EVALUATING THE EFFECT OF KIDNEY VOLUME ON DECLINE IN RENAL FUNCTION USING THE GENERALIZED PROPENSITY SCORE

by

Yaming Li

BA, Hebei Medical University, China, 1998

MS, Capital Medical University, China, 2003

Submitted to the Graduate Faculty of

the Department of Biostatistics

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Yaming Li

It was defended on

June 12th, 2017

and approved by

Committee Member: Gary M, Marsh, PhD
Professor of Biostatistics, Epidemiology and Clinical & Translational Science
Graduate School of Public Health and School of Medicine
University of Pittsburgh

Committee Member: Ada O. Youk, PhD
Associate Professor of Biostatistics, Epidemiology and Clinical &Translational Science
Graduate School of Public Health and School of Medicine
University of Pittsburgh

Thesis Director: Douglas P. Landsittel, PhD
Professor of Medicine, Biostatistics and Clinical and Translational Science
School of Medicine and Graduate School of Public Health
University of Pittsburgh

Douglas Landsittel, PhD

EVALUATING THE EFFECT OF KIDNEY VOLUME ON DECLINE IN RENAL FUNCTION USING THE GENERALIZED PROPENSITY SCORE

Yaming Li, M.S.

University of Pittsburgh, 2017

ABSTRACT

Autosomal-dominant polycystic kidney disease (ADPKD) is characterized by gradual renal enlargement and cyst growth prior to the loss of renal function. The Consortium for Radiologic Imaging Studies in Polycystic Kidney Disease (CRISP) is a longitudinal observational study ADPKD individuals using high-resolution magnetic resonance (MR) imaging to determine if a change in renal and cyst volumes can be detected over a short period of time, and if they correlate with a decline in renal function early in the disease. The aim of this study was to determine if height-adjusted total kidney volume (htTKV) had a causal effect on renal decline in the CRISP cohort by using a method for causal inference, namely the generalized propensity score (GPS) method, which is a generalization of the more common propensity score methods (applicable to binary treatments or exposures) for continuous data. Results provide further evidence that baseline htTKV may have a causal effect on subsequent renal function (measured at least a decade later). The study did however have limitations, as we could only consider limited factors available at birth to construct the GPS (and thus preserve temporal associations).

This study has a high degree of public health significance given the high incidence and cost of chronic kidney disease (CKD) and end stage renal disease (ESRD). CKD is identified as

a major public health concern requiring intervention, as nearly 20 million people are estimated to have CKD. ESRD also introduces a significant burden on patient, health care, and societal costs. Finding biomarkers that identify cases earlier are critical to reducing the disease burden.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to express my sincere and genuine gratitude to Dr. Douglas Landsittel, my thesis advisor, who introduced the topic of my thesis and has always been encouraging and illuminating me through my whole study and research. Without his insights and guidance, finishing this thesis would be impossible. My gratitude would also go to my academic advisor Dr. Gary Marsh, who gave me constructive suggestions and advise of my academic study and theses design and analysis. I also appreciate Dr. Ada O. Youk for not only sitting on my thesis committee but most importantly for sharing her valuable insights and inspiring comments on this thesis.

I also show my great appreciation to Renee Valenti, the Recruitment and Academic Affairs Administrator of biostatistics. Whenever I have questions and problem, she is always ready and willing to help me.

Last but not least, I would like to thank my husband for supporting my back both financially and mentally while I worked on my degree. He has shared both my joy and frustration during the past two years. If any glory, it should be dedicated to him as well.

1.0 INTRODUCTION

1.1 AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

Polycystic kidney disease (PKD) is an inherited disorder characterized by cystic expansion of the kidneys producing progressive kidney enlargement and renal insufficiency. Autosomal dominant polycystic kidney disease (ADPKD) is the most common form of polycystic kidney disease, occurring in 1 in 800 live births and the third most common single cause of end-stage renal disease (ESRD) in the United States.[1, 2] There are two types of ADPKD: type I is caused by mutations in the *PKD1* gene and accounts for 85 to 90 percent of cases[3] and type II is caused by mutations in the *PKD2* gene and accounts for 10 to 15 percent of cases.[4] Type II disease has a later onset of symptoms and a slower rate of progression to renal failure; therefore patients have a longer life expectancy (69.1 years) than those with type I disease (53.0 years).[5] Some patients with typical features of autosomal dominant polycystic kidney disease have no mutations in *PKD1* or *PKD2,* suggesting that there may be a rare third form of the disease [6] although the proposed gene —*PKD3*— has not been identified. Patients often present with hypertension, hematuria, polyuria, and flank pain and are prone to recurrent urinary tract infections and renal stones. In addition to the presence of hundreds to thousands of renal cysts, clinically significant cysts are also common in the liver (especially in women), pancreas, and intestine. Patients

frequently experience complications involving various extrarenal manifestations, such as intracranial aneurysms, colon diverticular disease, and liver cysts.

ADPKD is characterized by tremendous cystic growth of both kidneys resulting in bilateral kidney enlargement. Franz and Reubi[7] proposed that renal function remains stable in ADPKD patients for years, followed by a sharp decline once a critical renal size is reached. Increased renal volume predicts and is associated with loss of renal function in ADPKD.[8] However, standard radiographic imaging has not provided the resolution and accuracy necessary to detect small changes in renal volume or to reliably measure renal cyst volumes.

Renal function is measured by glomerular filtration rate (GFR) as a continuous measure, or by categorizing GFR values below 60 as chronic kidney disease (CKD). GFR can either be estimated with equations based on serum creatatine, age, sex, and race, or measured directly with iothalamate clearance. Another outcome related to renal decline is ESRD, where the patient requires a transplant or dialysis; stage 5 CKD, where GFR is below 15, may also be grouped with ESRD.

1.2 THE CRISP STUDY

The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) was established to develop innovative imaging techniques and analyses using magnetic resonance (MR) imaging to measure cyst and renal volume reliably and accurately in ADPKD individuals early in the course of their disease. The study started in 2001 and is concurrently ongoing. They enrolled 241 patients with ADPKD who were 15 to 46 years of age and who were evaluated annually (for a total of four visits) over a period of three years beginning in January 5, 2001.

CRISP was subsequently funded again in 2006, 2011, and 2017 (for 5 year periods). These subsequent 5-year periods included two to three clinic visits with subsequently described imaging and outcome measures. Eligible patients had received a diagnosis of ADPKD, had an actual or estimated (by the Cockcroft–Gault equation) creatinine clearance of at least 70 ml per minute, and had a serum creatinine level of 1.6 mg per deciliter (141 μmol per liter) or less in the case of male patients and 1.4 mg per deciliter (124 μmol per liter) or less in the case of female patients. Patients were ineligible if they had other medical conditions besides hypertension that could affect renal function (e.g., diabetes mellitus). The CRISP cohort study longitudinally observed ADPKD individuals using high-resolution magnetic resonance (MR) imaging to determine if change in renal and cyst volumes can be detected over a short period of time, and if they correlate with decline in renal function at early stage of the disease.

The CRISP study showed that MR measures of renal and cyst volume are reliable and accurate in patients with ADPKD. ADPKD is characterized by significant cystic involvement that increases with age. Structure (renal and cyst volume) and function (GFR) are inversely related and directly related with the presence of hypertension and urinary albumin excretion in individuals with normal renal function.[9] During the initial CRISP study period of 3 years, CRISP found that kidney enlargement resulting from the expansion of cysts in patients with ADPKD is continuous and quantifiable. Total kidney volume and total cyst volume increased exponentially, and the baseline total kidney volume predicted the subsequent rate of increase in volume, independently of age. Higher rates of kidney enlargement are associated with a more rapid decrease in renal function.[10] At the baseline initial study visit, adult men had greater mean TKV than adult women with a ratio of 1.15. In order to correct for other factors influencing, TKV was referenced initial study to baseline height, weight, body surface area, or

BMI in order to diminish the sex differences. From this analysis, height was the best reference for TKV (htTKV), with a male/female ratio of 1.037 and was used thereafter in CRISP study.[11] CRISP reported with 8 years of follow-up had found increasingly strong associations between baseline htTKV and the follow-up iothalamate clearances and progression through the K/DOQI stages.[11] These observations demonstrate that renal cyst burden, reflected by htTKV, is a very important determinant of renal functional decline in ADPKD.

This current analysis seeks to better assess causal effects of htTKV on renal decline. So far, the methods used in CRISP, and in other studies of predictors of renal function, have used generalized linear models (including mixed models for repeated longitudinal measurements) to assess associations. For this study, we consider htTKV as the exposure or "treatment" and describe and apply the concept of the generalized propensity score to better estimate the average causal effect. Estimating causal effects is challenging in observational studies, since the treatments of interest (e.g., screening by echocardiography and use of antidepressants in pregnancy) are not randomly allocated, and important characteristics differ between groups. Similarly, htTKV may be related to other factors that confound the association with the outcome. Direct comparisons of the outcomes between treated and untreated groups, or by kidney volume in CRISP, would have likely resulted in significantly biased estimates.

1.3 THE POTENTIAL OUTCOME FRAMWORK FOR CAUSAUL INFERENCE

1.3.1 Association versus Causation

Most analyses of observational studies examine the association between an exposure (e.g., a food, something in the environment, or a biomarker) and an outcome (often a disease or death). Because of all the other exposures occurring simultaneously in the complex lives of free-living humans that can never be completely accounted for, such studies cannot provide evidence of cause and effect using standard statistical methods. In other words, without using special methods for causal inference in the setting of a well-designed observational study, such studies cannot distinguish direction–whether exposure A influences outcome B, or B influences A, or both are influenced by something else, even if that association may be strong and consistent. In CRISP study, based on the data what we can say is: Increased kidney volume is associated with subsequently decreased renal function. But we want to say that increased volume causes subsequently decreased renal function. In order to use kidney volume as a surrogate marker for ADPKD, i.e. as a measure that can substitute for the outcome, we have to establish the causation between them.

1.3.2 The Potential Outcomes Framework for Causal Inference

Causal inference is the process of drawing a conclusion about a causal connection based on the conditions of the occurrence of an effect. A number of different frameworks exist for causal inference, and the assumptions depend on the framework being utilized. For this analysis, we define causal effects in terms of potential outcomes. [12] Briefly, in the simple case of a

dichotomous treatment with two levels A and B, the individual causal effect (which cannot be directly estimated) is defined as the difference between the observed outcome and the other potential outcome, i.e. the counterfactual (or some function of that difference depending on the outcome distribution). Say individual i receives treatment A at time 0; the outcome measured at time 1 is denoted as Y_i (A). The potential (unobserved) outcome for individual i on treatment B is defined as Y_i (B). We can then define the average treatment effect (ATE) as the expectation of those differences: ATE = E [Y_i (A) - Y_i (B)].

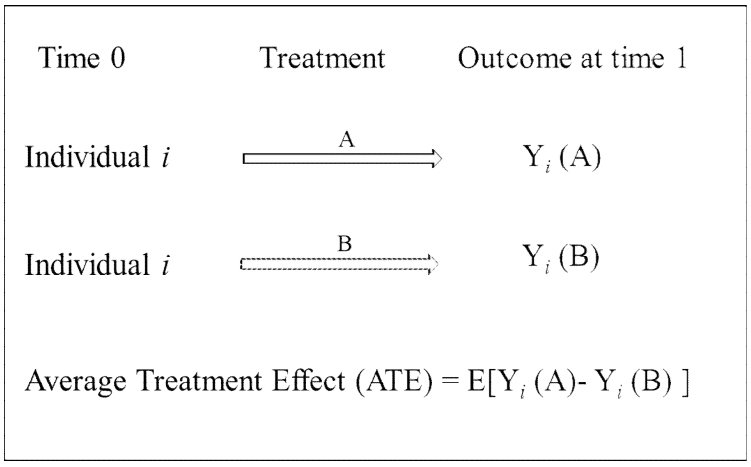

Figure 1. The concept of potential outcomes

In randomized trial studies, the ATE can be estimated without bias by simple differences in the group A and group B mean outcomes since treatments are assigned randomly. However, in observational studies, the causal effects (ATE) are not estimated by the simple difference in expectations: ATE ≠E [Y_i (A)| X] – E [Y_i (B)| X], because the propensity to one treatment versus another is typically affected by many other factors, such as the physician choice, patient factors and institutional factors.

1.4 METHODS FOR ESTIMATING TREATMENT EFFECTIVENESS

A wide range of methods are available for analyzing treatment effectiveness from observational data to reduce or eliminate the effects of confounding factors, such as stratifying results by a single or multiple covariates using standard statistical tests or statistics. For example, the 2×2 contingency tables were extended to multiple contingency tables with arbitrarily many rows and/or columns, where rows and columns are orderable, and may even be on a continuous scale. With scores assigned, a deviation of the sum of cross products from expectation, and its variance conditioned on all marginal totals, are computed for each table and a chi square is determined corresponding to the grand total of the deviations. The procedure extends or is equivalent to the asymptotic form of many known non-parametric techniques.[13]

A second approach is to match subjects on a specific characteristic(s) of interest and conduct paired analyses. For example a paired samples t-test based on a matched-pairs sample, which results from an unpaired sample that is subsequently used to form a paired sample, by using additional variables that were measured along with the variable of interest.[14] The matching is carried out by identifying pairs of values consisting of one observation from each of the two samples, where the pair is similar in terms of other measured variables. This approach is sometimes used in observational studies.

A third approach is to use regression models. Historically, applied researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between treated and untreated subjects. Logistic regression is a commonly used method to control for imbalances between groups. Its primary advantage is the ability to control for many variables simultaneously.[15]

The first two approaches offer very limited flexibility in terms of the number of potential confounders and manner in which the variable is characterized. The third method is also problematic of the above if too many variables need to be included in a model relative to the number of events, the estimates from these models can be incorrect. Regression and stratification methods also do not provide unbiased estimates of causal effects.[16]

One set of approaches that better estimate the causal effect under more realistic assumptions are propensity score (PS) methods. The propensity score was defined by Rosenbaum and Rubin [17] to be the probability of treatment assignment conditional on observed baseline covariates. The propensity score is a balancing score: conditional on the propensity score, the distribution of measured baseline covariates is similar between treated and untreated subjects. Thus, in a set of subjects all of whom have the same propensity score, the distribution of observed baseline covariates will be the same between the treated and untreated subjects. PSs are typically estimated with a logistic regression model that regresses the exposure variable on observed confounders. The estimated propensity score is the predicted probability of treatment derived from the fitted regression model. PS methods enable the investigators to create study groups that were similar to one another and more accurately measure the relationship between treatment and outcome.

Four different propensity score methods are most commonly used for removing the effects of confounding when estimating the effects of treatment on outcomes: matching on the propensity score, stratification on the propensity score, inverse probability of treatment weighting (IPTW) using the propensity score, and covariate adjustment using the propensity score [18]. Propensity score matching entails forming matched sets of treated and untreated subjects who share a similar value of the propensity score. Stratification on the propensity score

involves stratifying subjects into mutually exclusive subsets based on their estimated propensity score. Typically, subjects are stratified based on quantiles of the propensity score. Subjects are ranked according to their estimated propensity score. Subjects are then stratified into subsets based on previously defined thresholds of the estimated propensity score. IPTW using the propensity score uses weights based on the propensity score to create a synthetic sample in which the distribution of measured baseline covariates is independent of treatment assignment. With covariate adjustment using the propensity score, the outcome variable is regressed on an indicator variable denoting treatment status and the estimated propensity score. The choice of regression model would depend on the nature of the outcome. These propensity score methods allow one to design and analyze an observational study so as to mimic some of the characteristics of a randomized study.[19]

Much of the work on propensity score analysis has focused on the case where the treatment is binary. Imbens extends the method to the multi-group to reduce bias in observational studies where treatment can have several levels.[20] His work involves calculating an average treatment effect by weighting observations by the inverse of the probability of treatment level actually observed. Jiang and colleagues have used both binary and multi-group approaches to study the effects of breastfeeding initiation and duration on child cognitive outcomes.[21] The generalized propensity score (GPS) for continuous treatments is a straightforward extension of the well-established and widely used propensity score methodology for binary treatments and multi-valued treatments. Similar to the binary and multivalued treatment propensity score methods, it is assumed that – conditional on observable characteristics – the level of treatment received can be considered as random. In a setting in which doses (usually treatment) are not administered under experimental conditions, estimation of a dose-response function (outcome) is

possible using the GPS. Hirano and Imbens show that the GPS has a balancing property similar to the balancing property of the "classic" propensity score.[22] This implies that individuals within the same strata of the GPS should be identical in terms of their observable characteristics, independent of their level of treatment.

1.5 THE AIM OF THIS STUDY

In this study we examined an extension to the propensity score method: the GPS method, in a setting with a continuous measure, which is a biomarker (baseline htTKV) rather than a treatment or exposure. However, the use of the GPS applies equally well to the case of this biomarker as it does for the originally developed application of continuous treatment doses. We assessed the causal relationship between htTKV and progression of ADPKD (as measured by GFR at least a decade later) using a subgroup of the CRISP cohort study. To accomplish this, we apply essentially the same method as developed for estimating a dose-response function as proposed in Hirano and Imbens.[22] Specifically, in this study we estimated the causal relationship between the continuous response of GFR across the range of values of the continuous htTKV. In terms programming the method, we used a set of Stata programs to estimate the GPS, test whether the estimated GPS satisfies the balancing property, and predicted the dose–response function. We estimated the ADPKD patients' average GFR after at least 10 years from baseline. Participants were also included if they reached end-stage renal disease (ESRD). Baseline htTKV was recorded when they were enrolled in the study. These estimates were adjusted for differences in characteristics available at birth (to maintain temporal associations) using the generalized propensity score methodology.

2.0 METHODS

2.1 THE STUDY DESIGH

CRISP is a prospective, longitudinal cohort study, which began in 2001 and is currently ongoing. The data in this paper covered the study period until to 2014. CRISP enrolled participants ages 15 to 46 years who met the following criteria: (1) Diagnosis of ADPKD; (2) Actual or estimated creatinine clearance of at least 70 mL/min; (3) Serum creatinine level of ≤ 1.6 mg/deciliter for men and ≤ 1.4 mg/deciliter for women. Patients were ineligible if they had other medical conditions besides hypertension that could affect renal function (e.g., diabetes mellitus).The initial study enrolled 241 participants. Enrollees were interviewed by telephone every 6 months and were evaluated in standardized fashion during clinic visits every 12 months. Total kidney volume (TKV) and glomerular filtration rate (GFR) were measured each of the first 3 years of the study and approximately every 2 years thereafter. TKV was measured by MRI[23] and referenced to height (htTKV, cc/m). GFR was measured by a nonradiolabeled iothalamate clearance technique with sonographic monitoring of bladder emptying [24] and was referenced to body surface area adjusted to a fixed norm (milliliters per minute per 1.73 m^2). Our primary outcome was the GFR at the last visit for each subject (which was usually at 10-12 years after baseline). For subjects who reached ESRD at any time in the study, a value of 10 was imputed for their GFR, since it is effectively 10 at the time of kidney failure. For this analysis, we

excluded participants who were lost-to-follow-up (i.e. did not have a GFR and imaging measurement) before 10 years and did not yet reach ESRD.

There were a total of 186 participants who were followed for at least ten years or they reached the endpoint of end-stage renal disease (ESRD). CRISP includes a wide range of variables on demographics, medical conditions and hospitalizations, imaging measures (which includes htTKV), and urine and serum biomarkers measured at baseline and (for most variables) at the clinic visits. However, this analysis concentrates on variables which were measured at birth since we can definitively say they occurred before, and are not a result of, the patient's htTKV. More specifically, among the variables in the original data set, we chose gender, race, genotype (PKD1, PKD2, or no mutation detected), truncation (yes or no for whether the mutation includes frame shifting, nonsense or splicing) and birth season, which were the only variables that were from birth. The htTKV measurement was the main predictor of interest.

2.2 STATISTICAL ANALYSIS

2.2.1 Descriptive analyses

Means, standard deviations (SDs), minimum values, maximum values, medians and interquartile ranges (displayed as the median [IQR]) were provided to describe htTKV and GFR values over time. Gender, race, genotype, truncation and birth season were categorical variables and described as frequencies and percentages.

2.2.2 Associations with htTKV and GFR

To assess the association of each covariate with baseline htTKV and last visit GFR, simple linear regression models were fitted between gender, genotype, truncation, race, birth season and htTKV, GFR respectively. These analyses were part of the process of calculating the GPS, as described below. The coefficients were reported.

2.3 THE GENERALIZED PROPENSITY SCORE METHOD

2.3.1 GPS definition

As described in the introduction, the GPS is a variation of traditional propensity methods where we model the probability of treatment (T) given patient characteristics. In the binary treatment case, we postulate for each individual there are the potential outcomes Yi(t), for t ϵT, here T=(0,1). However, for a continuous treatment T is an interval $[t_0, t_1]$ and we are interested in the average dose-response function, u(t) = E $[Y_i(t)]$.

Let $r(t, x)$ be the conditional density of the treatment given the covariates X: $r(t, x) = f_{T|X}(t|x)$.

The generalized propensity score (R) is then defined as

(1) $R = r(T, X)$.

2.3.2 Assumption of the GPS

The key assumption of Hirano and Imbens [22] generalizes the unconfoundedness assumption for binary treatments made by Rosenbaum and Rubin [17] to the continuous case:

(2) $Y(t) \perp T \mid X$ for all $t \in T$

It is referred as weak unconfoundedness as we do not require joint independence of all potential outcomes, $\{Y(t)\}(25)_{t \in [t0, tI]}$. Instead, we require conditional independence to hold for each value of the treatment.

The GPS has a balancing property similar to the balancing property of the propensity score for binary treatments. Within strata with the same value of $r(t,X)$ the probability that $T=t$ does not depend on the value of covariates X.

This is an implication of the definition of the GPS and does not require weak unconfoundedness. In combination with weak unconfoundedness, it implies that assignment to treatment is unconfounded given the GPS. It has been showned that if assignment to treatment or exposure (or in our case the value of kidney volume) is weakly unconfounded given covariates X, then it is also weakly unconfounded given the GPS[22].

2.3.3 The procedures of GPS method

Given the result that if assignment to treatment is weakly unconfounded given covariates X, then it is also weakly unconfounded given the GPS, it is possible to use the GPS to remove bias associated with differences in covariates in the following steps [16]. To be consistent with the notation in the literature on GPS, we refer to treatment, denoted by T_i, but the same approach applies to estimating causal effects of the biomarker value.

2.3.3.1 Model and estimation of the GPS

We assume a normal distribution for the baseline htTKV given the covariates.

(3) $T_i | X_i \sim N(\beta_0 + \beta_1' X_i, \sigma^2)$

We use normal Quantile-Quantile (Q-Q) plots to check the normality of htTKV. Since htTKV was not normal (see Results), data was then be log transformed to give the following equation:

(4) $\log T_i | X_i \sim N(\beta_0 + \beta_1' X_i, \sigma^2)$

In the simple normal model we can estimate β_0, β_1, and σ^2 by maximum likelihood. The estimated GPS was calculated as

(5) $\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}(T_i - \hat{\beta}_0 - \hat{\beta}'_1 X_i)^2\right)$

2.3.3.2 Removal of bias by using the generalized propensity score

In the case of a continuous measurement it is also crucial to evaluate how well adjustment for the GPS works in balancing the covariates.

We assessed the covariates' balance by using K square test before the GPS adjustment, as suggested by Hirano and Imbens [22]. We divided the sample into three groups according to the distribution of length of baseline htTKV, cutting at 33th, 66th percentile. For each of the covariates, we investigated the balance by testing whether the frequency in one of the three treatment groups was different from the rest of the other samples.

Then we investigated how GPS affected the balance of the covariates. In the binary case the typical approach is to compare the covariate means for the treated and control units before and after matching, testing for covariate balance is more difficult with continuous measurements. We followed Hirano and Imbens' approach of "blocking on the score".

First, the sample was divided into three groups as described above. Within each group, we evaluated the GPS at the median of htTKV. In the second step we divided each group into five blocks by the quintiles of the GPS evaluated at the median. Within each of these blocks, we calculated the difference-in-means of covariates with respect to individuals that had a GPS such that they belonged to that block, but had a htTKV level different from the one being evaluated. This procedure tested for each of these blocks whether the covariate means of individuals belonging to the particular htTKV-level group were significantly different from those of individuals with a different htTKV level, but similar GPS. A weighted average over the five blocks in each htTKV level group can be used to calculate the t-statistic of the differences-in-means between the particular htTKV level group and all other groups. The procedure needed to be repeated for each of htTKV level group and for each covariate. If adjustment for the GPS properly balances the covariates, we would expect all those differences-in-means not to be statistically different from zero.

2.3.3.3 Estimate the conditional expectation of the outcome

With the GPS method, we estimated the conditional expectation of the outcome last visit's GFR as a function of two scalar variables, the htTKV level *T* and the GPS *R*, i.e.

(6) $\beta(t,r) = E[Y \mid T = t, R = r]$.

For the estimation of the equation, we had to assume some functional form of the relationship between the last GFR *Y*, the htTKV level *T*, and the GPS *R*.

Following the general approach proposed by Hirano and Imbens, we assessed the correlation pattern between GFR and T_i (htTKV), GFR and GPS (Ri) respectively, and tested for addition of the interaction term of T and R. Depending on those results, we then choose one of the following polynomials.

(7) $E(Y_i \mid T_i, R_i)= \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i$

(8) $E(Y_i \mid T_i, R_i)= \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 + \alpha_4 R_i + \alpha_5 R_i^2 + \alpha_6 R_i^3 + \alpha_7 T_i R_i + \alpha_8 T_i^2 R_i + \alpha_9 T_i R_i^2$

2.3.3.4 Estimate the dose-response function at each particular level of the treatment

This is implemented by averaging the conditional expectation function over the GPS at that particular level of the treatment,

(9) $\mu(t)= E[\beta(t,r(t,X))]$

The procedure does not average over the GPS $R=r(T, X)$, but instead it averages over the score evaluated at the htTKV level of interest $r(t,X)$.

For each individual the observed htTKV (T_i) and estimated GPS $\hat{R}_i$ were used, and the equation was estimated by ordinary least squares. Given the estimated parameters, if we used a quadratic approximation, the average potential outcome at htTKV level t was estimated as:

(10) $\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_0 + \hat{\alpha}_1 \cdot t + \hat{\alpha}_2 \cdot t^2 + \hat{\alpha}_3 \cdot \hat{r}(t, X_i) + \hat{\alpha}_4 \cdot \hat{r}(t, X_i)^2 + \hat{\alpha}5 \cdot t \cdot \hat{r}(t, X_i))$

where $\hat{\alpha}$ is the vector of the estimated parameters in the second stage.

The above function can then be obtained by estimating this average potential outcome for each level of the treatment. In our application, we used bootstrap methods to obtain standard errors that take into account estimation of the GPS and α parameters, i.e. we bootstrapped the entire estimation process.

3.0 RESIUTS

3.1 BASELINE CHARACTERISTICS

The mean of baseline htTKV was 638.13 cc/m, the minimum was193.82 cc/m, the maximum was 2113.12 cc/m, and the median was 507.37 cc/m. The mean of last visit's GFR was 62.87 ml/min per 1.73m^2, from the minimum value of 10 ml/min per 1.73m^2 to the maximum value of 172 ml/min per 1.73m^2 (Table 1).

Among the participants, 59.14% was female, 79.89% was PKD1 mutation, 69.94% of the patients had gene truncation, 10.75% was black and about 20-30% were born at each of the four seasons (Table 2).

Table 1. Summary statistics of the variables htTKV and last visit's GFR

Variable	n	Mean	Min	Q25	Q50	Q75	Max
htTKV(cc/m)	186	638.13	193.82	362.15	507.37	848.57	2113.12
GFR(ml/min/1.73m^2)	186	62.87	10	25	60	94	172

Table 2. Summary statistics of baseline characteristics

Variable		Frequency	Percent (%)
Gender*	Male	75	40.9
	Female	110	59.1
Genotype*	NMD/PKD2	37	20.1
	PKD1	146	79.9
Truncation	Non-Truncating	51	30.1
	Truncation	121	69.9
Race	Others	166	89.3
	Black	20	10.8
Season	Spring	54	29.0
	Summer	41	22.0
	Fall	55	29.6
	Winter	36	19.4

*Note: Gender: * one missing observation; ** three missing observations.*

3.2 REGRESSION COEFFICIENTS TEST

Among the baseline characteristics, only genotype is significant associated with baseline htTKV (p=0.003) (Table 3). Genotype is significantly associated with last visit's GFR (p=0.019) as well (Table 4).

Table 3. Association between each of the covariates and htTKV

Variable		Coef.	P value
Gender		-56.83	0.329
Genotype		212.49	0.003
Truncation		13.56	0.836
Race		-65.50	0.483
Season*	Spring	-47.38	0.931
	Fall	-48.81	
	Winter	-30.47	

*Note: *With Summer as the baseline*

Table 4. Association between each of the covariates and outcome of GFR

Variable		Coef.	P value
Gender		6.30	0.311
Genotype		-17.82	0.019
Truncation		5.35	0.448
Race		12.90	0.189
Season*	Spring	10.71	0.590
	Fall	8.70	
	Winter	9.63	

*Note: *With Summer as the baseline*

3.3 THE APPLICATION OF GPS

3.3.1 Modelling the conditional distribution of htTKV given the covariates

The distribution of the htTKV was skewed with a skewness of 1.40 and a kurtosis of 4.77. The Q-Q plots also showed a systematic deviation from normality. We therefore used a log transformation. The logarithm of the htTKV was approximately normal with a skewness of 0.29 and a kurtosis of 2.31 (Figure2 and 3). We then used a normal linear model for the logarithm of htTKV.

Histogram **Q-Q plots**

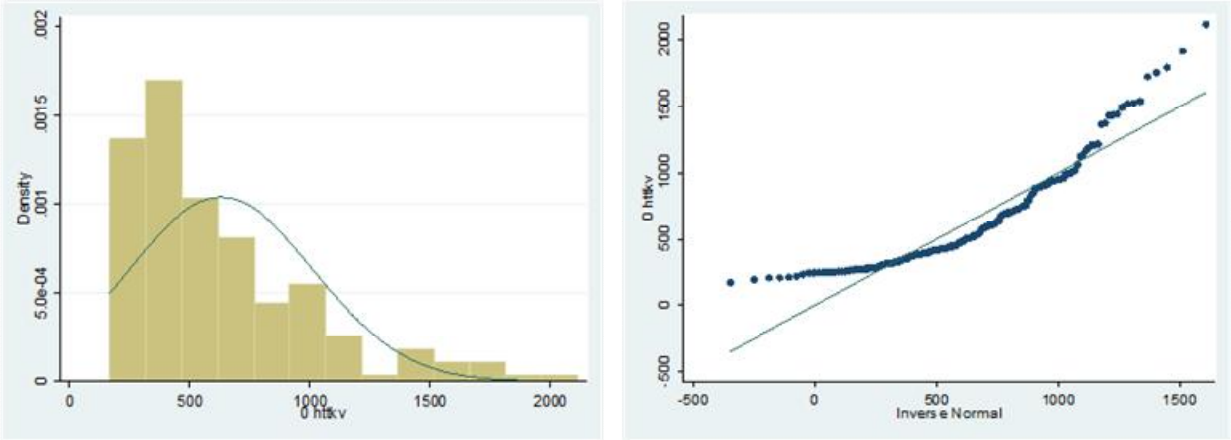

Figure 2. Histogram and Q-Q plots of htTKV before logarithms transformation

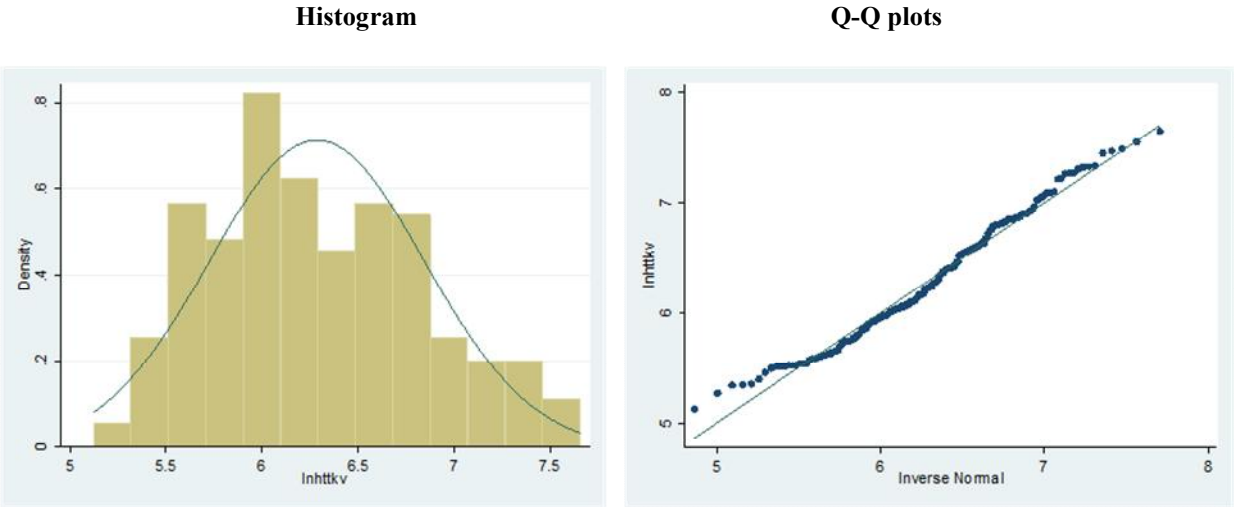

Figure 3. Histogram and Q-Q plots of htTKV after logarithms transformation

The estimated coefficients (GPS Est) and standard error of GPS in model (4) are presented in Table 5. There is no direct meaning to the estimated coefficients in this model, except that testing whether all coefficients for the GPS were equal to zero can be interpreted as a test whether the covariates introduce any bias.

Table 5. Estimated coefficients for the GPS

Variable		GPS Est.	GPS SE
Intercept		6.218	0.1672
Gender		-0.088	0.0854
Genetype		0.2628	0.1196
Truncation		0.0106	0.0906
Race		-0.0774	0.1440
Season*	Spring	-0.0819	0.1172
	Fall	-0.1135	0.1181
	Winter	-0.0650	0.1332

*Note: *With Summer as the baseline*

3.3.2 Removal of bias by using the generalized propensity score

In the case of a continuous measurement it is also crucial to evaluate how well adjustment for the GPS works in balancing the covariates. We assessed the covariates' balance by using K square test before the GPS adjustment.

In Table 6, we reported the k square test for each of the five covariates in each of the three groups before GPS adjustment. The results showed 4 statistic tests had significant differences.

Table 6. Balance of the covariates: k square test for equality of frequency

Variable		Unadjusted: chi2 (*p*)		
		[0; 395.15]	[395.15;697.83]	[697.83; 2113.12]
Gender		0.37(0.541)	0.09(0.769)	0.80(0.371)
Genetype		7.40(0.007)	0.08(0.774)	8.84(0.003)
Truncation		0.08(0.783)	0.01(0.963)	0.04(0.826)
Race		7.52(0.006)	5.28(0.022)	0.19(0.660)
Season	Spring	0.87(0.351)	1.28(0.258)	0.04(0.843)
	Summer	0.85(0.357)	0.03(0.866)	1.16(0.282)
	Fall	2.88(0.089)	1.08(0.299)	0.42(0.515)
	Winter	0.01(0.939)	0.01(0.939)	0.02(0.880)

Then we investigated how GPS affects the balance of the covariates. After the adjustment for the GPS, none of the *p* values were greater than 0.05, showing that the GPS eliminated any significant imbalances (Table 7).

Table 7. Balance given the generalized propensity score: t-statistics for equality of medians

Variable		Adjusted for GPS: t (p)		
		[0; 395.15]	[395.15;697.83]	[697.83; 2113.12]
Gender		0.723(0.235)	-0.532(0.702)	-0.718(0.763)
Genetype		1.414(0.795)	-0.297(0.617)	-1.211(0.886)
Truncation		-0.151(0.560)	0.014(0.494)	-0.199(0.579)
Race		-1.028(0.847)	1.522(0.065)	-1.153(0.875)
Season	Sping	0.623(0.267)	-1.145(0.873)	0.973(0.166)
	Summer	-0.192(0.576)	0.195(0.423)	-1.304(0.568)
	Fall	-0.991(0.161)	1.037(0.151)	-0.172(0.568)
	Winter	0.552(0.291)	0.044(0.482)	0.162(0.436)

3.3.3 Model the conditional expectation of Y_i and R_i

We checked the correlation pattern between GFR and T_i (htTKV), GFR and GPS (R_i) respectively, to determine whether to use linear regression or non-linear regression. From figure 4 we can see, GFR and T_i were negative non-linear correlated. Figure 5 showed that GFR and Ri were positive and non-linear correlated. Therefore we used a quadratic approximation of Ti and Ri. Because Ri and Ti are in opposite direction (Figure 4 and Figure 5), an interaction term of Ti and Ri was included in the model as well. Therefore model (7) in 3.3.3 of "methods" section was used.

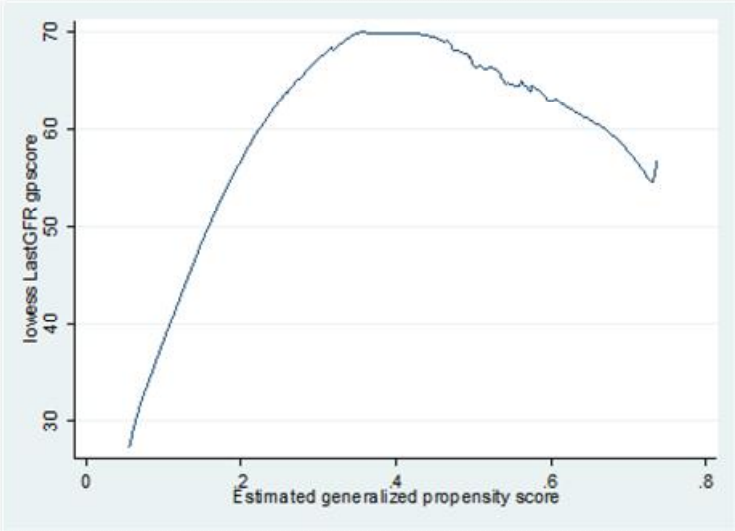

Figure 4. Fitted curve of GPS and GFR

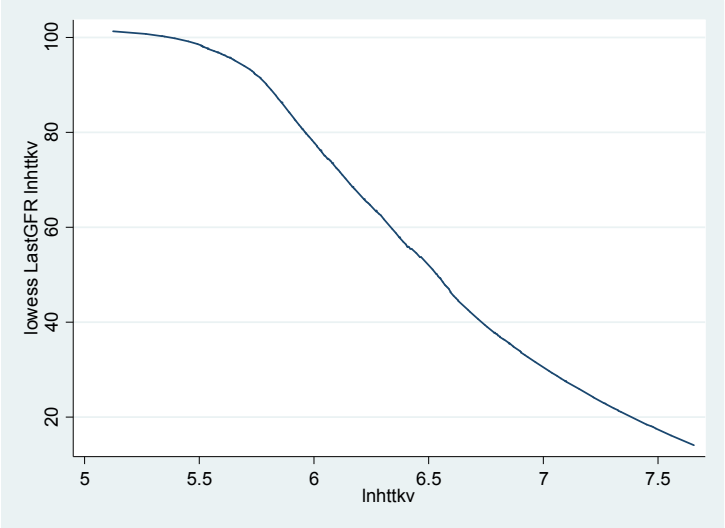

Figure 5. Fitted curve of htTKV and GFR

3.3.4 Averaging the estimated regression function over the score function evaluated at the desired level of the htTKV

The estimated GFR over each value of htTKV was obtained by estimating the average potential outcome (last visit GFR) over possible values of baseline htTKV. The equation (10) in 3.34 of "methods" section was used in this procedure. In our application, we use bootstrap methods to obtain standard errors that take into account estimation of the GPS and α parameters, i.e. we bootstrap the entire estimation process.

Figures 6 shows the shape for estimated GFR outcomes based on the baseline htTKV, which indicates a negative correlation between baseline htTKV and last visit's GFR. As we would expect, the expected GFR decreases with increasing htTKV until leveling out at higher kidney volumes. The apparent increase near the end is likely chance with little data and wide confidence intervals in that range. The estimated curve shows that if htTKV value is more than 520 cc/m at the first visit, participants would progress to at least stage 3 CKD or an GFR<60ml/min per 1.73m^2 in 10 years.

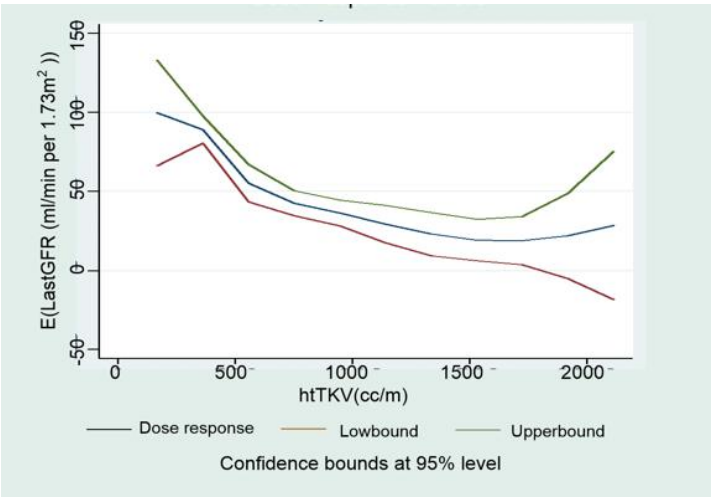

Figure 6. Expected GFR and Confidence Interval Based on the GPS

4.0 CONCLUSION AND DISCUSSION

Kidney enlargement resulting from the expansion of cysts in patients with ADPKD is continuous and quantifiable and is associated with the decline of renal function. Usually physicians and patients monitor changes in serum creatinine levels to determine the extent of progression. However, serum creatinine levels do not typically rise in patients with ADPKD until the fourth or fifth decade of life, after the noncystic parenchyma has incurred serious, irreversible damage. Therefore, creatinine levels are usually ineffective for early detection and prevention. Our results showed that the larger values of baseline htTKV appear to have a causal relationship with subsequent renal decline within the following decade. Therefore, htTKV offers a high potential for earlier detection and may be very useful for prevention efforts and early interventions to reduce incidence of ESRD. htTKV may therefore also be useful in clinical practice and for risk stratification in designing clinical trials for this disorder.

There are several practical reasons for preferring the use of propensity score based methods to regression-based methods when estimating treatment effects using observational data. First, it is simpler to determine whether the propensity score model has been adequately specified than to assess whether the regression model relating treatment assignment and baseline covariates to the outcome has been correctly specified.[19] The diagnosis of whether the propensity score model has been adequately specified diagnostics were based on comparing the distribution of measured baseline covariates between treated and untreated subjects. In contrast,

it is much more difficult to determine whether the regression model relating treatment selection and baseline covariates to the outcome has been correctly specified. Goodness-of-fit measures, such as model R^2, do not provide a test of whether the outcome model has been correctly specified. Furthermore, goodness-of-fit tests do not allow one to determine the degree to which the fitted regression model has successfully eliminated systematic differences between treated and untreated subjects. Second, it has been reported that approaches using the propensity score estimate less biased estimators than regression analysis when there are seven or fewer events per confounder variable in simulation studies.[26] Moreover, the propensity score approaches do not need unrealistic assumptions to estimate causal effects in an unbiased manner, and these estimates are robust with regard to model misspecification.[27]

In this study, we applied the approach developed by Hirano and Imbens[22] who propose estimating the entire dose-response function (DRF) of a continuous treatment. This approach fits perfectly with the objective of our analysis, since we are interested in the response (declining renal function) associated with each value of the continuous measure of htTKV. Alternatively we could categorize the continuously distributed variable for volume and apply propensity score methods for multi-valued treatments. But the GPS has the advantage that it makes use of the entire information contained in the distribution of htTKV. The GPS has balancing properties that can be used to assess the adequacy of particular specifications of the score.

In summary, this study illustrates the use of the GPS for making inferences about levels of a biomarker and its causal effects on renal function. While the variables we collected at birth were limited and often not significantly associated with the outcome, the study still serves to illustrate the use of this method in the setting of a continuous biomarker, which is a novel application of these methods. Future studies should explore considering other variables which

are measured earlier in life (but after birth) that may still precede kidney growth. The significance of these methods and the CRISP study is emphasized by continuing interest in htTKV as a biomarker for early detection and prevention.

APPENDIX A. DOSE-RESPONSE FUNCTION ANALYSIS WITH AGE STRATIFICATION

Polycystic kidney disease is a life-long condition, the cysts start grow at birth and patients reach ESRD at the median age of 54 years old for PKD1 mutation and 74 years for PKD2 mutation. As we showed in this study that the larger values of baseline htTKV appear to have a causal relationship with renal decline possibly a decade later. So we suspect that those subjects with an earlier age of onset of an enlarged htTKV would have a more severe disease and hence earlier age of onset of renal impairment than subjects who manifested the disorder later in life. Therefore we stratified our observations into two sets by median age 44.7 years old (Table 8), and obtained expected GFR based on the GPS (dose-response function analysis) and the results were shown at figure 7.

Figure 7 indicated a negative correlation between baseline htTKV and last visit's GFR: slowly and monotonously decreasing GFR response to the increasing of the htTKV in both of the younger than median age and older than median ones. With the same baseline htTKV, the participants in the older group had a lower last visit GFR than the younger group, indicating that age was an important factor in determining renal function.

The graphs also showed that the slope of figure 7a was deeper than figure 7b, which indicated that those subjects with an earlier age of onset of enlarged kidney volume would have a more severe disease and hence earlier age of onset of renal impairment than subjects who manifested the disorder later in life. It was also shown that children with onset in utero or in the first year of life appeared to do worse.[25, 28]

Table 8. Summary statistics of the variable age

Variable	n	Mean	Min	Q25	Q50	Q75	Max
Age (Year)	186	44.08	26.41	28.37	44.70	51.18	59.15

Figure 7a.Age younger than median age

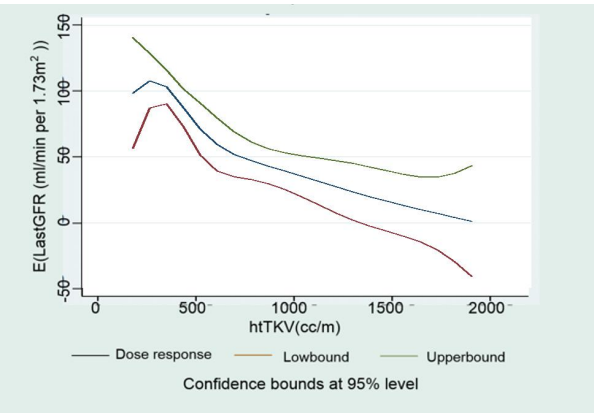

Figure 7b. Age older than median age

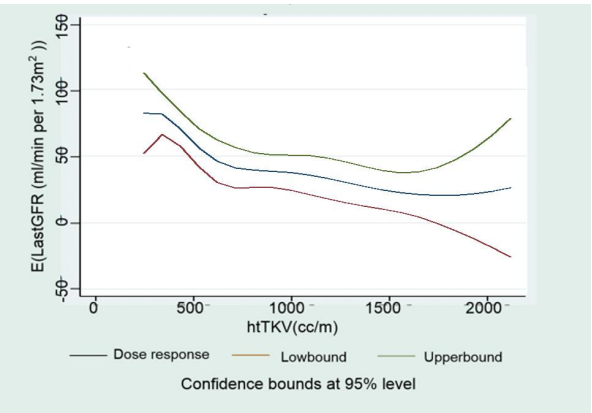

Figure 7. Expected GFR and Confidence Interval Based on the GPS stratified by median age

APPENDIX B. STATA CODE

```
use "C:\Users\Yaming\Desktop\Thesis\Yaming thesis dataset4.dta" ,clear
 *clean data
 sort pkdid vis
 by pkdid: gen httkv0=httkv[1]
 by pkdid: gen N=_N
 by pkdid: gen n=_n
 keep if n==N
 keep if visc>=10 | esrd==1
 drop cic_c race gender birthdate genetype trunc_grp race4 httkv brwgt UMOD_creat months
httkv_18 N n
 * Summary statistics of  htTKV and last visit's GFR
sum httkv0 LastGFR, detail
* Summary statistics of  baline characteristics
gen summerseason=0
replace summerseason=1 if Fallseason==0 & winterseason==0 & Springseason==0
tabulate gender1
tabulate genetype1
tabulate trunc_grp1
tabulate race1
tabulate Springseason
tabulate summerseason
```

```
tabulate Fallseason

tabulate winterseason

*Normality test

histogram httkv0, norm

sum httkv0,detail

gen lnhttkv=ln(httkv0)

histogram lnhttkv,norm

sum lnhttkv,detail

qnorm httkv0

qnorm lnhttkv

*regression coefficciants test

reg lnhttkv gender1 genetype1 trunc_grp1 race1

foreach v of varlist gender1 genetype1 trunc_grp1 race1 {

reg  httkv0 `v'

}

foreach v of varlist gender1 genetype1 trunc_grp1 race1 {

reg LastGFR `v'

}

reg httkv0 i.seasons

reg LastGFR i.seasons

* Balance test bedore GPS

egen p33 = pctile( httkv0 ), p(33)

egen p66 = pctile( httkv0 ), p(66)
```

```
quietly generate cut = 395.15 if httkv0<=395.15
quietly replace  cut = 697.84 if httkv0>395.15 & httkv0<=697.84
quietly replace  cut =2113.12  if httkv0>697.84
tab cut
gen group=1 if cut<=395.15
replace group=2 if cut<=697.85 & cut>395.15
replace group=3 if cut>697.85
gen group1=1 if group==1
replace group1=2 if group==2 | group==3
gen group2=1 if group==1 | group==2
replace group2=2 if group==3
gen group3=1 if group==1 | group==3
replace group3=2 if group==2
tab cut
foreach v of varlist gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason
winterseason {
tabulate  group1 `v', chi2
}
foreach v of varlist gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason
winterseason {
tabulate  group2 `v', chi2
}
```

```
foreach v of varlist gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason {
tabulate  group3 `v', chi2
}
*dose-response function test
gpscore gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason,t(httkv0) gpscore(gpscore) predict(y_hat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) detail
drop sd
doseresponse gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason, outcome(LastGFR)t(httkv0)gpscore(psore) predict(hat_treat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) dose_response(dose_response) npoints(160) delta(1) reg_type_t(quadratic) reg_type_gps(quadratic) interaction(1) bootstrap(yes) boot_reps(100) filename("output_wide") analysis(yes) graph("graph_output_wide") detail

sort httkv0
twoway lowess LastGFR httkv0
sort lnhttkv
twoway lowess LastGFR lnhttkv
*Stratified by age
use "C:\Users\Yaming\Desktop\Thesis\Yaming thesis dataset4.dta" ,clear
sum age, detail
keep if age<=44.7
```

```
egen p33 = pctile( httkv0 ), p(33)
egen p66 = pctile( httkv0 ), p(66)
gpscore gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason,t(httkv0) gpscore(gpscore) predict(y_hat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) detail
drop sd
doseresponse gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason, outcome(LastGFR)t(httkv0)gpscore(psore) predict(hat_treat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) dose_response(dose_response) npoints(160) delta(1) reg_type_t(quadratic) reg_type_gps(quadratic) interaction(1) bootstrap(yes) boot_reps(20) filename("output_wide") analysis(yes) graph("graph_output_wide") detail
use "C:\Users\Yaming\Desktop\Thesis\Yaming thesis dataset4.dta" ,clear
keep if age>44.7
egen p33 = pctile( httkv0 ), p(33)
egen p66 = pctile( httkv0 ), p(66)
gpscore gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason,t(httkv0) gpscore(gpscore) predict(y_hat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) detail
drop sd
doseresponse gender1 genetype1 trunc_grp1 race1 Springseason summerseason Fallseason winterseason, outcome(LastGFR)t(httkv0)gpscore(psore) predict(hat_treat) sigma(sd) cutpoints(cut) index(p50) nq_gps(5) t_transf(ln) dose_response(dose_response) npoints(160)
```

delta(1) reg_type_t(quadratic) reg_type_gps(quadratic) interaction(1) bootstrap(yes) boot_reps(20) filename("output_wide") analysis(yes) graph("graph_output_wide") detail

BIBLIOGRAPHY

1. Chapman AB. Approaches to testing new treatments in autosomal dominant polycystic kidney disease: insights from the CRISP and HALT-PKD studies. Clinical journal of the American Society of Nephrology : CJASN. 2008;3(4):1197-204.
2. Wuthrich RP, Serra AL, Kistler AD. Autosomal dominant polycystic kidney disease: new treatment options and how to test their efficacy. Kidney & blood pressure research. 2009;32(5):380-7.
3. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. The European Polycystic Kidney Disease Consortium. Cell. 1994;78(4):725.
4. Mochizuki T, Wu G, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, et al. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science. 1996;272(5266):1339-42.
5. Hateboer N, v Dijk MA, Bogdanova N, Coto E, Saggar-Malik AK, San Millan JL, et al. Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. Lancet. 1999;353(9147):103-7.
6. Ariza M, Alvarez V, Marin R, Aguado S, Lopez-Larrea C, Alvarez J, et al. A family with a milder form of adult dominant polycystic kidney disease not linked to the PKD1 (16p) or PKD2 (4q) genes. Journal of medical genetics. 1997;34(7):587-9.
7. Franz KA, Reubi FC. Rate of functional deterioration in polycystic kidney disease. Kidney international. 1983;23(3):526-9.
8. Fick-Brosnahan GM, Belz MM, McFann KK, Johnson AM, Schrier RW. Relationship between renal volume growth and renal function in autosomal dominant polycystic kidney disease: a longitudinal study. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2002;39(6):1127-34.
9. Chapman AB, Guay-Woodford LM, Grantham JJ, Torres VE, Bae KT, Baumgarten DA, et al. Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort. Kidney international. 2003;64(3):1035-45.

10. Grantham JJ, Torres VE, Chapman AB, Guay-Woodford LM, Bae KT, King BF, Jr., et al. Volume progression in polycystic kidney disease. The New England journal of medicine. 2006;354(20):2122-30.
11. Chapman AB, Bost JE, Torres VE, Guay-Woodford L, Bae KT, Landsittel D, et al. Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease. Clinical journal of the American Society of Nephrology : CJASN. 2012;7(3):479-86.
12. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annual review of public health. 2000;21:121-45.
13. Mantel N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association. 1963;58(303):690-700.
14. David HA, Jason L. Gunnink. The paired t test under artificial pairing. The American Statistician. 1997;51(1):9-12.
15. Harrell FE, Kerry L. Lee, Robert M. Califf, David B. Pryor, Robert A. Rosati. Regression modelling strategies for improved prognostic prediction. Statistics in medicine. 1984;3(2):143-52.
16. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology. 1996;49(12):1373-9.
17. Rosenbaum PR, Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983:41-55.
18. Li L, Kleinman K, Gillman MW. A comparison of confounding adjustment methods with an application to early life determinants of childhood obesity. Journal of developmental origins of health and disease. 2014;5(6):435-47.
19. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate behavioral research. 2011;46(3):399-424.
20. Imbens GW. The Role of the Propensity Score in Estimating Dose-Response Functions. Biometrika. 2000;87(3):706-10.
21. Jiang M, Foster EM, Gibson-Davis CM. Breastfeeding and the child cognitive outcomes: a propensity score matching approach. Maternal and child health journal. 2011;15(8):1296-307.
22. Hirano K, Imbens, G. W. The propensity score with continuous treatments. Applied Bayesian modeling and causal inference from incomplete-data perspectives. 2004;226164:73-84.
23. Brosnahan GM. Volume progression in polycystic kidney disease. The New England journal of medicine. 2006;355(7):733; author reply -4.
24. Rule AD, Torres VE, Chapman AB, Grantham JJ, Guay-Woodford LM, Bae KT, et al. Comparison of methods for determining renal function decline in early autosomal dominant polycystic

kidney disease: the consortium of radiologic imaging studies of polycystic kidney disease cohort. Journal of the American Society of Nephrology : JASN. 2006;17(3):854-62.

25. Gabow PA, Johnson AM, Kaehny WD, Kimberling WJ, Lezotte DC, Duley IT, et al. Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease. Kidney international. 1992;41(5):1311-9.

26. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. American journal of epidemiology. 2003;158(3):280-7.

27. Drake C. Effects of misspecification of the propensity score on estimators of treatment effect. Biometrics. 1993;49:1231–36.

28. Sedman A, Bell P, Manco-Johnson M, Schrier R, Warady BA, Heard EO, et al. Autosomal dominant polycystic kidney disease in childhood: a longitudinal study. Kidney international. 1987;31(4):1000-5.