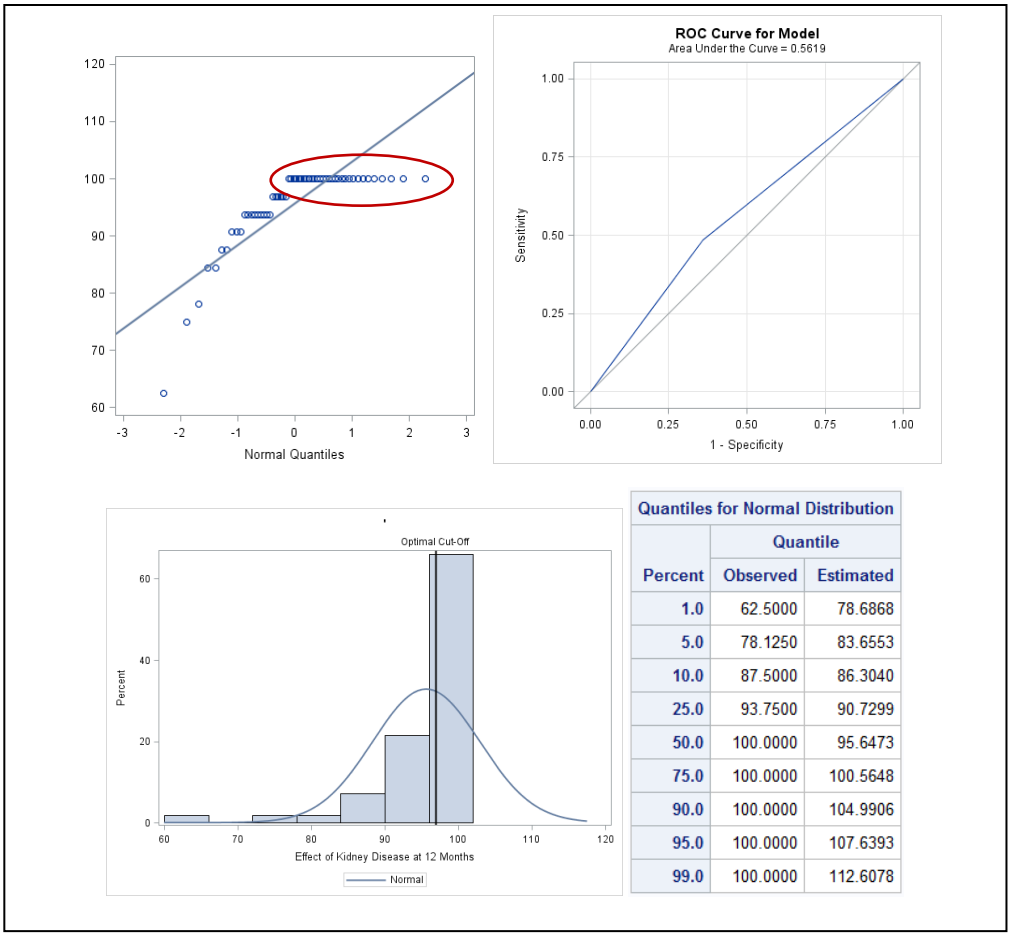
Table 3.5 Corresponding probabilities with corresponding optimal cut-offs at maximized *J* and at minimized *D.*

Figure 3.7 visually illustrates the formal and informal ways of identifying the optimal threshold for Symptoms/Problems List at 12 months. In particular, the QQ-plot in the top left panel showed an inflection point around the 25th percentile which was also highlighted in the distribution/histogram of the variable at the bottom left panel at 80. Formally, this threshold was indeed found to give the minimal Euclidean distance *D* between the ROC curve and the (0, 1) point. Note that quantiles of this variable emphasized the ceiling effect and hence justifying the use of optimal cut-off.



Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	40.9091	59.4618
5.0	65.9091	67.4431
10.0	68.1818	71.6979
25.0	81.8182	78.8075
50.0	88.6364	86.7067
75.0	95.4545	94.6059
90.0	97.7273	101.7155
95.0	100.0000	105.9703
99.0	100.0000	113.9515

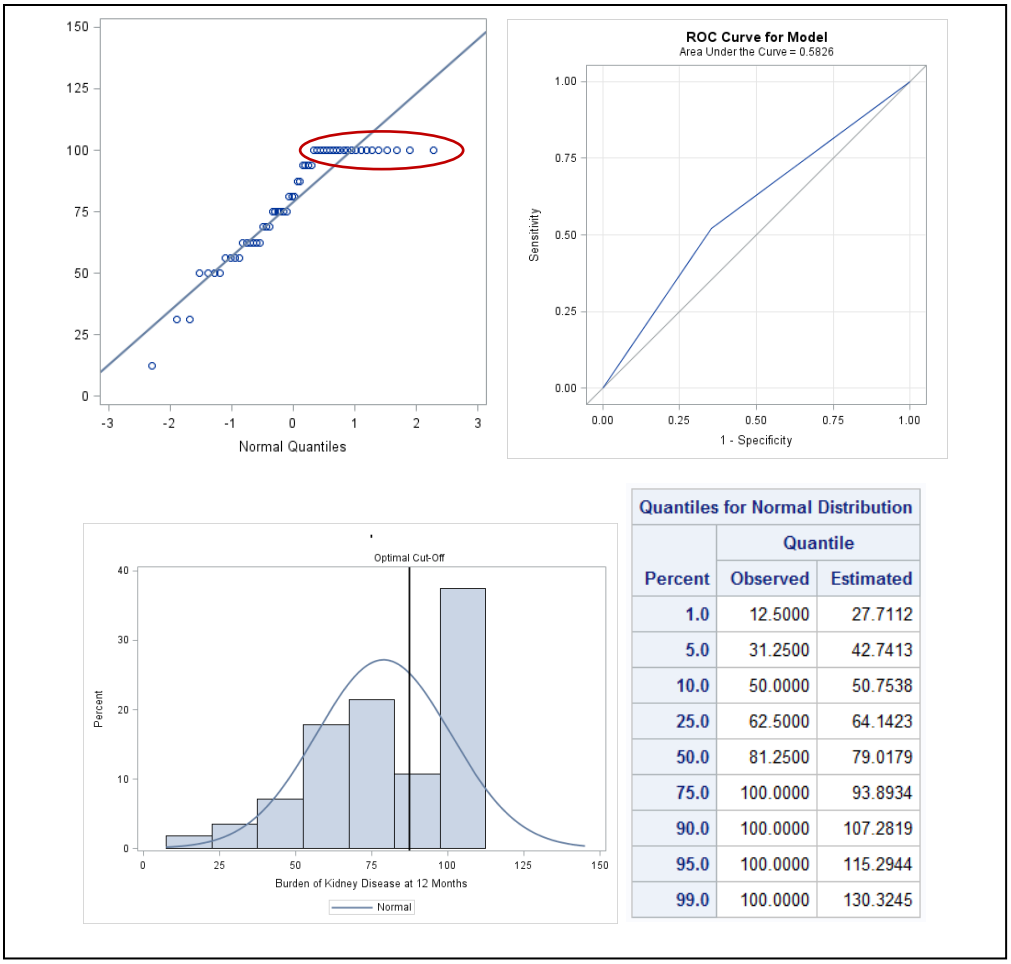
Figure 3.7 Identifying optimal cut-off for Symptoms/Problems List. *Top left panel*: QQ-plot of Symptoms/Problems List at 12 months, *Top right panel*: ROC curve indicating classification between improvement and lack of improvement for the Symptoms/Problems List based on optimal cut-off between *Treatment* groups; *Bottom left panel*: The distribution of Symptoms/Problems List at 12 months with a reference line at the optimal threshold, and *Bottom right panel*: Quantiles of the Symptoms/Problems List at 12 months.



Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	62.5000	78.6868
5.0	78.1250	83.6553
10.0	87.5000	86.3040
25.0	93.7500	90.7299
50.0	100.0000	95.6473
75.0	100.0000	100.5648
90.0	100.0000	104.9906
95.0	100.0000	107.6393
99.0	100.0000	112.6078

Figure 3.8 Identifying optimal cut-off for Effect of Kidney Disease. *Top left panel*: QQ-plot of Effect of Kidney Disease at 12 months, *Top right panel*: ROC curve indicating classification between improvement and lack of improvement for the Effect of Kidney Disease based on optimal cut-off between *Treatment* groups; *Bottom left panel*: The distribution of Effect of Kidney Disease at 12 months with a reference line at the optimal threshold, and *Bottom right panel*: Quantiles of the Effect of Kidney Disease at 12 months.

Figure 3.8 visually illustrates the formal and informal ways of identifying the optimal threshold for Effect of Kidney Disease at 12 months. Informally, in the QQ-plot in the top left panel showed an inflection point around the 50th percentile which was also highlighted in the distribution/histogram of the variable at the bottom left panel at 95. Formally, this threshold was indeed found to give the minimal Euclidian distance *D* between the ROC curve and the point (0, 1) and the maximal Youden Index *J* between the ROC curve and Y = X line. Note that quantiles of this variable emphasized the ceiling effect and hence justifying the use of optimal cut-off.



Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	12.5000	27.7112
5.0	31.2500	42.7413
10.0	50.0000	50.7538
25.0	62.5000	64.1423
50.0	81.2500	79.0179
75.0	100.0000	93.8934
90.0	100.0000	107.2819
95.0	100.0000	115.2944
99.0	100.0000	130.3245

Figure 3.9 Identifying optimal cut-off for Burden of Kidney Disease. *Top left panel*: QQ-plot of Burden of Kidney Disease at 12 months, *Top right panel*: ROC curve indicating classification between improvement and lack of improvement for the Burden of Kidney Disease based on optimal cut-off between *Treatment* groups; *Bottom left panel*: The distribution of Burden of Kidney Disease at 12 months with a reference line at the optimal threshold, and *Bottom right panel*: Quantiles of the Burden of Kidney Disease at 12 months.

Figure 3.9 visually illustrates the formal and informal ways of identifying the optimal threshold for Burden of Kidney Disease at 12 months. Informally, in the QQ-plot in the top left panel showed an inflection point between the 50th and 70th percentile which was also highlighted in the distribution/ histogram of the variable at the bottom left panel at 85. Formally, this threshold was indeed found to give the minimal Euclidian distance *D* between the ROC curve and the point (0, 1) and the maximal Youden Index *J* between the ROC curve and Y = X line. Note that quantiles of this variable emphasized the ceiling effect and hence justifying the use of optimal cut-off.

Linear Regression Model

The primary outcome of interest, PAM score, was a continuous response variable ranging from 0 to 100. There were also 16 other continuous variables (secondary outcomes) of interest including but not limited to body mass index (BMI) and cholesterol. To obtain the treatment effect, multiple linear regression models were employed to examine the effect of the primary exposure (i.e. home-based renal care) on the mean of outcome of interest while adjusting for (A) outcome scores at baseline only, and (B) outcome scores and imbalanced covariates at baseline including Cholesterol, LDL, and hsCRP. The functional form for Model-A and Model-B were presented below and illustrated only for the primary outcome noting that the same functional form was applied for the secondary outcomes.

Model-A: Adjusting for outcome scores at baseline only:

When modeling PAM score, the outcome response was PAM score at 12 months minus the baseline measurements for PAM, the primary exposure was *Treatment* group, and IV that was adjusted-for was PAM scores at baseline. The results from these linear models were presented in the top of Table 3.6.

Let Z be an indicator of the binary *Treatment* with 1 for home-based renal care (HBRC) and 0 for the standard clinical care (SC), X_1 be the baseline measurement of the continuous variable of interest and Y be the difference between the 12-month measurement and the baseline measurement. Mathematically, the linear regression model for estimating the treatment effect was:

$$[3]\ Y = \beta_0 + \gamma Z + \beta_1 X_1 + \varepsilon\ ,$$

where γ is the estimated parameter for the treatment effect.

Model-B: Adjusting for outcome scores and imbalanced covariates at baseline:

When modeling PAM score, the outcome response was PAM score at 12 months minus PAM score at baseline, the primary exposure was *Treatment* group, and our adjusted-for IVs were PAM score

at baseline, Cholesterol level (continuous measure), LDL (continuous measure), and hsCRP (continuous measure)—the imbalanced baseline covariates. The results from these linear regression models were presented in top of Table 3.7.

Let Z be an indicator of the binary *Treatment* with 1 for home-based renal care (HBRC) and 0 for the standard clinical care (SC), X_1 be the baseline scores of the primary outcome of interest, X_2 be the baseline cholesterol level, X_3 be the baseline LDL level, X_4 be the baseline hsCRP level, and Y be the primary outcome of interest (i.e. the difference between measurement at 12 month and baseline). Then, mathematically, the linear regression model for estimating the treatment effect was:

$$[4]\ Y = \beta_0 + \gamma Z + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon,$$

where γ is the estimated parameter for the treatment effect.

Log Transformation

Even though normality and constant variance were not required assumptions due to using GEE, we still log-transformed variables that were not normally distributed to improve efficiency as GEE is more efficient with normal data.[43] We log-transformed three variables—Triglycerides, Urine ACR and High sensitivity CRP. We note that the interpretation of the effect switches from change in units in the outcome to percentage change (i.e. $100\%*\hat{\gamma}$) instead due to the log-transformation of the outcome.[44]

Estimating Treatment Effect using Linear Regression and Logistic Regression

We constructed Table 3.6 of estimated treatment effect with respect to all of the variables of interest. For the estimated parameters from the logistic regression model, at $\alpha = 0.05$, the testing hypotheses were:

$$\mathbf{H_0}: \text{OR} = 1 \text{ vs. } \mathbf{H_1}: \text{OR} \neq 1.$$

For estimated patient activation level, we rejected the null hypotheses and concluded that on average, the odds of patient activation level were 9.7 times higher in the intervention group compared to the control

group when adjusting for outcome baseline levels (OR=9.7; 95% C.I. 1.7-54.3) and 13.2 times higher in the intervention group compared to the control group when adjusting for outcome baseline levels and imbalanced covariates at baseline (OR=13.2; 95% C.I. 1.1-166.1).

For the estimated parameters from the linear regression model, the testing hypotheses, at $\alpha = 0.05$, were

$$\mathbf{H_0}: \gamma = 0 \text{ vs. } \mathbf{H_1}: \gamma \neq 0.$$

For estimated patient activation total score, we rejected the null hypothesis and concluded that on average, patient activation score was 16 points higher in the intervention group compared to the control group when adjusting for outcome baseline scores ($\hat{\gamma}$=16; 95% C.I. 8.8-23.1; p<0.0001), and 15.7 points higher in the intervention group compared to the control group when adjusting for outcome baseline scores and imbalanced covariates at baseline ($\hat{\gamma}$=15.7; 95% C.I. 7.6-23.8). For estimated body mass index (BMI), we rejected the null hypothesis and concluded that on average, BMI was 1.2 lower in the intervention group compared to the control group when adjusting for outcome baseline scores ($\hat{\gamma}$=-1.2; 95% C.I. -2.2 to -0.2; p=0.0212), and -1.0 points lower in the intervention group compared to the control group when adjusting for outcome baseline scores and imbalanced covariates at baseline ($\hat{\gamma}$=-1.0; 95% C.I. -2.0 to -0.1; p=0.0392). For high sensitivity CRP, we rejected the null hypothesis and concluded that on average, hsCRP was lower in the intervention group compared to the control group by 70% ($\hat{\gamma}$=-0.7; 95% C.I. -1.0 to -0.3; p<0.0001), and similarly when adjusting for outcome baseline scores and imbalanced covariates at baseline ($\hat{\gamma}$=-0.7; 95% C.I. -1.0 to -0.3; p=0.0001). For all other variables, we failed to reject the null hypothesis. We note that when running the same models while removing the influential points we obtained similar results as seen in the right most columns of Tables 3.6 and 3.7.

		With Influential Point		# of Influential points	Without Influential Points	
	Characteristic	**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**		**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**
Linear Regression	**Primary outcome measures**					
	Patient activation total score	***16 (8.8 to 23.1)***	***<.0001***	1	***14.5 (8.1 to 20.9)***	***<.0001***
	Secondary outcome measures					
	Body mass index, kg/m^2	***-1.2 (-2.2 to -0.2)***	***0.0212***	0		
	BP, mm HG					
	Systolic	-3.7 (-11.2 to 3.7)	0.3237	5	1.6 (10.3 to 7.1)	0.7120
	Diastolic	-1.9 (-7.4 to 3.6)	0.4954	0		
	HbA1c, %	-0.8 (-1.9 to 0.4)	0.1896	2	***-1.1 (-2.0 to -0.20)***	***0.0161***
	Glucose, mg/dl	-3 (-47.5 to 41.6)	0.8957	0		
	Serum total protein, mg/dl	-0.1 (-0.3 to 0.1)	0.3495	0		
	Serum cholesterol, mg/dl	4.8 (-15.9 to 25.5)	0.6513	0		
	Serum triglycerides, mg/dl[a]	-0.1 (-0.3 to 0.2)	0.5573	0		
	Serum HDL cholesterol, mg/dl	-0.1 (-7.9 to 7.7)	0.9706	0		
	Serum LDL cholesterol, mg/dl	4.9 (-11.2 to 21)	0.5529	0		
	eGFR, ml/min per 1.73 m^2	9.1 (-12.6 to 30.9)	0.4110	0		
	Urine ACR, mg/g[a]	-0.5 (-1.2 to 0.3)	0.2631	0		
	hsCRP, mg/L[a]	***-0.7 (-1 to -0.3)***	***<.0001***	0		
	Morisky score[b]	-0.2 (-1.1 to 0.8)	0.6903	0		
	KDQDL measures					
	SF-12 physical score	2 (-3.2 to 7.1)	0.4531	0		
	SF-12 mental score	1.2 (-3.3 to 5.8)	0.5975	5	2.3 (-1.9 to 6.4)	0.2841
Logistic Regression	**Primary outcome measures**	**OR (95% C.I.)**	**p-value**			
	Patient activation level ≥ 3	***9.7 (1.7 to 54.3)***	***0.0097***	0		
	Secondary outcome measures					
	KDQDL measures					
	Symptom/ problem list	0.8 (0.2 to 2.9)	0.7209	0		
	Effects of kidney disease	2.5 (0.7 to 8.7)	0.1519	0		
	Burden of kidney disease	*2.7 (0.9 to 8.3)*	*0.0862*	0		

Table 3.6 Estimated treatment effect of intervention compared control for Model-A with and without influential points

		With Influential Point		# of Influential points	Without Influential Points	
	Characteristic	**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**		**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**
Linear Regression	**Primary outcome measures**					
	Patient activation total score	***15.7 (7.6 to 23.8)***	***0.0001***	1	***14.0 (7.5 to 20.5)***	***<.0001***
	Secondary outcome measures					
	Body mass index, kg/m^2	***-1.0 (-2.0 to -0.1)***	***0.0392***	0		
	BP, mm HG					
	Systolic	-4.4 (-11.5 to 2.6)	0.2182	5	-2.3 (-12.7 to 8.0)	0.6569
	Diastolic	-4.5 (-10.3 to 1.2)	0.1207	4	-0.9 (-6.5 to 4.6)	0.7415
	HbA1c, %	-0.2 (-1.1 to 0.8)	0.7098	5	-0.6 (-1.6 to 0.4)	0.2374
	Glucose, mg/dl	10.8 (-40.2 to 61.7)	0.6790	0		
	Serum total protein, mg/dl	*-0.2 (-0.4 to 0)*	*0.0712*	5	-0.1 (-0.4 to 0.3)	0.6956
	Serum cholesterol, mg/dl	12 (-10 to 34.1)	0.2841	2	2.2 (-22.3 to 26.8)	0.8590
	Serum triglycerides, mg/dl[a]	0.1 (-0.2 to 0.4)	0.6466	0		
	Serum HDL cholesterol, mg/dl	-2.3 (-10.5 to 6)	0.5920	0		
	Serum LDL cholesterol, mg/dl	9.8 (-8.3 to 27.9)	0.2904	1	1.6 (-16.3 to 19.6)	0.8582
	eGFR, ml/min per 1.73 m^2	-0.4 (-27.9 to 27)	0.9755	0		
	Urine ACR, mg/g[a]	0.2 (-0.8 to 1.2)	0.6997	1	0.2 (-0.8 to 1.2)	0.6360
	hsCRP, mg/L[a]	***-0.7 (-1.0 to -0.3)***	***0.0001***	0		
	Morisky score[b]	-0.5 (-1.6 to 0.7)	0.4191	1	-0.6 (-1.6 to 0.5)	0.2810
	KDQDL measures					
	SF-12 physical score	2.7 (-3.5 to 8.9)	0.3950	0		
	SF-12 mental score	0.9 (-3.1 to 4.9)	0.6608	5	1.4 (-2.9 to 5.7)	0.5196
Logistic Regression	**Primary outcome measures**	**OR (95% C.I.)**	**p-value**		**OR (95% C.I.)**	**p-value**
	Patient activation level ≥ 3	***13.2 (1.1 to 166.1)***	***0.0454***	0		
	Secondary outcome measures					
	KDQDL measures					
	Symptom/ problem list	1.1 (0.3 to 4.1)	0.8547	0		
	Effects of kidney disease	*3.7 (0.8 to 17.4)*	*0.0951*	0		
	Burden of kidney disease	1.8 (0.5 to 6.3)	0.3749	0		

Table 3.7 Estimated treatment effect of intervention compared control for Model-B with and without influential points

Cross Validation

To assess if the results of a statistical analysis could be generalizable to an independent dataset, cross-validation could be used. This technique estimates how accurately a predictive model is performing in practice. Through cross-validation, one firstly partitions the sample data, repeatedly, into complementary subsets called the training sets and testing sets respectively. Secondly, one performs the analysis on the training set, and thirdly validates the analysis on the other subset (called the validation set or testing set).[40]

For the purpose of this study, we have conducted cross validation for the models that only showed statistically significant treatment effect in Table 3.6 (while adjusting for outcome baseline measurements only) since they didn't substantially differ than the results from Table 3.7 (while adjusting for outcome baseline measurements plus imbalanced covariates at baseline).

The matric we use to evaluate cross validation for the linear regression are Root Mean Square Error (RMSE), the mean absolute error (MAE), and R^2 in the test set.[45] Here, the RMSE is the standard deviation of the differences Y-$\hat{Y}$, while the MAE is the average of their absolute values. For logistic regression, we used Mean Percentage of Correct Classifications (MPCC) and its corresponding Standard Error (SE) as a measure of reliability and validity for the cross validation.[46]

		Treatment Effect $\hat{\gamma}$ (95% C.I.)	**p-value**	**RMSE**	**MAE**	**R^2**
Linear Regression	**Primary outcome measures**					
	Patient activation total score	*16 (8.8 to 23.1)*	*<.0001*	*14.2*	*9.7*	*0.39*
	Secondary outcome measures					
	Body mass index, kg/m^2	*-1.2 (-2.2 to -0.2)*	*0.0212*	*2.0*	*1.5*	*0.09*
	hsCRP, mg/L[a]	*-0.7 (-1 to -0.3)*	*<.0001*	*0.61*	*0.47*	*0.38*
Logistic Regression	**Primary outcome measures**	**OR (95% C.I.)**	**P-value**	**MPCC (SE) For training sets**		**MPCC (SE) For testing sets**
	Patient activation level ≥ 3	*9.7 (1.7 to 54.3)*	*0.0097*	*78.7 (0.08)*		*77.5 (0.34)*

Table 3.8 Cross validation results

For patient activation score, the R^2 reflects that 39% of the variation in patient activation score can be attributed to the *Treatment* while adjusting for the baseline measurement of PAM score. For BMI, 9% of the variation in BMI can be attributed to the *Treatment* while adjusting for the baseline measurement of BMI. For high sensitivity CRP levels, 38% of the variation in hsCRP levels can be attributed to the *Treatment* while adjusting for the baseline measurement of hsCRP. For patient activation level, MPCC for the testing set reflects that 77.5% of the predictions made using the logistic regression model for PAM level was correct.

Chapter 4

Propensity Score

Though the treatment assignment was allotted through randomization, there remained three variables at baseline that were statistically significantly different between groups—cholesterol, LDL and hsCRP. This imbalance may have introduced treatment selection bias to the estimated parameters in the model. As a remedy, we used inverse probability of treatment weighting derived from propensity scores as a method to reduce the potentially induced bias. Propensity score has been defined as the probability of being assigned to the intervention group given a set of observed independent variables.[47]

Let Z be an indicator of the binary *Treatment* with 1 for home-based renal care and 0 for standard clinical care—hence primary exposure, $\mathbf{x}_i$ be a row vector of observed values of imbalanced variables, and $Pr(Z = 1| X = \mathbf{x}_i)$ be the propensity towards assignment to *Treatment* 1 given the observed i^{th} values of imbalanced variables (i.e. $\mathbf{x}_i$).

$$[5]\ PS_i = Pr(Z = 1| X = \mathbf{x}_i)$$

Estimating Propensity Scores

To obtain propensity scores (PSs), we used a multiple logistic regression model including the three imbalanced variables and the baseline measurement as independent variables (IVs), and *Treatment* Z as dependent variable (DV). Note that the primary exposure— *Treatment*—became the dependent variable for this intermediate calculation. We note that the choice of only four predictors to calculate PSs is due to the small sample size especially for the events of interest in the dichotomized primary outcome. According to Vittinghoff and McCulloch (2006), we need at least 6-7 observation in logistic regression per event to avoid overfitting and bias in parameters' estimates. Mathematically, the logistic regression model for estimating the propensity score was:

$$[6]\ \ln\left(\frac{PS_i}{1-PS_i}\right) = \boldsymbol{\beta}\mathbf{x}_i+\varepsilon$$

$$\Leftrightarrow PS_i = \frac{\exp(\boldsymbol{\beta}\mathbf{x}_i+\varepsilon)}{1+\exp(\boldsymbol{\beta}\mathbf{x}_i+\varepsilon)},$$

where $\boldsymbol{\beta}$ was a vector of parameters for imbalanced variables and PS_i as described in [2].

Properties of Propensity Scores

An important property of propensity score is that if *Treatment* is independent given the realization $\mathbf{x}_i$, then it is also independent given the propensity score. That is:

$$[7.1]\ (y_1, y_0) \perp\!\!\!\perp Z \mid \mathbf{x}_i$$

$$[7.2] \Rightarrow (y_1, y_0) \perp\!\!\!\perp Z \mid PS_i,$$

where y_0 is the response that would have resulted if the participant were assigned to *Treatment* Z=1 and y_1 is the response that would have resulted if the participant were assigned to *Treatment* Z=0. Equation [7.1] is referred to as the conditional independence assumption.

Another property of propensity score is that if every participant has a positive probability of receiving each *Treatment* given the realization $\mathbf{x}_i$, then this is also true given the propensity score. That is:

$$[8.1]\ 0 < \Pr(Z = 1 \mid X = \mathbf{x}_i) < 1$$

$$[8.2] \Rightarrow 0 < \Pr(Z = 1 \mid PS_i) < 1$$

Equation [8.1] is referred to as the common support assumption. Thus, when these two properties are true, it is said that *Treatment* is strongly ignorable given the realization $\mathbf{x}_i$.

Inverse Probability of Treatment Weighting (IPTW)

Though the concept of propensity scores was introduced in 1983 by Rosenbaum, it was only in 1987 when he introduced IPTW.[48] The idea behind IPTW is that one could minimize treatment selection

bias by weighting the observations in both groups of the intervention while utilizing the estimated propensity scores.[49] Going from PSs to weights, denoted by W, is done as follows.

$$[9]\ W_i = \frac{Z_i}{PS_i} + \frac{(1-Z_i)}{1-PS_i}$$

$$\Leftrightarrow W_i = \begin{cases} \frac{1}{PS_i} & \text{when } Z = 1 \\ \frac{1}{1-PS_i} & \text{when } Z = 0 \end{cases}$$

Assumptions for conducting IPTW:

1. *Common Support Assumption*: This assumption entails having an overlap in the support of the distribution of PSs between the two groups of the intervention (i.e. between Z=1 and Z=0).

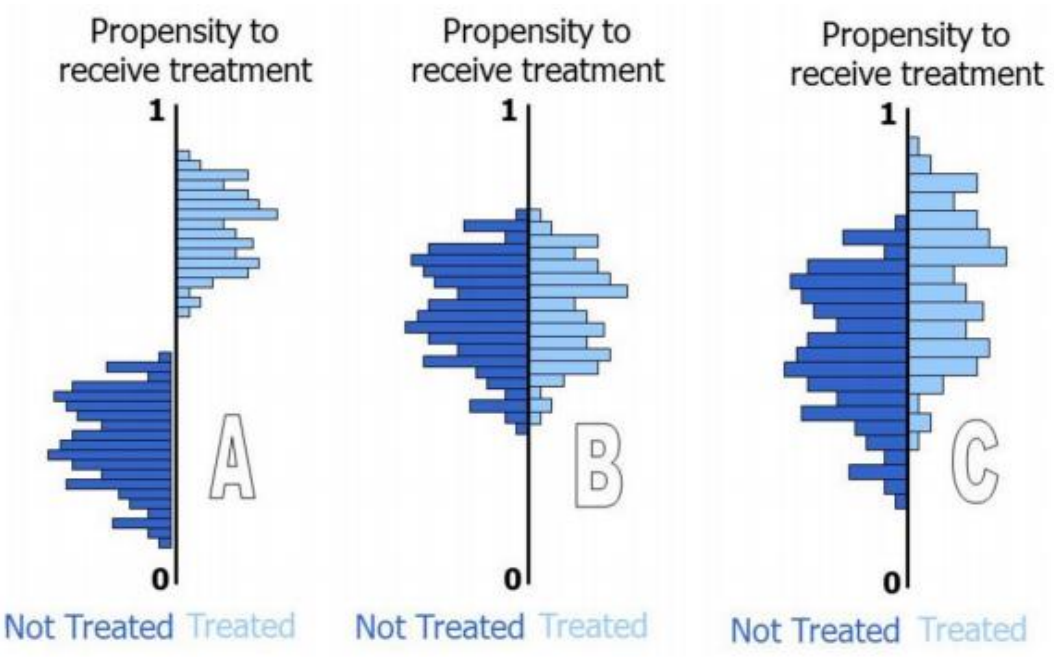


Figure 3.10 Three scenarios for the common support assumption. A: no overlap indication significant violation of the common support assumption, B: complete overlap indicating compliance with the common support assumption, and C: partial overlap indicating minor violation of the common support assumption that could be remedied via some techniques suggested below.[50]

A situation as in A of Figure 3.10 suggests that propensity scores is less likely to be helpful in reducing treatment selection bias; the imbalance is so drastic to the extent that such approach cannot bring any

benefits. However, in a situation as in C of Figure 3.10 one could resolve the violation of the common support assumption by either (a) truncate the non-overlapping support segments or (b) utilize the Stabilized Inverse Probability of Treatment Weighting (SIPTW) as shown below.

2. *Balance Assumption:*

This assumption entails having similar distribution for the IPTW between the two groups of the intervention (i.e. between Z = 1 and Z = 0). Figure 3.11 demonstrates a real example showing a graphical examination for this assumption.

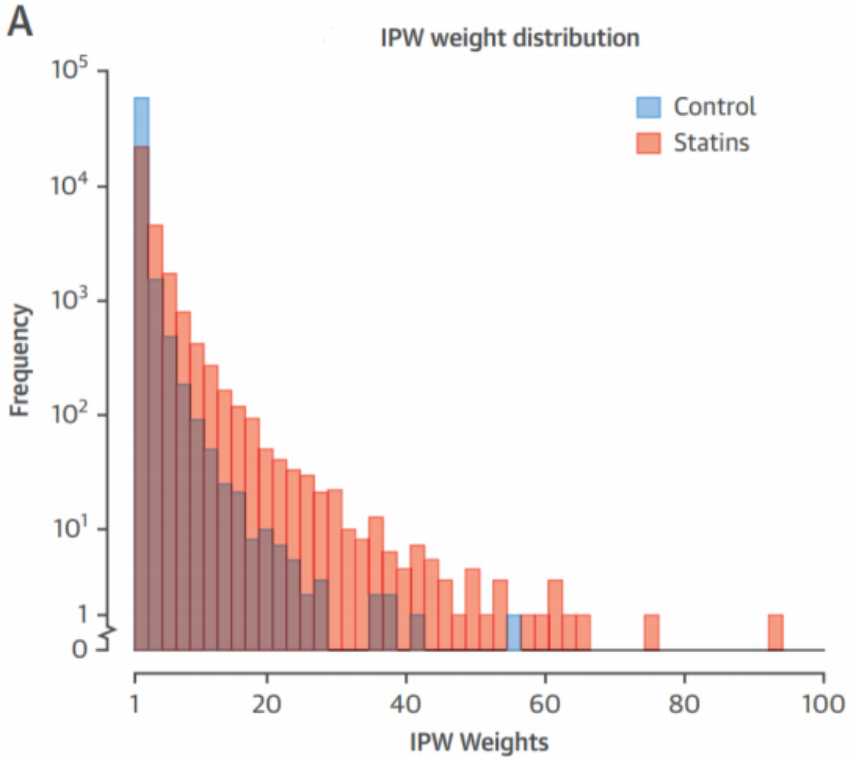


Figure 3.11 An example for the distribution of IPTW from a real example.[51]

Stabilized Inverse Probability of Treatment Weighting (SIPTW)

Using the unstabilized IPTW has been previously reported to underestimate the variance of the treatment effect producing narrow confidence intervals that leads to inflation of the probability of a type-I error.[52] The SIPTW can be obtained by multiplying the IPTW by the marginal probability of *Treatment* to

Z=1 without considering independent variables. Let θ be the marginal probability of *Treatment* to 1. The stabilized SIPTW was calculated as follows:

$$[10]\ \mathrm{SIPTW_i} = \begin{cases} \frac{\theta}{\mathrm{PS_i}} & \text{when } z = 1 \\ \frac{1-\theta}{1-\mathrm{PS_i}} & \text{when } z = 0 \end{cases}$$

Application of Propensity Scores: An Example from Nephrology

Applying these principles to the HBRC data. Firstly, the propensity score was estimated according to equation [6] using PROC LOGISTIC in SAS where we used *Treatment* as the primary outcome and cholesterol, LDL, hsCRP, and baseline PAM score as IVs. Secondly, stabilized weights were created according to equation [10]. Thirdly, weighted simple logistic regression model with PAM as the primary outcome, using the SIPTW weights, with *Treatment* as the primary exposure was employed to produce parameter estimates with minimal treatment selection bias.

1. *Testing if SIPTW could balance baseline characteristics between intervention groups:*

To assess if the generated weights were able to fix the imbalance at baseline, we have conducted multiple unweighted logistic regression model in which *Treatment* (Z) as the primary outcome and cholesterol, LDL, hsCRP, and baseline PAM score as IVs. The results of this model are presented in Table 3.1 (and Figure 3.1) and indeed showed the existing imbalance between the two groups with respect to these IVs via unweighted odds ratio and their corresponding 95% confidence intervals. In particular, a 10 units increase in *Cholesterol* level was associated with 26.77% (1.024^10=1.2677) higher odds of being in the intervention group (OR=1.024, 95% C.I. 1.004-1.064; P=0.0102) which is an indication of imbalance. Furthermore, a one units increase in hsCRP level was associated with 13.8% higher odds of being in the intervention group (OR=1.138, 95% C.I. 1.038-1.315; P=0.0024) which is another indication of imbalance.

To demonstrate that SIPTWs were able to balance baseline characteristics between *Treatment* groups, we have employed multiple weighted logistic regression model, using the SIPTWs weights, in which *Treatment* (Z) as the primary outcome and cholesterol, LDL, hsCRP, and baseline PAM score as IVs. The results of this model are presented in Table 3.9 (and Figure 3.12) and indeed showed the existing imbalance between the two groups with respect to these IVs was diminished. In particular, none of PAM score at baseline, Cholesterol, LDL, or hsCRP had a significant association with higher odds of being in the intervention group (p=0.8201, p=0.3195, p=0.7314, and p=0.3098 respectively) which is an indication of balance.

	Estimate	**95% Confidence Intervals**	**p-value**
Pam Score at Baseline	0.964	0.913 to 1.009	0.1163
Cholesterol	**1.024**	**1.004 to 1.064**	**0.0102**
LDL	0.992	0.946 to 1.021	0.6582
hsCRP	**1.138**	**1.038 to 1.315**	**0.0024**

Table 3.9 Unweighted OR Estimates for being in the intervention group with 95% C.I.

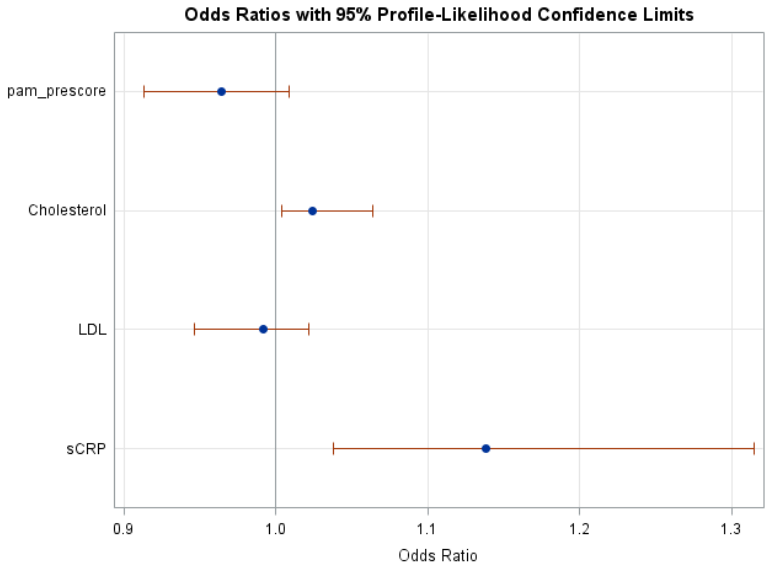


Figure 3.12 Unweighted ORs for being in the intervention group with 95% C.I.

	Estimate	95% Confidence Intervals	p-value
PAM Score at Baseline	0.995	0.952 to 1.039	0.8201
Cholesterol	1.009	0.992 to 1.039	0.3195
LDL	0.995	0.957 to 1.022	0.7314
hsCRP	1.041	0.964 to 1.149	0.3098

Table 3.10 Weighted OR Estimates of being in the Treatment group with 95% C.I.

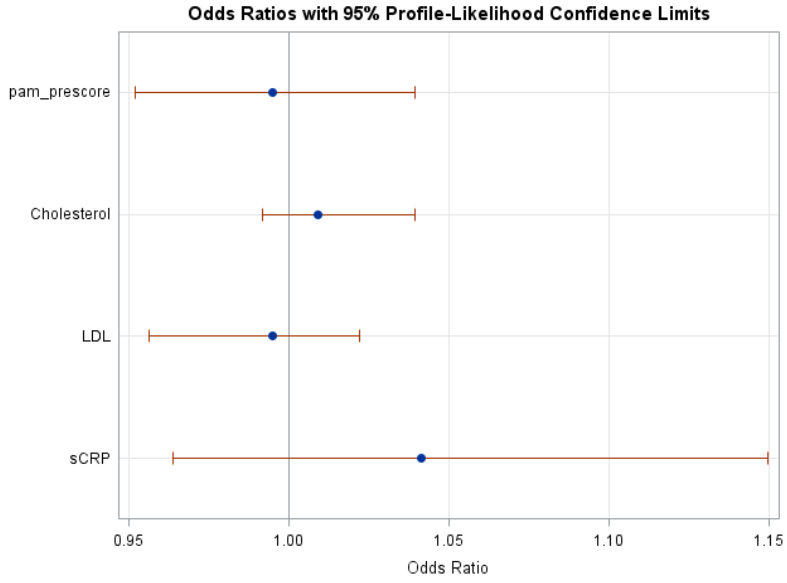


Figure 3.13 Weighted ORs of being in the Treatment group with 95% C.I.

2. Checking Assumptions

Before using the estimated propensity score, we checked model assumptions. First, the conditional independence assumption was met because the outcome of one participant did not affect that of another participant. This was assumed by random selection of the sample participants. Then, we checked the common support or the overlap condition assumption. Each participant must have a non-zero probability of receiving either of the *Treatment* options in a randomized controlled trial.

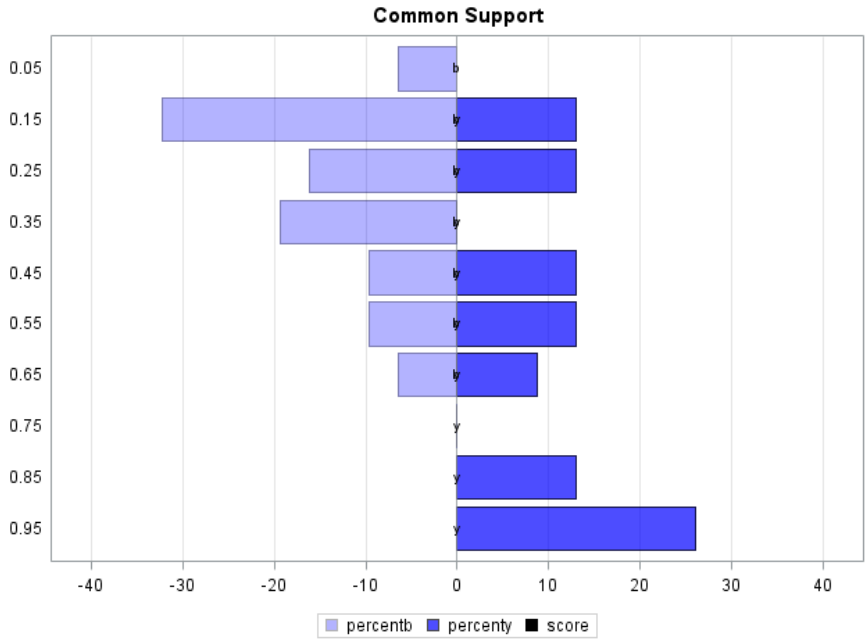


Figure 3.14 Common Support Assumption

As Figure 3.14 shows, there was a partial overlap in the distribution of PSs for the control and intervention groups. No overlap would suggest that there are too many pre-existing differences between groups for causal inference. Because of the partial overlap in the distribution of PSs, we proceeded to calculate the inverse probability of treatment weight using PSs.

Calculating the Stabilized Inverse Probability of Treatment Weights

To calculate the stabilized inverse probability of treatment weighting, we first calculated the marginal probability of *Treatment* $Z = 1$. Using the HBRC data, we found that θ was 0.4259. In other words, the probability of being assigned to the home-based renal care group was 42.9%. In turn, the probability of being assigned to the standard clinical care group was 57.1%. Then, we calculated SWs using equation [10]. Figure 3.15 showed that the distribution of the SIPTWs between groups were similar, which is an indication of balance.

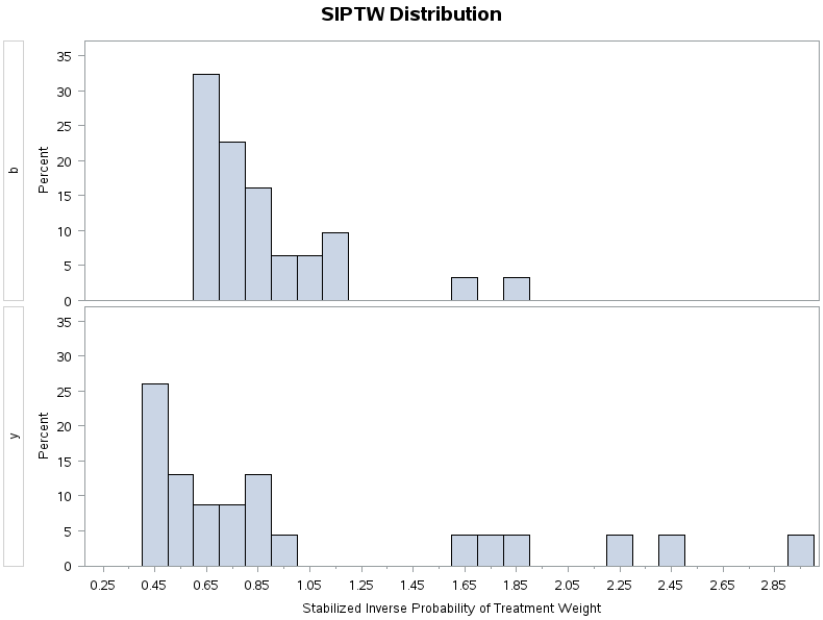


Figure 3.15 Distribution of stabilized inverse probability of treatment weight

Before using the SIPTWs to estimate the parameters again, we checked the odds ratio estimates to ensure that we have corrected the imbalance of the treatment assignment. Table 3.10 showed the estimated odds ratio and the corresponding 95% confidence intervals. The testing hypotheses were H_0: odds ratio = 1 vs. H_1: odds ratio $\neq$ 1 at 95% confidence. Each of the 95% confidence intervals for the baseline PAM score, cholesterol, LDL and hsCRP included 1. We failed to reject the null hypotheses and concluded that the treatment assignment was independent from the outcomes of interest.

Applying the SIPTWs to reduce treatment selection bias

Using SIPTWs, we estimated the parameters of the original regression models again to correct the treatment selection bias. Table 3.11 showed a side-by-side comparison of estimated parameters given by regression models with and without the SIPTWs.

		Without SIPTW		With SIPTW	
	Characteristic	**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**	**Treatment Effect $\hat{\gamma}$ (95% C.I.)**	**p-value**
Linear Regression	**Primary outcome measures**				
	Patient activation total score	***16 (8.8 to 23.1)***	***<.0001***	***15.6 (8.3 to 22.9)***	**<.0001**
	Secondary outcome measures				
	Body mass index, kg/m^2	***-1.2 (-2.2 to -0.2)***	***0.0212***	-0.7 (-1.7 to 0.3)	0.1446
	BP, mm HG				
	Systolic	-3.7 (-11.2 to 3.7)	0.3237	-2.5 (-9.7 to 4.7)	0.4980
	Diastolic	-1.9 (-7.4 to 3.6)	0.4954	-1.1 (-6.2 to 3.9)	0.6635
	HbA1c, %	-0.8 (-1.9 to 0.4)	0.1896	-0.2 (-1.2 to 0.9)	0.7425
	Glucose, mg/dl	-3 (-47.5 to 41.6)	0.8957	11.7 (-34.2 to 57.6)	0.6177
	Serum total protein, mg/dl	-0.1 (-0.3 to 0.1)	0.3495	-0.1 (-0.4 to 0.1)	0.2398
	Serum cholesterol, mg/dl	4.8 (-15.9 to 25.5)	0.6513	4.6 (-16 to 25.1)	0.6633
	Serum triglycerides, mg/dl[a]	-0.1 (-0.3 to 0.2)	0.5573	0.04 (-0.2 to 0.3)	0.7785
	Serum HDL cholesterol, mg/dl	-0.1 (-7.9 to 7.7)	0.9706	-1.7 (-9.7 to 6.3)	0.6782
	Serum LDL cholesterol, mg/dl	4.9 (-11.2 to 21)	0.5529	3.1 (-12.1 to 18.4)	0.6886
	eGFR, ml/min per 1.73 m^2	9.1 (-12.6 to 30.9)	0.4110	7.8 (-13.4 to 29.0)	0.4724
	Urine ACR, mg/g[a]	-0.5 (-1.2 to 0.3)	0.2631	-0.08 (-1.0 to 0.8)	0.8600
	hsCRP, mg/L[a]	***-0.7 (-1 to -0.3)***	***<.0001***	***-0.6 (-0.9 to -0.3)***	***0.0001***
	Morisky score[b]	-0.2 (-1.1 to 0.8)	0.6903	-0.6 (-1.6 to 0.3)	0.2107
	KDQDL measures				
	SF-12 physical score	2 (-3.2 to 7.1)	0.4531	1.9 (-3.3 to 7.0)	0.4765
	SF-12 mental score	1.2 (-3.3 to 5.8)	0.5975	2.6 (-1.1 to 6.2)	0.1635
Logistic Regression	**Primary outcome measures**	**OR (95% C.I.)**	**P-value**	**OR (95% C.I.)**	**P-value**
	Patient activation level ≥ 3	***9.7 (1.7 to 54.3)***	***0.0097***	***8.4 (1.9 to 37)***	***0.0049***
	Secondary outcome measures				
	KDQDL measures				
	Symptom/ problem list	0.8 (0.2 to 2.9)	0.7209	0.9 (0.2 to 3.2)	0.8572
	Effects of kidney disease	2.5 (0.7 to 8.7)	0.1519	*3.0 (0.9 to 10.3)*	*0.0810*
	Burden of kidney disease	*2.7 (0.9 to 8.3)*	*0.0862*	2.7 (0.8 to 8.9)	0.1105

Table 3.11 Estimated treatment effect of intervention compared control with and without SW

Chapter 5

Discussion

In conclusion, after using SIPTW to adjust for the imbalanced variable at baseline, the estimated average change in the difference in PAM scores is 15.6 points higher in the treatment group compared to the control group while holding the baseline PAM scores constant. When PAM score is categorized into PAM level, we observe that participants in the treatment group are 8.4 times more likely to be activated at 12 months compared to the control group while holding the baseline PAM scores constant. SIPTWs could be used to reduce treatment selection bias by 2.5% for the continuous outcome variable and 13.4% for the binary outcome variable. The impact of SIPTWs in reducing treatment selection bias could be more pronounced in observational studies. The selection of participants with diabetes mellitus in the beginning of the study may have contributed to the observed bias in the parameter estimation.

As previously mentioned, the HBRC data were collected through a randomized controlled trial. Because of the randomization, the implementation of propensity score may not have been substantially impactful given that among 21 baseline covariates 18 of them were balanced to start with. The subject matter expert was interested in analyzing the data only for the diabetic subgroup from the collected data. We did not perform another round of randomization for the selection of this subset participants with diabetes. Though it was likely that a subset selected from the randomized groups was still random, it was not guaranteed. This was supported by the three variables that showed statistically significant difference at baseline.

Additionally, there were 56 total participants of which 24 participants were in the intervention group. This was a small sample size for implementation of propensity score. There might be insufficient data to reach meaningful results. The small sample size also limited the number of covariates used to generate the propensity scores. Vittinghoff and McCulloch (2007) found that it is acceptable to relax the

rule of thumb of 10 or more Event Per Variable (EPV) to 5-9 EPV in using a logistic regression model.[53] They defined the acceptable threshold as confidence interval coverage of greater than 93 percent, type 1 error rate of lower than 7 percent or relative bias less than 15 percent. By this finding, we were limited to having a parsimonious model with 2 to 4 covariates in the logistic regression model that was used to generate the propensity scores. To avoid overfitting, we only adjusted for 4 covariates to stay within this acceptable range of EPVs.

Another limitation brought on by the small sample size in regard to the propensity score is that it was not appropriate to trim the dataset. The restriction of treatment comparisons to subjects with a common range of covariates can improve the validity of estimated treatment effects.[54] With large sample size, propensity score trimming can increase the validity of the treatment effect estimates.[55] However, any analysis done using the trimming method will not be causal in the sense that they do not apply to any clearly defined population because of the range restriction.[55] Again, this technique did not apply well to this data because of its small sample size.

Thirdly, another limitation of the analysis of this study is the potentially inflated *Treatment* estimates. This is because we estimated the *Treatment* effect on the difference between baseline to 12 months measures while adjusting for the baseline measurements. Glymour (2005) reported that even though the baseline adjustments improve efficiency and eliminates confounding, adjusting for baseline measures could introduce bias to the estimated parameters in situations where the primary exposure predicts baseline level of the outcome.[56] In particular, it gives two common situations where this could occur—(1) when the measurement reliability is imperfect or the latent variable is instable and (2) when the change has already occurred prior to the baseline measurement, the rate of change experienced in the past predicts that of the future, and exposure is unaffected by baseline function. When either of these situation holds true, the baseline adjusted model induces a spurious correlation between the exposure and the change score because the exposure is likely to be a predictor on the outcome even under the null assumption of no causal effect.

Finally, in using the ROC analysis to arrive at the optimal cut-off point, we opted to use the Euclidean distance *D* rather than the Youden Index *J* in situations where the two values were not in agreement. While both methods are very practical in dichotomizing a continuous variable, there are advantages to using the Youden Index *J* over the Euclidean distance *D*. According to Perkins and Schisterman (2005), the Youden Index *J* is more robust against measurement error compared to the Euclidean distance *D* given by an approximate confidence interval generated using the delta method.[57] This may be of concern in our study because the data collection was not done by a laboratory machine. Also, *J* has easier clinical interpretation because it does not involve a quadratic term in its calculation. However, another study by Unal (2017) shows that the Euclidean distance *D* consistently reduced the relative bias and MSE compared to the Youden Index *J*.[35] For this reason, we used the Euclidean distance D. It would be interesting to compare the treatment estimates given by the two different methods.

For a future study, it would be beneficial to conduct the randomized controlled trial on the diabetic patients rather than the CKD patients with diabetes to have a better interpretation of the treatment effect on the participants. It would provide better insight into the direct effect that the home-based renal care has on diabetic patients rather than the effect that it has on CKD patients with diabetes. Also, a concern that a local Zuni chief had was that people with compromised health did not get the treatment. This community would benefit from conducting a delayed randomized trial rather than the traditional randomized controlled trial because all participants in the study would be able to get an intervention at some point in the trial.

References

1. Dieleman, J. L. *et al.* US Spending on Personal Health Care and Public Health, 1996-2013. *JAMA* **316**, 2627–2646 (2016).
2. U.S. Diabetes Surveillance System. Available at: https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html. (Accessed: 19th March 2019)
3. Stevens, P. E., Levin, A. & Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann. Intern. Med.* **158**, 825–830 (2013).
4. Kramer, H. J., Nguyen, Q. D., Curhan, G. & Hsu, C.-Y. Renal insufficiency in the absence of albuminuria and retinopathy among adults with type 2 diabetes mellitus. *JAMA* **289**, 3273–3277 (2003).
5. Norris, K. & Nissenson, A. R. Race, gender, and socioeconomic disparities in CKD in the United States. *J. Am. Soc. Nephrol.* **19**, 1261–1270 (2008).
6. Crews, D. C., Bello, A. K., Saadi, G. & World Kidney Day Steering Committee. Burden, Access, and Disparities in Kidney Disease. *Am. J. Hypertens.* **32**, 433–439 (2019).
7. Shah, V. O. *et al.* Identifying barriers to healthcare to reduce health disparity in Zuni Indians using focus group conducted by community health workers. *Clin Transl Sci* **7**, 6–11 (2014).
8. Whitbeck, L. B., Adams, G. W., Hoyt, D. R. & Chen, X. Conceptualizing and measuring historical trauma among American Indian people. *Am J Community Psychol* **33**, 119–130 (2004).
9. MacCluer, J. W. *et al.* Heritability of measures of kidney disease among Zuni Indians: the Zuni Kidney Project. *Am. J. Kidney Dis.* **56**, 289–302 (2010).

10. Williams, G. C. *et al.* Promoting glycemic control through diabetes self-management: evaluating a patient activation intervention. *Patient Educ Couns* **56**, 28–34 (2005).
11. Hibbard, J. H., Mahoney, E. R., Stockard, J. & Tusler, M. Development and testing of a short form of the patient activation measure. *Health Serv Res* **40**, 1918–1930 (2005).
12. Mei, Z. *et al.* Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am. J. Clin. Nutr.* **75**, 978–985 (2002).
13. Hsu, C., McCulloch, C. E., Iribarren, C., Darbinian, J. & Go, A. S. Body mass index and risk for end-stage renal disease. *Ann. Intern. Med.* **144**, 21–28 (2006).
14. Kramer, H. *et al.* Obesity and prevalent and incident CKD: the Hypertension Detection and Follow-Up Program. *Am. J. Kidney Dis.* **46**, 587–594 (2005).
15. Catapano, F. *et al.* Antiproteinuric response to dual blockade of the renin-angiotensin system in primary glomerulonephritis: meta-analysis and metaregression. *Am. J. Kidney Dis.* **52**, 475–485 (2008).
16. Nathan, D. M., Singer, D. E., Hurxthal, K. & Goodson, J. D. The clinical information value of the glycosylated hemoglobin assay. *N. Engl. J. Med.* **310**, 341–346 (1984).
17. Diabetes Control and Complications Trial Research Group *et al.* The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986 (1993).
18. American Diabetes Association. (2) Classification and diagnosis of diabetes. *Diabetes Care* **38 Suppl**, S8–S16 (2015).
19. Abboud, H. & Henrich, W. L. Clinical practice. Stage IV chronic kidney disease. *N. Engl. J. Med.* **362**, 56–65 (2010).

20. Bammens, B., Verbeke, K., Vanrenterghem, Y. & Evenepoel, P. Evidence for impaired assimilation of protein in chronic renal failure. *Kidney Int.* **64**, 2196–2203 (2003).
21. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* **106**, 3143–3421 (2002).
22. Cox, R. A. & García-Palmieri, M. R. Cholesterol, Triglycerides, and Associated Lipoproteins. in *Clinical Methods: The History, Physical, and Laboratory Examinations* (eds. Walker, H. K., Hall, W. D. & Hurst, J. W.) (Butterworths, 1990).
23. Grundy, S. M. *et al.* Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation* **110**, 227–239 (2004).
24. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
25. Hostetter, T. H., Olson, J. L., Rennke, H. G., Venkatachalam, M. A. & Brenner, B. M. Hyperfiltration in remnant nephrons: a potentially adverse response to renal ablation. *Am. J. Physiol.* **241**, F85-93 (1981).
26. Levey, A. S. & Coresh, J. Chronic kidney disease. *Lancet* **379**, 165–180 (2012).
27. Ginsberg, J. M., Chang, B. S., Matarese, R. A. & Garella, S. Use of single voided urine samples to estimate quantitative proteinuria. *N. Engl. J. Med.* **309**, 1543–1546 (1983).
28. Levey, A. S. *et al.* The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int.* **80**, 17–28 (2011).

29. Kushner, I. The phenomenon of the acute phase response. *Ann. N. Y. Acad. Sci.* **389**, 39–48 (1982).
30. Scirica, B. M. & Morrow, D. A. Is C-reactive protein an innocent bystander or proatherogenic culprit? The verdict is still out. *Circulation* **113**, 2128–2134; discussion 2151 (2006).
31. Wilson, P. W. F. *et al.* C-reactive protein and reclassification of cardiovascular risk in the Framingham Heart Study. *Circ Cardiovasc Qual Outcomes* **1**, 92–97 (2008).
32. Zimmermann, J., Herrlinger, S., Pruy, A., Metzger, T. & Wanner, C. Inflammation enhances cardiovascular risk and mortality in hemodialysis patients. *Kidney Int.* **55**, 648–658 (1999).
33. Ricardo, A. C. *et al.* Validation of the Kidney Disease Quality of Life Short Form 36 (KDQOL-36) US Spanish and English versions in a cohort of Hispanics with chronic kidney disease. *Ethn Dis* **23**, 202–209 (2013).
34. Morisky, D. E., Ang, A., Krousel-Wood, M. & Ward, H. J. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens (Greenwich)* **10**, 348–354 (2008).
35. Unal, I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Computational and Mathematical Methods in Medicine* (2017). doi:10.1155/2017/3762651
36. Liang, K.-Y. & Zeger, S. L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika* **73**, 13–22 (1986).
37. Ziegler, A. *Generalized Estimating Equations*. (Springer Science & Business Media, 2011).
38. 12.1 - Introduction to Generalized Estimating Equations | STAT 504. Available at: https://newonlinecourses.science.psu.edu/stat504/node/180/. (Accessed: 16th April 2019)

39. Gong, G. & Samaniego, F. J. Pseudo Maximum Likelihood Estimation: Theory and Applications. *The Annals of Statistics* **9**, 861–869 (1981).
40. Qeadan, F. Longitudinal Data Analysis by Example. (2016).
41. Christensen, R. *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. (CRC Press, 1996).
42. Indrayan, A. & Malhotra, R. K. *Medical Biostatistics*. (CRC Press, 2017).
43. Biswas, A., Datta, S., Fine, J. P. & Segal, M. R. *Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics*. (John Wiley & Sons, 2007).
44. *Interpreting Coefficients in Regression with Log-Transformed Variables*. (Cornell University, 2012).
45. Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* **7**, 1247–1250 (2014).
46. Saeed, A. I. *et al.* A novel cytokine profile associated with cancer metastasis to mediastinal and hilar lymph nodes identified using fine needle aspiration biopsy - A pilot study. *Cytokine* **89**, 98–104 (2017).
47. Rosenbaum, P. R. & Rubin, D. B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**, 41–55 (1983).
48. Rosenbaum, P. R. Model-Based Direct Adjustment. *Journal of the American Statistical Association* **82**, 387–394 (1987).
49. Qeadan, F. The How and Why of Propensity Scores in Comparative Analyses of Observational Data. (2017).

50. Newman, T. B. & Kohn, M. A. *Evidence-Based Diagnosis*. (Cambridge University Press, 2009).

51. Elze, M. C. *et al.* Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *J. Am. Coll. Cardiol.* **69**, 345–357 (2017).

52. Xu, S. *et al.* Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health* **13**, 273–277 (2010).

53. Vittinghoff, E. & McCulloch, C. E. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am. J. Epidemiol.* **165**, 710–718 (2007).

54. Miettinen, O. S. Stratification by a multivariate confounder score. *Am. J. Epidemiol.* **104**, 609–620 (1976).

55. Stürmer, T., Rothman, K. J., Avorn, J. & Glynn, R. J. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. *Am. J. Epidemiol.* **172**, 843–854 (2010).

56. Glymour, M. M., Weuve, J., Berkman, L. F., Kawachi, I. & Robins, J. M. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am. J. Epidemiol.* **162**, 267–278 (2005).

57. Schisterman, E. F., Perkins, N. J., Liu, A. & Bondell, H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* **16**, 73–81 (2005).