PREDICTIVE MODELING USING A NATIONALLY REPRESENTATIVE DATABASE TO
IDENTIFY PATIENTS AT RISK OF DEVELOPING MICROALBUMINURIA

by

Lorenzo Andrés Villa Zapata

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF PHARMACEUTICAL SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2014

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Lorenzo Andrés Villa Zapata, titled “Predictive Modeling Using a Nationally Representative Database to Identify Patients at Risk of Developing Microalbuminuria” and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Terri Warholak PhD Date: (August 11, 2014)

Marion Slack PhD Date: (August 11, 2014)

Daniel Malone PhD Date: (August 11, 2014)

Anita Murcko MD Date: (August 11, 2014)

George Runger PhD Date: (August 11, 2014)

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Terri Warholak PhD Date: (August 11, 2014)

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Lorenzo Andrés Villa Zapata

ACKNOWLEDGEMENTS

There are many people whose help inspired me to pursue this path. To them, I owe all my gratitude. It is impossible to name them all; however, several people deserve a special mention because of their significant contribution.

Thank you to the members of my dissertation committee: Marion Slack, PhD; Daniel Malone, PhD; Anita Murcko, MD; and George Runger, PhD. Their advice and guidance has been invaluable and has helped me to have an open vision through this entire process. I want to particularly express my profound gratitude to my major advisor Terri Warholak, PhD. Thank you so much for your guidance and friendship along this entire journey; your words will be in my mind always.

To my family in Chile who sent me all their energy and good vibes. You were always there for me when I needed it. This is for you Loro, Mary, Pauly, Chocho, Pablo, Camila, Matias, Feña and Homero. We walked this adventure together and I am eternally grateful for your support. I love you all.

I would like to say thank you to all my friends in Tucson. You guys were my “Gringo-Chilean” family and I know our lives will cross again in the future—I can’t wait for that. Without your support and friendship, this would not have been possible. Thanks to Ilse, Daniela, Judith, Juan, Amanda, Queeny, Prasadini, Ana, Diana, Denis and so many others that have shared part of their life with me. I feel fortunate!

To my partner of life Michael, you have been a blessing for me. You know we are starting a new chapter together. Let’s discover and enjoy new experiences. Thank you for all your support, words and unconditional love. I love you.

Finally, I would like to thank your family as well, which is now my family too. Thank you dear Levengood, Basile, and Wetherington families for all your support, guidance, and jokes, and for accepting me into your wonderful family.

TABLE OF CONTENTS

LIST OF FIGURES.....	8
LIST OF TABLES.....	9
ABSTRACT.....	11
LIST OF ABBREVIATIONS.....	13
CHAPTER 1: INTRODUCTION.....	15
1.1 Background.....	17
1.1.1 Predictive modeling for general disease management.....	17
1.1.2 Microalbuminuria.....	19
1.2 Statement of the problem.....	20
1.3 Purpose of the study.....	23
1.4 Study objectives.....	25
1.4.1 Objective 1.....	25
1.4.2 Objective 2.....	27
1.4.3 Objective 3.....	29
1.5 Dissertation outline.....	30
CHAPTER 2: REVIEW OF THE LITERATURE.....	31
2.1 Prediction modeling.....	31
2.1.1 Definition.....	31
2.1.2 Model development.....	32
2.1.3 Validation of the model.....	36
2.1.4 Overview of predictive modeling processes and outcomes.....	38
2.2 Microalbuminuria.....	41
2.2.1 Definition, measurement and prevalence of microalbuminuria in the general population.....	41
2.2.2 Pathophysiology of microalbuminuria.....	43
2.2.3 Therapeutical interventions to control the presence of microalbuminuria.....	45
2.2.4 Microalbuminuria and predictive modeling.....	46
2.3 Use of predictive modeling in urology.....	50
2.4 NHANES, National Health and Nutrition Examination Survey.....	52
2.5 Theoretical Framework.....	53

TABLE OF CONTENTS (continued)

CHAPTER 3: METHODS.....	59
3.1 Study overview.....	59
3.2 Study design.....	60
3.3 Database.....	62
3.3.1 NHANES.....	62
3.4 Study outcomes.....	64
3.5 Candidate predictor.....	65
3.5.1 Variable definitions.....	66
3.6 Sample size considerations.....	67
3.7 Analysis: statistics and modeling process.....	69
3.7.1 Study of correlations.....	69
3.7.2 Variable transformation.....	70
3.7.3 Multivariate logistic regression and model construction.....	72
3.7.3.1 Variable selection method.....	72
3.7.3.2 Predicted probability.....	76
3.7.3.3 Odds ratio.....	77
3.7.3.4 Risk score calculation.....	78
3.7.4 Assessing the quality of the model, overall performance, discrimination and calibrations.....	80
3.7.4.1 Overall performance.....	81
3.7.4.2 Goodness of fit test.....	81
3.7.4.3 Receiver operator characteristic curve.....	82
3.7.4.4 Validation.....	84
3.7.5 Statistical package and test.....	85
CHAPTER 4: RESULTS.....	87
4.1 Descriptive statistics.....	87
4.1.1 NHANES dataset.....	87
4.2 Inferential statistics.....	94
4.2.1 Study of correlation and univariate analysis.....	94
4.2.1.1 Conclusions to hypotheses for objective one.....	97
4.2.2 Model development.....	98
4.2.3 Risk score.....	103
4.2.3.1 Conclusions to hypotheses for objective two.....	108
4.3 Model validation.....	110
4.3.1 Model diagnostic assessment.....	115
4.3.1.1 Goodness of fit test.....	115
4.3.1.2 Specification errors.....	116
4.3.1.3 Pseudo R squared.....	117
4.3.1.4 Collinearity.....	118

TABLE OF CONTENTS (continued)

CHAPTER 5: DISCUSSION.....	120
5.1 Prevalence of microalbuminuria.....	120
5.2 Correlation of predictors with microalbuminuria.....	120
5.2.1 Blood pressure.....	122
5.2.2 Sex.....	126
5.2.3 Age.....	128
5.2.4 Diabetes mellitus.....	129
5.2.5 Triglycerides.....	133
5.2.6 C-reactive protein.....	135
5.2.7 Blood urea nitrogen.....	136
5.2.8 Uric acid.....	138
5.2.9 Creatinine.....	141
5.2.10 Estimated glomerular filtration rate.....	143
5.2.11 Globulins.....	144
5.2.12 Ethanol.....	145
5.2.13 Lactate dehydrogenase.....	147
5.3 Predictive modeling and microalbuminuria.....	148
5.3.1 Variables included in the model A and B.....	153
5.3.1.1 Systolic blood pressure.....	153
5.3.1.2 Fasting glucose.....	155
5.3.1.3 HbA1c.....	157
5.3.1.4 Blood urea nitrogen.....	159
5.3.1.5 C-reactive protein.....	160
5.3.1.6 Smoking.....	161
5.3.1.7 Ethanol.....	164
5.4 Prediction of an individual patient risk.....	167
5.5 Comparing predictive model for microalbuminuria.....	170
5.6 Publication process.....	172
5.7 Summary of findings	172
5.8 Limitations.....	173
5.9 Lessons learned and future research.....	174
5.10 Conclusions.....	179
5.11 References.....	180

LIST OF FIGURES

Figure 2.1	Example of receiver operator curve (ROC).....	37
Figure 2.2	Rationale in the development of predictive model based upon clinical data.....	40
Figure 2.3	Microalbuminuria and its relationship with cardiovascular and renal diseases.....	44
Figure 2.4	Conceptual scheme for cause of a disease, sufficient-component cause model theory.....	55
Figure 3.1	Transformation options according with variable distribution.....	71
Figure 3.2	Area under the curve and its interpretation.....	84
Figure 4.1	Area under the curve. Model A.....	110
Figure 4.2	Area under the curve. Model B.....	111
Figure 4.3	Performance of model A in both datasets.....	112
Figure 4.4	Performance of model B in both datasets.....	114

LIST OF TABLES

Table 3.1	Sample size necessary to create a predictive model.....	68
Table 4.1	Demographic and clinical characteristics for the overall, development and validation datasets.....	88
Table 4.2	Albuminuria prevalence in included subjects: whole datasets.....	90
Table 4.3	Albuminuria prevalence in included subjects: model development dataset.....	90
Table 4.4	Albuminuria prevalence in included subjects: model validation dataset.....	90
Table 4.5	Predictor variables in the model development dataset.....	92
Table 4.6	Correlation and univariate logistic regression relationship of albuminuria and predictors in model development dataset.....	95
Table 4.7	Summary of hypothesis and conclusions for objective one.....	97
Table 4.8.1	Results of the multivariate logistic regression analysis exploring patient characteristics associated with microalbuminuria. Model A.....	100
Table 4.8.2	Results of the multivariate logistic regression analysis exploring patient characteristics associated with microalbuminuria. Model B.....	101
Table 4.9	Risk Score for model A.....	106
Table 4.10	Predicted probabilities for each score.....	107
Table 4.11.1	Summary of the hypothesis for objective two for model A and conclusions.....	108
Table 4.11.2	Summary of the hypothesis for objective two for model B and conclusions.....	109
Table 4.12	Comparisons between areas under the curve for model A in development and validation datasets.....	113
Table 4.13	Comparisons between areas under the curve for model B in development and validation datasets.....	114
Table 4.14	Goodness of fit test (Hosmer-Lemeshow) for model A and B.....	115

Table 4.15	Specification errors for model A and B.....	116
Table 4.16	Measures of fit, pseudo-R squared by model.....	117
Table 4.17	Study of collinearity in both models.....	118
Table 4.18.1	Summary of the hypothesis for objective three and conclusions for model A.....	119
Table 4.18.2	Summary of the hypothesis for objective three and conclusions for model B.....	119

ABSTRACT

Background: Predictive models have been used in other sciences for years but have only recently been explored in health care. By analyzing the conditions that may lead to a particular health outcome, predictive models offer the benefit of allowing clinicians to more accurately identify higher and lower risk patients and make more targeted treatment decisions, offering cost savings and improving efficiency in health systems. Microalbuminuria (MA) is a condition characterized by the presence of albumin in the urine below the threshold detectable by a standard dipstick (Albumin Urine Creatinine Ratio (AUCR) between 30 and 300 mg/g). Its presence involves the dysfunction of the glomerular filtration barrier and it is understood to be an early marker for cardiovascular disease. Therefore, identifying patients at risk for MA and intervening to treat or prevent variables associated with dysfunction of the glomerular filtration barrier, such as high blood pressure or high blood glucose, may support more cost-effective treatment for cardiovascular disease.

Methods: Data from the 2007-2008 National Health and Nutrition Examination Survey (NHANES) was utilized to create predictive models for MA. This cross-sectional database, representative of the U.S. population, includes clinical patient assessments, medical and laboratory tests, and self-reported data on numerous variables that may be associated with MA. To develop the model, the dataset was split into thirds; one-third was used to develop the model, while the other two-thirds were utilized to validate the model. Univariate logistic regression was initially performed to identify variables related with MA. Then, step-wise multivariate

logistic regression was performed to create the models. Model performance was evaluated using three criteria: 1) receiver operator characteristic (ROC) curves; 2) pseudo R-squared values; and 3) goodness of fit (Hosmer-Lemeshow). The predictive models were then used to develop risk-scores.

Results: Two models were developed using variables for which there were significant correlations in the univariate analysis ($p\text{-value} < 0.05$). For Model A variables included in the final multivariate model included; systolic blood pressure (SBP); fasting glucose; C-reactive protein; blood urea nitrogen (BUN); and alcohol consumption. For Model B the final multivariate model included: SBP; glycohemoglobin; BUN; smoking status; and alcohol consumption. Both models performed well in the creation dataset and no significant difference between the models was found when they were evaluated in the validation set. A 0-18 risk score was developed utilizing Model A (excluding alcohol consumption) and the predictive probability of developing MA for each score was calculated.

Conclusion: The predictive models developed in this dissertation provide new evidence about which variables are related with MA and may be used by clinicians to identify at-risk patients and to tailor treatment. Furthermore, the risk score developed using Model A may allow clinicians to more easily measure patient risk. Both predictive models have a reasonable discriminative capacity but will require external validation before they can be applied to other populations.

List of Abbreviations

AHRQ	Agency for Healthcare Research and Quality
AIC	Akaike Information Criterion
AIDS	Acquired Immunodeficiency Syndrome
AUC	Area Under the Curve
AUCR	Albumin Urine Creatinine Ratio
BMI	Body Mass Index
BUN	Blood Uric Nitrogen
BNP	B-type Natriuretic Peptide
DBP	Diastolic Blood Pressure
df	Degrees of Freedom
HbA1c	Glycated Hemoglobin
HIV	Human Immunodeficiency Virus
LDL-c	Low Density Lipoprotein Cholesterol
LR	Logistic Regression
MA	Microalbuminuria
MLE	Maximum Likelihood Estimates
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Survey
OR	Odds Ratio
Pr	Probability
PREVEND	Prevention of Renal and Vascular End-stage Disease

RCT	Randomized Controlled Trial
RSS	Residual Sum of Squares
SBP	Systolic Blood Pressure
SCC	Sufficient Component Cause model
T2DM	Type 2 Diabetes Mellitus
TB	Tuberculosis
UAE	Urinary Albumin Excretion
U.S.	United States

CHAPTER 1

1. INTRODUCTION.

The United States (U.S.) spends more on healthcare per capita than any other country in the world (Anderson and others 2005; OECD ; Squires 2012). In recent years, this spending has come under unprecedented scrutiny. (Moses and others 2013) Considering the intense discussion that surrounded the passage of the Affordable Care Act in 2010, the cost of health care has since become a focal point of debates concerning government expending, entitlement reform, and ways to reduce the federal deficit (Gruber 2011; Howard 2013). Healthcare resources are unfortunately finite; therefore, appropriate resource allocation is essential for any health system to be successful. This is especially relevant in the US considering that 5% of the population is responsible for the almost half of all healthcare expenditures (Stanton 2005).

As the cost of healthcare rises, health programs must face the challenge of running suitable health services while concurrently controlling expenses. One way to address this problem is to identify patients with special needs and enroll them in tailored care plans. This type of care management could lead better health outcomes or cost savings results, because it encourages preventive efforts that could improve health outcomes while decreasing the need for pricy care.

Compared to other Organization for Economic Co-operation and Development countries (OECD), the performance of the U.S. healthcare system is relatively poor, despite extraordinary expenditures and the economic success of many healthcare providers. Despite improvements in recent years, the U.S.

healthcare system has become trapped in an iron triangle of incompatible patient, clinician, and policy maker expectations. Patients' preferences for choice of physician or setting, for example, are often incompatible with insurers' cost-reduction goals. One approach to harmonize this situation may be to identify patient health status and healthcare utilization through forecasting using mathematical models. (Moses and others 2013)

Healthcare providers are under pressure to cut healthcare costs and improve patient outcomes, satisfaction and safety, which requires a more cost-effective use of resources. (Paper 2009) Under this scenario, predictive modeling could serve as a powerful tool to help healthcare providers achieve these goals. While they have been used for decades in other industries, predictive models are just now starting to be used in healthcare. (OCS 2008; Wharam and Weiner 2012) Predictive modeling has the potential to help healthcare providers and managers improve the efficiency and quality of care, thus reducing costs and improving patient outcomes. (Harrell 2001; Hodgman 2008) Predictive modeling generally works through a statistical process by which retrospective data are analyzed and used to create an algorithm or model that can determine the likelihood of a future event. In other words, predictive modeling helps identify the future occurrence of an outcome, based on an in-depth understanding and analysis of what has happened in the past using patient information already collected. (Steyerberg and others 2010) This approach could help healthcare systems to predict outcomes and provide appropriate patient-specific care. Also, predictive modeling is likely to be more accurate than the predictions of clinical experts because it can statistically account for bias,

accommodate multiple variables, and incorporate data simultaneously. Because of its accuracy when predicting clinical outcomes, modeling can be useful in patient level clinical decision-making as well as for boarder health policy decisions.

1.1 Background

1.1.1 Predictive modeling for general disease management

Chronic diseases are a principal source of disability, morbidity and mortality in the US and globally. It is predicted that by 2020, around 51% of people worldwide will suffer from a chronic disease, (Bodenheimer and others 2009) which will cause healthcare costs to increase. Since healthcare resources are finite and high-risk patients often consume expensive technology, the utilization of predictive modeling, by identifying patients at high-risk, could improve therapeutic results and focus resources on patients at highest risk from their medical conditions.

From this perspective, the idea of using predictive modeling is attractive because it could incorporate an algorithm utilizing patient history data to anticipate the occurrence of a certain outcome in the future, with the advantage of allowing also for actualization and adaptation over time. The model produced can help to improve staff efficiency, reduce administrative utilization and increase patient focus. (Michael S. Cousins 2004) However, models, like people, are complex, and often the inclusion of too many variables or characteristics makes it difficult to identify a single source or outcome.

A model is an abstraction of the real world that attempts to capture complex human behavior in simple mathematical and statistical terms. (Michael S. Cousins

2004) The more complex a model, generally the more accurately it can predict outcomes. Complex predictive models can be developed using any number of statistical approaches, perhaps the most common of which are decision trees and logistic regression. The advantage of these approaches is that they can deal with the complexity of different scenarios by taking into consideration multiple factors or characteristics when calculating anticipated risk or the likelihood of an event.

(Harrell 2001; Hodgman 2008) There are several reasons why predictive modeling may be desirable over predictions made by physicians; a model can screen an entire population on a repeated basis and is more accurate than clinicians because it avoids cognitive bias and includes patient information from any part of the healthcare system. (Lewis 2011)

Currently, clinical applications of predictive modeling in healthcare are not common; however, they are beginning to gain the attention of healthcare policy makers and clinicians. Therefore, a comprehensive exploration using a dataset that contains clinical and demographic patient information could calculate the probability of an outcome on a patient or population level. Predictive modeling can also help clinicians, policy makers, and researchers more efficiently and thoroughly understand patients and their risks. (Actuaries 2010; Sernyak and Rosenheck 2003) This understanding could be translated into more successful allocation of resources, implementation of best practices, and a focus on the patients most in need. Predictive modeling in urology is becoming increasingly popular not only for academic and scientific purposes but also to address budget constraints, making its incorporation into clinical practice a distinct possibility. (Cestari 2013)

1.1.2 Microalbuminuria

Microalbuminuria (MA) is the presence of albumin above the normal level but below the detection limit of simple assays. Some epidemiological investigations in the general population have set values of microalbuminuria between 30 to 300 mg/g. (Khosla and others 2006) Microalbuminuria is a consequence of altered glomerular permeability, which could be the result of pathophysiological events like hypertension or diabetes mellitus. (Bakris 1996; Scheven and others 2013c; Yudkin and others 1988) Recent studies have shown that both microalbuminuria and isolated microalbuminuria (microalbuminuria occurring once or twice) could be associated with an increased risk of cardiovascular events and mortality, but further research is needed. (Gumbinger and others 2012; Hillege and others 2001) In addition to albumin, recently a urinary albumin/creatinine ratio has been used to measure renal function; however, it is necessary to use caution when interpreting this ratio because it could be affected by biological factors such as an increase in salt and protein intake, as well as muscle mass, race, and gender. Patients with large muscle mass, males, and African Americans, for example, have higher levels of creatinine excretion. (Khosla and others 2006; Mattix and others 2002b)

Variation in prevalence of microalbuminuria also depends on the presence of high blood pressure, race, age or renal disease. Some studies estimate a prevalence of microalbuminuria of 12% to 36% in type 2 diabetes patients and of 5% to 40% in non-diabetic patients. (Khosla and others 2006) Clinical implications of the presence of microalbuminuria are related with its role as a marker of organ damage, such as left ventricular hypertrophy and atherosclerosis, as well as its prognostic

implication in the line of hypertension organ damage and diabetes complications. (Echouffo-Tcheugui and Kengne 2012) Even though it has been recently available more evidence about the link between microalbuminuria and the occurrence of the diseases described above is needed, there is no evidence about which patient variables, from demographics to other comorbidities, are related with the presence of microalbuminuria. Furthermore, normalizing or reducing its level could be directly or indirectly related with reducing the occurrence of chronic diseases. The utilization of predictive modeling under this premise emerges as an opportunity to identify risk variables closely related with microalbuminuria appearance. Furthermore, a predictive model would allow for the identification of patients with a high probability of suffering microalbuminuria. These patients, in turn, could be placed under a preventive therapeutical path.

1.2 Statement of the problem.

Healthcare systems worldwide are moving toward a clinical integrative model of care. This new scenario will require accessible health data from providers, patients, and other participants involved directly in healthcare. Scientists and researchers must have the ability to transform available data into useful information for a diversity of purposes ranging from improving patient outcomes to offering public health interventions that impact population health. In order to be effective, a predictive model must consider not only the design of the model but also a subsequent intervention that has to be driven by the disease, patients, and

physicians. Also, prediction modeling should consider clinical priorities and measurable events, as well as clinical protocols and patient outcomes.

Predictive modeling has been gaining more attention recently for a number of reasons. The quantity of health information is increasing due to the creation of and access to complex databases and electronic health records, better understanding of statistical predictors of health, and increased utilization of multifactorial models. Additionally, these models are capable of producing risk profiles for different types of patients and of modeling future health care resources consumption. This technique could use retrospective data available from patient profiles to predict which members of a certain population have a high probability of experiencing outcomes of interest (e.g., declines in health or consumption of health services in the future). This predictive information could be valuable for providers, clinicians, and patients because it may allow the targeting of those at highest risk for an event.

Microalbuminuria is a condition linked with cardiovascular risk especially in patients diagnosed with diabetes mellitus or high blood pressure. (Jarraya and others 2013; Khosla and others 2006) Even though enough proof has been collected to support the role of microalbuminuria as a predictor of cardiovascular events in high-risk populations, there is no local evidence until now linking patient characteristics with the occurrence of MA. Screening on a regular basis for microalbuminuria could help to identify a subgroup of patients with MA risk factors; however, laboratory tests are not able to identify values of protein classified as microalbuminuria. Developing a model that could identify patients at risk will allow

a more intensive therapy or a closer follow-up because they could benefit from early MA intervention and treatment.

Furthermore, using predictive modeling to identify patients at risk of microalbuminuria will accelerate potential benefits, including the introduction of innovative patient care strategies that target at risk populations so that proactive interventions and more real time recommendations for individuals can be made based on data about their health. Also, results of predictive modeling will allow for a feedback loop that aggregates additional evidence regarding the real world impact of forecasts on clinical decisions. While some may think it is too soon to determine whether predictive modeling will improve patient outcomes and quality, recent initiatives from Agency for Healthcare Research and Quality's (AHRQ) are trying to translate research into practice, thus demonstrating the potential for predictive models to be used in healthcare in conjunction with education and direct patient care.

The results generated in this research effort could be used by physicians to guide clinical care and target specific MA treatment paths; providers to better plan for the allocation of MA resources; and people who wish to stay healthy and to anticipate the occurrence of MA.

1.3 Purpose of the Study

Predictive modeling is a thriving field of study. In the U.S., as well as other countries, there has been a significant increase in the amount of collected data that are waiting to be processed. New techniques to process large volumes of data repositories and terabits are now the norm. (Duncan 2011) Healthcare systems, on the other hand, are seeking computational abilities that allow for the provision of higher quality care to at risk patient populations. Data generated within a health intervention are collected, compiled and protected for the generation of new knowledge. In 2009, the Health Information Technology for Economic and Clinical Health Act (HITECH) was passed and introduced new challenges in this big data era. (Goldstein and Thorpe Jane 2010) Despite this new legislation, a lack of standardized data across settings and systems, inadequate system infrastructure, and the cost of accessing information are clear barriers for healthcare systems. Nevertheless, there has been an increase in exploration of this new field to produce predictive algorithms that can accurately predict future behavior or outcome patterns. The consequences for clinical and economic decision-making could be immeasurable. Integrating clinical and administrative data will provide a better understanding of disease evolution and anticipation of trends.

Protein in urine is not normal, even in small quantities. Investigators have found that patients with microalbuminuria – protein at a level too low to be detected with standards laboratory tests – are at higher risk not only for kidney but also for cardiovascular disease. (Bakris 1996; Chen and others 2012; Folsom 2013) In people with controlled hypertension, whether microalbuminuria is a good

forecaster of cardiovascular disorder remains to be determined. To date, no evidence has identified patient variables related with the occurrence of microalbuminuria in the United States.

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults in the United States. This survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country.

The survey is unique because it combines interviews and physical examinations. The interviews include demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests.

Findings from this survey have been used to determine the prevalence of major diseases and risk factors for diseases. NHANES results are also the basis for national standards for such measurements as height, weight, and blood pressure. Data from this survey has been used in health sciences research to develop public health policy, design health programs and services, and expand health knowledge for the nation.

This investigation aims to use data from patients belonging to the NHANES dataset to develop a predictive model able to identify patients at risk of developing MA. The models created were validated in the same dataset.

1.4 Study Objectives

The following specific objectives were evaluated through the utilization of NHANES national dataset to develop and test a model capable of identifying and predicting the occurrence of microalbuminuria in patients belonging to this database.

1.4.1 Objective 1

To identify a temporal variable relationship, using health record data stored in NHANES national database in patients at risk of having microalbuminuria.

To identify the strength (weak, medium, strong) and direction (positive, negative) of patient variables included in the NHANES database and the occurrence of microalbuminuria through correlation and univariate association.

Specific null Hypotheses are as follows:

Hypothesis 1.1: Systolic blood pressure is not correlated with microalbuminuria.

Hypothesis 1.2: Diastolic blood pressure is not correlated with microalbuminuria.

Hypothesis 1.3: Glycosylated hemoglobin is not correlated with microalbuminuria.

Hypothesis 1.4: Age is not correlated with microalbuminuria.

Hypothesis 1.5: Male sex is not correlated with microalbuminuria.

Hypothesis 1.6: Type 2 diabetes mellitus is not correlated with microalbuminuria.

Hypothesis 1.7: Lack of health care access is not correlated with microalbuminuria.

Hypothesis 1.8: Race is not correlated with microalbuminuria.

Hypothesis 1.9: Body mass index is not correlated with microalbuminuria.

Hypothesis 1.10: Total cholesterol level is not correlated with microalbuminuria.

Hypothesis 1.11: Triglyceride level is not correlated with microalbuminuria.

Hypothesis 1.12: Fasting glucose level is not correlated with microalbuminuria.

Hypothesis 1.13: Smoking is not correlated with microalbuminuria.

Hypothesis 1.14: Alcohol consumption is not correlated with microalbuminuria.

Hypothesis 1.15: C-reactive protein is not correlated with microalbuminuria.

Hypothesis 1.16: Estimated glomerular filtration rate (eGFR) is not correlated with microalbuminuria.

Hypothesis 1.17: Systolic blood pressure, HbA1c, age and male sex are not the strongest predictors of microalbuminuria.

Rationale: A reliable predictive factor or variable must have a sensitivity adequate to detect those patients who are destined to have the outcome of interest and a specificity adequate to correctly identify those patients who will not have the outcome of interest. Although a variable labeled as a risk factor could be identified using the relative risk (or odds ratio) of an outcome, a variable has a desirable predictive power based not only on the magnitude of that association but also on the

prevalence of both the variable and the outcome. Some variables have shown a strong correlation by themselves, however in a model these relationships could no longer be true.

1.4.2 Objective 2

To develop a patient data driven predictive model and a risk score assessment to improve the identification of MA and MA risk.

The following null hypotheses are: when included in the predictive model, we hypothesize that:

Hypothesis 2.1: Systolic blood pressure does not predict the occurrence of microalbuminuria.

Hypothesis 2.2: Diastolic blood pressure does not predict the occurrence of microalbuminuria.

Hypothesis 2.3: Glycosylated hemoglobin does not predict the occurrence of microalbuminuria.

Hypothesis 2.4: Age does not predict the occurrence of microalbuminuria.

Hypothesis 2.5: Male sex does not predict the occurrence of microalbuminuria.

Hypothesis 2.6: Type 2 diabetes does not predict the occurrence of microalbuminuria.

Hypothesis 2.7: Lack of health care access does not predict the occurrence of microalbuminuria.

Hypothesis 2.8: Race does not predict the occurrence of microalbuminuria.

- Hypothesis 2.9: Body mass index does not predict the occurrence of microalbuminuria.
- Hypothesis 2.10: Total cholesterol level does not predict the occurrence of microalbuminuria.
- Hypothesis 2.11: Triglyceride level does not predict the occurrence of microalbuminuria.
- Hypothesis 2.12: Fasting glucose level does not predict the occurrence of microalbuminuria.
- Hypothesis 2.13: Smoking does not predict the occurrence of microalbuminuria.
- Hypothesis 2.14: Alcohol consumption does not predict the occurrence of microalbuminuria.
- Hypothesis 2.15: C-reactive protein does not predict the occurrence of microalbuminuria.
- Hypothesis 2.16: Estimated glomerular filtration rate does not predict the occurrence of microalbuminuria.
- Hypothesis 2.17: A model consisting of SBP, HbA1c, age and male sex does not predict the development of MA.
- Rationale: Due to recent advances in both information and technology, administrative data are becoming a powerful source of information and offer the potential to improve quality of care while limiting expenditures. The purpose of this study is to identify MA in patients before the actual occurrence of that outcome using a predictive model that incorporates national real life data, including a wide

range of tools to stratify a population according with its risk of developing or presenting an outcome.

1.4.3 Objective 3

To validate the performance of the predictive model created to identify patients at risk of microalbuminuria.

Hypothesis 3.1: An overall measure of model accuracy using Hosmer-Lemeshow test does not show a good model fit.

Hypothesis 3.2: The predictive power of the model assessed through receiving operating characteristic curves (ROC) does not show a higher area under the curve (AUC), or a good discriminative performance of the model.

Hypothesis 3.3: ROC curves generated from the model development and model validating portions of the NHANES are not significantly different.

Rationale: There is a lack of consistency in the literature regarding the performance of predictive models. Though statistical characteristics such as discrimination and calibration are extensively suggested to evaluate, however in reality, calibration is hardly assessed. (Collins and Moons 2012; Steyerberg and others 2010) To validate the statistical characteristics of the model, it is necessary to test it using new individuals who were not considered in the model development process.

1.5 Dissertation Outline

This dissertation is organized as follows. Chapter 2 presents background for this research, introduces information on microalbuminuria, utilization of logistic regression in predictive modeling, and predictive modeling itself as an approach to generate new evidence and various related concepts. Chapter 3 presents the methodology utilized in the development of this investigation including a description of the dataset used to find the predictors of microalbuminuria in a regional population. Chapter 4 presents results as descriptive and analytical statistical approaches used to describe the dataset, the sub dataset created to develop the model, the proposed model, the selection procedure and the construction of the predictive model and its validation. It discusses the patients' demographics and microalbuminuria's predictors. It also describes the results of the model validation process and risk score creation. Chapter 5 describes the final models and its utilization and discusses future improvements that can be done, summarizing the main findings of this investigation and concludes the dissertation.

CHAPTER 2

LITERATURE REVIEW

2.1 Prediction Modeling.

2.1.1 Definition.

Predictive modeling uses data available in patient profiles to predict which member of a population has a certain (in this case high) chance of experiencing losses or becoming a more frequent user of healthcare facilities in the future due to the occurrence of a certain disease. (Duncan 2011; Hendriksen and others 2013) In clinical practice, physicians incorporate clinical history, examination data, laboratory test results and disease state knowledge. In essence, prediction models simulate, but do not replace medical diagnostics, and for each combination of variables, models provide classification of patients according to their calculated risk. Prediction models can be used as a new instrument to guide medical decision-making regarding the necessity of new tests, assessments and treatments for specific patients.

The goal of our research is to identify whether people with a high probability of developing microalbuminuria according to their patient characteristics could be treated before that MA develops, resulting in cost savings and improved health outcomes. This predictive modeling can be used to answer the clinical question of how to best tailor healthcare. To be useful for providers, a predictive model needs to have transparency in development (i.e. evidence should be provided to support clinical necessity, variables included, clinical outcomes, and modeling process), have validity evidence (external and internal) and documented impact of implementation.

(Grobman and Stamilio 2006; Royston and others 2009) Moreover, the prediction of a specific outcome allows patients to make informed decisions when selecting from a set of treatment alternatives.

2.1.2 Model Development

An initial step in the development of a predictive model is to identify a source of data that may be capable of forecasting an outcome, and allocate a comparative weight for each variable. It is assumed that the data are sufficiently accurate to determine the outcome of interest. The individuals who provide these variables generally have the same disease, diagnosis and risk of the same health outcome. The cohort used to obtain the information is preferably prospective, but retrospective cohort databases with already collected data were employed in this research.

Data used for predictive models are usually gathered for other purposes, such as routine hospital records or surveys to get an instant picture of health population status and are usually those less systematic sources of data that may include several years of evidence. (Grobman and Stamilio 2006; Hendriksen and others 2013; Moons and others 2012b) The data included should be comparable among sites included, and if there is uncertainty regarding a variable, it is better to reconsider its inclusion because instead of contributing to the model, it may dilute model forecasting ability. If there is missing data for predictors, which is inevitable, it is necessary to identify the pattern and decide whether it is necessary to remove that person in order to preserve power and avoid inaccuracy (Vergouwe and others 2005). It may also be necessary to impute the data if the missing data follows a

random pattern. (Gorelick 2006; Wood and others 2008) However, any predictor with a large amount of missing data suggests that that particular variable is hard to measure in current or real-life conditions or could have a significant amount of error associated with it, therefore making it prudent to re-evaluate whether the variable should be included in the model. (Hendriksen and others 2013; Royston and others 2009) Another important consideration is that some variables, particularly continuous ones, may require some type of transformation (e.g., square root, logarithm, quadratic, restricted cubic splines) since linearity cannot automatically be assumed. Furthermore, to avoid incorrect interpretation or incorrect predictions, transformation must be tested to explore non-linearity. (Moons and others 2012b)

Suitable outcomes must be relevant to the healthcare system, but the emphasis is clearly on patient needs. Outcomes vary from situation-to-situation and range symptom occurrence, disease remission and disease complication to death. A concise outcome is evaluated to avoid bias, meaning that outcomes must be measured without knowledge of the predictors under study.

Predictors are variables that have the power to predict or contribute to the prediction or the occurrence of the outcome of interest and can be used to test the model's performance. Most of the time, the development of a predictive model involves the inclusion of several predictor variables in a multivariate analysis; however, not all of the predictor variables will be included in the final model. Predictor variables are mostly from the following categories: demographics, clinical history, comorbidities, clinical test results, and treatments. In addition to these

categories, any variable that may be a predictor should be included; however, its association with the outcome may or may not be causal. (Moons and others 2009) Some researchers suggest that for reliable predictive modeling, it is necessary to include ten patients with the event of interest per variable in the final model; often, it may be necessary to combine or exclude some predictors that could be highly correlated. (Peduzzi and others 1996) All the predictors included in the model need to be clearly defined and measured in a reproducible manner to improve model generalizability. For example, a model that includes a biomarker that is not widely available may have limited utility.

Despite these concerns, there is no consensus on how to evaluate the quality of a predictor, and most often researchers use their knowledge and judgment. When possible, predictors must be standardized among settings that contributed data, and variables with a high incidence of measurement error or inter-observer variability might be less appropriate to include in the final model because they may wrongly influence the prediction itself. (Moons and others 2012b)

There is no one best method to obtain a final model, but the literature present two main model types: (1) the full model and (2) the predictor selection approach. When the full model is chosen, all the variables from the database are used in the final model. Those in favor of this approach argue that this prevents selection bias; however, this approach requires prior knowledge about all the variables possible being related with the outcome and able of being included. A downside to this approach is that a model with many variables is often difficult to replicate in many settings.

In the predictor selection approach, those variables that are hard to assess in common practice or that do not contribute to the model must be either removed from the final model or not considered. Backward elimination is one of the preferred approaches of accomplishing this because at the beginning, all the variables are included, and subsequent tests are run to remove variables or predictors according to a predefined significance level. Forward selection, the other approach, is less utilized because it cannot account for all of the variables and their effects. A significance level of p-value less than 0.05 or 0.01 in general generates a model with few variables missing. However, setting significance levels above these values will produce a model that includes variables without a clear predictive power.

Researchers also recommend not excluding predictors for multivariable analyses on the sole consideration that a predictor is not related with the outcome from a statistical consideration because this lack of relation could be the result of a poor sample size. When a multivariate approach is used, each predictor included in the final model is reciprocally corrected for the other predictor also included in the model. The coefficients from these variables provide the occurrence or the probability of experiencing the outcome. The interpretation of these coefficients is one unit of increment in the level of a predictor on the estimated outcome when all other predictors in the model are held constant. (Harrell 2001; Moons and others 2012b) The final model must include all the variables considered to be associated with the targeted outcome. Then the probability can be calculated using the coefficient of the predictor. This probability should be easy to implement and

interpret using the original equation to create, for example, a web based calculator or nomogram to calculate individual probability of developing the outcome.

2.1.3 Validation of the Model

Independent of the approach chosen, internal and external validation of the final model is recommended to increase understanding of the extent that the model contains important variables, is well fitted, or predictable. (Harrell 2001; Steyerberg 2009)

Discrimination and calibration is the next step that every predictive model must consider. Calibration is the agreement between the likelihood of developing the outcome in certain period of time and the observed outcome frequencies—in other words, to distinguish disease from non-disease correctly with probabilities that are in line with the current outcome frequencies. One way to assess discrimination is to use the area under the receiver operator curve (ROC). (Grobman and Stamilio 2006) In this curve, there is a representation of the chance that of two individuals, one with and one without the outcome, the predicted outcome probability must be higher for the individual with the outcome rather than the individual without the outcome. The higher the area under the curve (AUC), the better the discriminative performance of the model. Figure 2.1 showed an example of receiver operator (ROC) curve.

Figure 2.1 Example of receiver operator curve (ROC)

(Grobman and Stamilio 2006)

Calibration is also an important in the process of developing a predictive model. One popular approach is to consider the Hosmer and Lemeshow test, which examines the “goodness-of-fit” of the model, by checking the lack of fit of the model using logistic regression. This test partitions the observations into equal size groups according with their predictive probabilities and uses Chi-square and p-values of 0.05 or less to indicate a lack of fit. Higher values indicate that the model fits the data well. (Duncan 2011) Independent of the technique selected to validate the final model, all techniques may produce an over-fitted—or overly optimistic—model. As a consequence, it is essential to check the final model to provide the most stable and parsimonious model version. The most common approach is to randomly split the data and use one portion to develop the model itself and the second portion to validate it. Recently, a more advanced method to skip this process is to use

bootstrapping in order to simulate the people from the source population, resulting in the creation of several models. This exercise can be repeated several times.

This technique allows for a comparison of new models and their coefficients generated and for an evaluation of the amount of over-fitting of the original model. (Steyerberg 2009) Once the model has shown an internal validation, it is necessary to test whether it will perform well for a different group of patients. This can be accomplished by running an external validation or cross-validation. This does not represent the development of a new model or refitting of the developed model; rather, model validation takes the final model with all its predictors and applies this to new individuals, thus quantifying the model performance. (Carlin and Xia 1999; Moons and others 2012a) The individuals included in this process may be from the same institution or chosen by simple non-random sampling from the original dataset. When a lower prediction is found, then the model is rejected and a new model is proposed or fitted, most of the time by repeating the selection of predictors. A better approach is to update the “old model” to a particular setting because new information given by new patients will be added.

2.1.4 Overview of Predictive Modeling Processes and Outcomes

A model is a generalization of reality that attempts to capture a certain phenomena that occurs in real life situations. It tries to capture features of human behavior and explain them in a mathematical or statistical way. A flaw, however, is that no matter how extensive or exhaustive the process to develop a model, models will never capture the whole situation they seek to explain. Nevertheless, models

are able to capture the main relationship and predict certain outcomes in a variety of situations and environments. Each member of the population used to create a model provides independent variables (e.g., demographics); then, each independent variable generates a value for the outcome (i.e., dependent variable), which corresponds to the likelihood of the event occurring. Once the model is built, tested, and validated, it can be used to generate a prediction score for the dependent variable.

Data mining is the procedure used to identify and isolate patterns of data from a dataset and is the step taken before producing the predictive model itself. Applications of models will result in values that are in reality projections of the dependent variable. The analysis of these values plus actual values will provide insight into the ability of the model to identify appropriate independent variables and accurately predict outcomes. The appropriate model will depend on the nature of the data available and the scale in which the variables used are presented (e.g. categorical, continuous). Logistic regression, for example, will work for both types of scales.

Another important factor is the structure of the dataset. The dataset could contain, for example, correlated variables because of the nature of the clinical process in which more specific and complicated approaches are necessary to produce reliable results. Choosing the right analytical approach will depend on the objectives of the research, whether it be exploring the relationship between several predictors and the outcome or predicting the occurrence of certain outcome using a

statistical model. (Duncan 2011) Even though these two approaches are related, depending on the situation one is going to be more relevant than the other.

Figure 2.2: Rationale in the development of a predictive model based upon clinical data

Predictive modeling is a powerful tool capable of identifying patients who can benefit significantly from an aggressive treatment path or intervention that can be applied in a healthcare setting. Also, the probability calculated for each patient can be used to determine if the patient is at increased likelihood of seeking care, which could help to project costs of care. Current evidence suggests that in order for predictive modeling to be successful, there must exist a net that provides frequent interventions, education and behavior modification approaches, as this tool only determines the chance that certain outcomes will happen but does not provide coordinate care or facilitate appointment or provide education or any intervention that the identified patient may require. (Hodgman 2008) It is also important to mention that some researchers oppose the utilization of this approach because it has not produced measurable improvements in quality of life. However, this may be because research has not yet measured the approach's impact on quality of life. Also, some argue that this technique could lead health plans and pharmaceutical

companies to choose to market specific interventions and allow employer groups to set individual premiums depending on the health risk. (Nass and others 2009)

2.2 Microalbuminuria.

2.2.1 Definition, measurement and prevalence of microalbuminuria in the general population

According to the National Kidney Foundation, microalbuminuria (MA) is the presence of albumin between 30 mg/day and 299 mg/day. Values below this range are considered normal, and values above this range are considered to be macro-albuminuria, which is indicative of serious renal disease. (Khosla and others 2006) MA is frequently referred to as a subclinical increase in urinary albumin excretion; however, nowadays it is linked as a predictor of cardiovascular disease, diabetic nephropathy and endothelial dysfunction. (Khosla and others 2006; Scheven and others 2013c) In fact, MA is now considered more frequently to be a risk marker rather than risk factor for the diseases mentioned above, though more investigation is required to find a causal relationship. (Jarraya and others 2013) The value of isolated MA without cardiovascular history or diabetes in the prediction of developing a related disease remains unclear. (Konno and others 2013; Scheven and others 2013b)

Currently, it is possible to measure MA using sensitive techniques like enzyme-linked assays or high-performance liquid chromatography; however, it is also important to consider the urine collection technique because ideally measurement of MA can be performed in the first urine void taken in the morning or

a random sample during the day. Although the 24-hour urine collection sample is still considered as a gold standard, the new techniques mentioned above can be an accurate measurement for MA in a random urine sample. MA could be influenced by variation in urine volume and is likely to be found in a concentrated urine sample. (Jarraya and others 2013; Khosla and others 2006) For this reason, it is better to utilize the albumin-creatinine ratio to estimate MA. (Gansevoort and others 2005; Tagle and others 2012)

The prevalence of MA could vary depending on ethnic groups, method of specimen collection, analytical method, gender, age, and the presence of high blood pressure, diabetic mellitus, or renal disease. (Metcalf and Scragg 1994) However, recent reports reveal prevalence in the general population between 5 to 20%. (Konno and others 2013) A pioneering study based on the NHANES III (National Health and Nutrition Examination Survey) showed that the MA prevalence in the U.S. population depends on several factors; for instance, among patients without diabetes or hypertension, MA prevalence is around 5%. Furthermore, prevalence was higher in children than young adults and increased constantly after 40 years of age. MA was greater in non-Hispanic blacks and Mexican-Americans compared with similar-aged non-Hispanic whites. MA prevalence was 29% in the diabetic population and 16% in patients with hypertension, and in the same study, age, sex, race/ethnicity, and concomitant disease contributed to the variability of MA prevalence estimates. (Jones and others 2002b)

2.2.2 Pathophysiology of microalbuminuria

Nowadays, MA is considered a marker for the presence of other diseases rather than a pathogenic factor, even though recent studies have shown a certain relationship between its presence and the occurrence of disease. (Khosla and others 2006) MA is now being associated with the classical and emerging protein risk factor, linking its occurrence and the presence of atherosclerosis associated with capillary impairment. (Jarraya and others 2013) The presence of albumin in urine is by itself a predictor of cardiovascular disease, and for this reason, the measurement of MA is becoming an important risk factor for cardiovascular disease not only by common risk factor but also by a common pathophysiologic process that is essential to control. (Diercks and others 2002) Endothelial dysfunction, through deficient endothelial nitric oxide synthesis, plays an important role in the initiation and progression of atherosclerosis. Accordingly, MA could explain why its presence in urine explains not only renal disease but also cardiovascular disease. The PREVEND study (Smink and others 2012) showed that MA is a strong predictor of cardiovascular mortality even independent of the presence of other risk factors. This finding was confirmed by the Framingham Heart Study, which added that MA provides an increased risk for cardiovascular disease in normotensive patients without diabetes and renal impairment. (Arnlov and others 2005) Other reports have established a link between the presence of MA and the occurrence of stroke; however, this outcome could be influenced by other factors, such as the presence of atrial fibrillation, diabetes and hypertension in those patients studied. Despite these limitations, the results provide the opportunity to study the likelihood that MA and

stroke may share a related pathophysiologic process. (Cho and others 2012; Elyas and others 2013; Gumbinger and others 2012) MA in patients with diabetes mellitus may also contribute to a more activated immune system, as diabetes leads to endothelial inflammation related with immune system activation plus the glycated state in which albumin is present. (Khosla and others 2006) The injured glomerular membrane subsequently allows the passage of albumin and others molecules, explaining the presence of MA. MA is no longer considered to be a specific risk marker, and some propose the creation of a new era in medicine in order to control MA in hypertensive patients with or without diabetes to detect subclinical organ damage.

Figure 2.3 Cardiovascular and renal disease in patients with a variety of disorders related with insulin resistance and/or glucose intolerance.

MA has been found as a marker of these conditions and disease activity (e.g., diabetes mellitus). (Lane 2004)

2.2.3 Therapeutical interventions to control the presence of microalbuminuria

Controlling patients with comorbidities already identified with a high correlation of having MA is undeniable. If we decide to treat or control the presence of MA indirectly, we will contribute to a reduction in the stage or the progression of related diseases. Under this principle, diet considerations, such as employing low-protein diets and better glycemic control, may be used to preserve renal function in the diabetic population and prevent the progression to initial renal dysfunction. Controlling high blood pressure is by far the most important approach to reducing urinary excretion of albumin. Reduction in MA must be a treatment goal in addition to blood pressure control. For this reason, the presence of high blood pressure in many patients is a main factor to consider, and the concomitant use of antihypertensive agents, such as an angiotensin- converting enzyme and an angiotensin receptor blocker known to have good results controlling MA, is now accepted as a good therapeutic approach because it can reduce the intraglomerular pressure and prevent the development of atherosclerosis. Some investigations have shown that these agents have a direct antiproteinuric effect rather than prevent MA occurrence by reducing blood pressure.(Agruss and others 2007; de Jong and Curhan 2006)

2.2.4 Microalbuminuria and predictive modeling

The use of predictive modeling arises as a very attractive tool that on a large scale could allow for the identification of segments of the population that would benefit from the identification and treatment for a particular outcome. It is crucial that every model developed to predict some health outcome deliver precise realistic estimates with no methodological inconsistencies.

MA is accepted as an early risk factor/predictor of renal and cardiovascular disease. Methods to measure MA levels are easy and inexpensive to run and are available in most primary care units. Advancement in this area has been connected by a corresponding detection of a strong, independent association between the presence of MA and several risks.

In recent years, several researchers have shown the link between the presence of MA and cardiovascular or metabolic diseases, positioning MA as a predictor in the occurrence of those diseases. Nevertheless, it is necessary to be cautious and understand that the interpretation of MA requires knowledge of several factors, such as the onset of MA in the urine and the variability of its presence in the urine sample. Vigorous exercise, urinary tract infection, heart failure, and also circadian rhythm can produce a transitory increase in the albumin. Taking these factors into consideration, MA assessment should be performed using two first morning urine samples, with a short period of time between sampling. Recent influential studies have supported microalbuminuria as a predictor. Morbidity and mortality studies have confirmed the prognostic value of the changes in urinary albumin excretion in diabetes mellitus, hypertension and cardiovascular disease.

(Abougambou and Abougambou 2013; Khan and others 2013; Olsen and others 2011; Redon and Martinez 2012; Schmieder and others 2011)

Existing evidence indicates that MA could be utilized as a predictor for some diseases; the question is to investigate what happens in the previous step. Three studies have conducted an assessment of the predictive value of a combination of the variables and risk factors for microalbuminuria. One mathematical model of the overweight Chinese male population used logistic regression to find that body mass index (BMI), blood pressure, blood uric acid, and fasting plasma glucose were predictors of the occurrence of MA; however, several limitations were present in this study. For example, the study was only conducted within Chinese obese men and no females or patients of other races were included, and there was no follow up to see if some of the variables identified as a predictors changed over time to determine whether MA remained constant. The model was also validated using only a cross-sectional data. (Chen and others 2011)

Another study examined circulating biomarkers representing several biological pathways were associated with MA over time in the community. Researchers fixed their attention on C-reactive protein, aldosterone, renin, B-type natriuretic peptide (BNP), plasminogen-activator inhibitor type 1, and homocysteine. They linked aldosterone, BNP, and homocysteine with incidence of MA. An implication of these findings is that targeting investigation in BNP and aldosterone pathways and their molecular mechanisms could reduce the occurrence of MA. Other limitations were also present: the sample of patients only included white people, and patients excluded had more risk factors at baseline, which may

have influenced findings. (Fox and others 2010) Clinical trials and cost effectiveness assessments are needed before considering the inclusion of these biomarkers in clinical settings.

The PREVEND study (Scheven and others 2013a) was a community prospective cohort study with a median follow-up of 9 years. With almost 6,000 patients included in the final analysis, patients at baseline did not have albuminuria and renal disease. The multivariate logistic regression included demographic variables (i.e., age, gender, race), comorbidities, and the use of an angiotensin receptor blocker. The model was validated by analyzing the Receiver Operating Characteristic (ROC) curve. The authors found that urinary albumin excretion (UAE) at baseline, along with blood pressure and plasma glucose, were the most important predictors. The limitations of this model are numerous: the population was exclusively white, and some variables included were obtained through self-report (non explicitly mentioned). The authors also considered albuminuria to be a progressive phenomenon with values above 150mg/day. Despite these limitations, this was the first study to collect data from a significant number of patients and to provide evidence that baseline albumin is the best indicator of progressive albumin levels. Accordingly, if a patient has elevated values, even without cardiovascular risk, that patient should be under an albumin-monitoring program.

Another study from 2012 developed a score to predict MA in Thai patients with type 2 diabetes mellitus (T2DM). (Mongkolsomlit 2012) The authors conducted a case-control study including 225 patients in each group, baseline characteristics, including age, gender, education, duration of diabetes and its

complications, treatment, family history, blood pressure, BMI, consumption of alcohol, and smoking habits. Laboratory data (i.e., creatinine, fasting glucose, HbA1c) were also collected. Stepwise selection was considered when building the model and regression coefficients were transformed into scores. The scores were tested using the ROC approach. Validation was carried out using an independent set of data. Predictors of MA were duration of diabetes, Systolic Blood Pressure (SBP), creatinine, LDL-C, and alcohol drinking. The score obtained was tested in a new group of patients, demonstrating an ability to separate those patients who had developed MA from those who had not. The main limitations of this study include: its design, no score validation was conducted; small sample size and exclusively Thai population. Despite these flaws, the study provides an interesting approach that may be improved with our investigation.

In 2012, a systematic review and meta-regression of risk factors for microalbuminuria in the type 2 diabetes population were published. This study included observational and randomized control trials (RCT) published between 2000 and 2009. After the identification and selection of articles, 22 studies were finally selected (17 of them were cross sectional studies, 2 were case-control, and one was a RCT). Results from this study are in the same direction as studies presented before. Duration of diabetes, blood pressure, and HDL levels were positively associated with the occurrence of MA after adjusting for sex and age. Even though other articles have shown that BMI is associated with MA, this study did not find a positive relationship. The study also emphasized the importance of assessing MA in diabetes population; though this study did not evaluate all possible factors

because, according to the authors, it was difficult to include them from the screened studies. (Mongkolsomlit and others 2012)

There have been few efforts to provide a model to healthcare providers or decision makers to perform an accurate assessment of the occurrence of microalbuminuria. Our investigation uses, for purposes of creating the predictive model, a nationally based database (NHANES) and patient data that belongs to local urologic clinics for validating purposes. This considers approaches already performed and improves them by providing a more parsimonious approach. If some of these methodologies and approaches are combined, we will be able to create a predictor with a reasonable ability to forecast microalbuminuria.

2.3 Use of predictive modeling in urology

In recent years, databases and computer technology have become more advanced in several fields. This same phenomenon has begun to influence medical research as well. The possibility to put together an extensive quantity of data, collected in different healthcare settings and institutions, is starting to provide the power needed to develop models and statistical evaluations, using the urological field and its patients as an extensive source of diagnosis, image of their pathologies, records of the natural history of the disease, results of their treatment, complications, adverse effects, comorbidities, laboratory test, etc. The immense amount of information generated in this field creates the perfect conditions to expand the possibilities of creating, for the most prevalent conditions and diseases, appropriate models in order to better define, understand, and anticipate outcomes

of pathologies themselves and to create accurate and coherent predictive models with significance for patients, physicians, and the healthcare system itself. (Cestari 2013)

Within the new potential use of this large amount of data, predictive modeling is emerging as an interesting approach to better understand patient health behavior and to find people who could potentially fall within a certain risk category. This risk modeling can inform future health consumption decisions and early interventions to achieve some optimal state. A tailored model could also be used by physicians to directly guide patient care; by managers to control supply planning; and by patients to stay healthy. The urology field has been a part of this trend, and many predictive models have been developed in the area, specifically considering oncological diseases such as prostate cancer. The first attempt by a model to predict prostate cancer was made in 1993. (Partin and others 1993) Since then, many models have been created using statistical techniques that employ data stored in different healthcare facilities to predict the result of treatment, surgeries or any interventions used in this population.

The most common tool used is the nomograms, which allows predicting several endpoints using information that is easy to interpret and use in clinical settings. (Stephenson and Kattan 2006) Nomograms give the probability of a specific event based on what has been observed or known about a certain disease in the population under study. This instrument, in all cases, has provided more accurate predictions than other models.(Guillonneau 2007) Nowadays, biomarkers as a surrogate endpoint have begun to be used to predict clinical recurrence, cause

of death, and chance of survival. Several prognostic variables, such as clinical data and biomarkers, have been tested with different success and validity, considering whether they are predictive tools or part of a predictive model.

Furthermore, the investigation is just now starting to move toward other diseases beyond purely oncology. Our state of knowledge in urology outside of cancer is still quite young, but urology is starting to gain more attention from researchers, and it is expected that urology research will evolve using relatively sophisticated approaches. As a result, these models will support clinical decision-making with an expected high accuracy.

In urology, new variables, interactions, and biomarkers have been discovered recently. A new user interface is also being designed to provide a better understanding of the current data and easier access to laboratory test and results to improve care and clinical outcomes.

2.4 National Health and Nutrition Examination Survey, NHANES

The NHANES series that we included for the purpose of this research corresponded to data collected in 2007-2008 (data and documentation have been updated in 2012, and further updates will occur in the future as the NCHS revises the NHANES data collection frequently; however the 2007-2008 data is the most complete for the purposes of this investigation). Every year, around 7,000 persons are cross-examined in their homes and of these, approximately 5,000 complete the health examination component of the survey. The NHANES target population is the civilian and non-institutionalized population. In 2007-2008 a new sampling

methodology was implemented and all Hispanics were oversampled, not just Mexican Americans. Originally, households were identified for inclusion in the NHANES sample and an advance letter was mailed to each address informing the occupant(s) that an NHANES interviewer would visit their home. Trained examiners administered all of the questionnaires. The interview data was collected using the computer-assisted personal interview (CAPI) system. Once eligible individuals were identified, the interviewer proceeded with efforts to recruit these individuals. All cross-examined persons were asked to complete the health examination component. Those who agreed to participate were asked to sign additional consent forms for the health examination component. The examiner informed the participants that they would receive payment as well as compensation for transportation and childcare expenses, if needed.

2.5 Theoretical framework

The use of predictive modeling in urology started around the beginning of the 1990s and has been focused primarily on the prediction of prostate cancer. Evidence has shown that prediction of urologic cancer using the following three variables: serum prostate specific antigen, Gleason score (measured pathological stage through biopsy) and digital rectal examination, is limited. Research published in 1993 was the first to utilize a logistic regression approach, including several combinations of these three variables as predictors, concluding that the combination of variables anticipated the final pathology stage with better accuracy than any variable alone. (Partin and others 1993)

This reinforces the idea that certain pathology is the result of several variables acting together with a variety of strengths and the necessity of weighted predictors (causes) being included in a model (regression), which requires the use of software to obtain the relationships among variables. In 2012, a risk score predicting microalbuminuria in patients with T2DM was published. (Mongkolsomlit 2012) This research stated that no study had yet assessed the predictive value of a combination of risk factors for microalbuminuria and recognized the importance of using these factors (predictors) of different frequency and strengths to develop a probability score with a tested validity to predict microalbuminuria, which would provide early detection and treatment. In our research we want to use this nationally database (NHANES) and statistical software to explore variables (predictors) and the occurrence of MA (effect/outcome), avoiding an overly narrow inclusion criteria.

A cause is an event or a state occurring in nature that starts alone or with the intervention and or synergism of other events and produces an effect. A disease requires the occurrence of one or multiple causes and sometimes requires time to manifest. Most causes in health are considered sufficient, meaning they can by themselves produce an effect. For instance, in order to have Acquired Immunodeficiency Syndrome (AIDS), one must first be infected by Human immunodeficiency virus (HIV). Microalbuminuria (MA), on the other hand, requires more than one cause acting for certain period of time. A particular effect may result from a wide variety of causes (a causal mechanism that inevitably produces disease). Consequently, a cause is not necessarily a single factor, but a minimum set of factors

and circumstances that, if present in a given individual, will produce the disease. These components and their interactions may or may not have related qualities. Figure 3 illustrates that a “sufficient cause” is composed of various factors and that a combination of factors may result in disease. For example, in the figure, “sufficient cause” I accounts for 60 percent of the presence of disease; “sufficient cause” II accounts for 10 percent and “sufficient cause” III accounts for 30 percent.

Figure 2.4 Conceptual scheme for causes of a disease. (Rothman 1976)

A given disease can be produced by more than one causal mechanism and every one could involve the joint action of multiple component causes. Some causes are directly related with the patient (weight, genetic predisposition, behavior, etc.) and other factors are environmental. The importance of multi-causality in the occurrence of diseases has been gaining increased attention because advances in science have identified disease factors that may explain the occurrence of a disease or predict occurrence over a short timeframe. A distinction is sometimes necessary among component to identify those that are “closer” to the disease, implying a more

direct relationship with the outcome. However, causal mechanisms in several diseases remain unclear. The figure also implies that several variable components act together to produce a certain effect, but this does not imply that all of these factors are acting at the same time. Nevertheless, it does provide a biological model for a concept of causal interaction to produce an effect. (Rothman and others 2008)

Individual risk is another aspect that is necessary to include within this reasoning. This type of risk can be considered as illustrating a likelihood of a sufficient cause for a disease existing within the proper timeframe. From an epidemiological point of view, a weak cause confers only a small increase in disease risk for a certain patient, while a strong cause will significantly increase disease risk. Accordingly, the causal risk factor depends on the prevalence or occurrence of the complementary factor. This prevalence can be measured by a relative risk parameter and is dependent on the distribution of the population and of others causal factors.

The sufficient-component cause (SCC) model proposed by Rothman (Rothman 1976) is a useful framework for this investigation because it sets the definition and framework to provide an explanation of the biologic effects of several variables, comorbidities and risk factors, particularly strengths of association and their joint effects. It is necessary to consider that because this research utilized observational datasets, causation cannot be identified; however, it is possible to identify variables or predictors with a high probability of being involved in the occurrence of microalbuminuria.

The SCC model contains the following key characteristics:

- A cause is often not a single component, but a minimal set of conditions or events that produces the effect (outcome).
- Each component in a sufficient cause is called a “cause”. Scientists tend to refer to components as “causes” because the outcome will not occur by that pathway if any one of the components is missing (or prevented).
- There may be a number of sufficient causes for a given disease or outcome (it varies for every scenario).
- A component cause that must be present in every sufficient cause of a given outcome is referred to as a “necessary cause”. For example, tuberculosis (TB) exposure is necessary for TB infection to occur. Diabetes and hypertension may be linked with the occurrence of MA, but are not component causes.
- The completion of a sufficient cause is synonymous with the biologic occurrence of the outcome (this is not clear in the case of MA).
- The components of a sufficient cause need not act simultaneously; they can act at different times. In contrast, the latent period is the interval between disease onset and the clinical detection of disease, either by screening or as a result of symptoms and diagnosis.

The SCC may be considered to be a particular response configuration of a potential outcome model. The configuration in which certain diseases occur is defined based on a logical, biologic conceptualization of the way different variables act in causing disease. (Flanders 2006) Under this premise, predictive modeling will capture the

association between the predictor and the outcome of interest. Sometimes prediction modeling is incorrectly referred to as causal inference or causal reasoning, and it involves the using of a causal model (SCC) to find the probability of future events and, ideally, determine what will result from an intervention in the healthcare system. (Kleinberg and Hripcsak 2011) Today healthcare informatics could benefit from both methodological developments in causal inference as well as more explicit discussions of how causal claims can be supported. Algorithms for constructing predictive models and identifying links or relationships between presumed variables and their corresponding outcomes could have broad applications for healthcare policies. Causation then could be evaluated through an alternate, appropriate study design, e.g. a double blind randomized, controlled, clinical trial.

Predictive models can be applied to deal with some of the most expensive healthcare problems we face, such as late diagnosis and lack of preventive care, and to develop understandings and policies capable of enhancing care while minimizing costs. In particular, these models could be used to enhance the management of chronic disease, improve medical procedures and safety, and leverage real-time detection and prediction in local care settings.

CHAPTER 3

METHODS

3.1 Study Overview

The ability to predict clinical outcomes is of highest importance for the patient-physician relationship. (Grobman and Stamilio 2006) If physicians were able to predict of a patient's clinical course with high precision, a treatment could be modified, and outcomes could be improved. Alternatively, patients could participate in their own treatment by making the most conscious and educated choices from the different treatment strategies available according with their risk.

This research utilized a national retrospective database (NHANES) to develop a model able to predict the occurrence of microalbuminuria. A predictive model was developed through collaboration with research and healthcare professionals. This modeling research is relevant because represents the first attempt to develop a model to investigate a urologic disease other than to predict a cancer outcome, using a retrospective nationally database. This model could have an impact on patients with certain characteristics that put them at risk of developing microalbuminuria (MA) in the short, medium or long-term. This information will allow healthcare providers to put these patients on an intervention path, for which a decision analytic methodology can be used to evaluate certain interventions and their consequences.

3.2 Study design

This investigation was a retrospective database study. The data utilized for the creation of the model and validation was the data from the National Health and Nutrition Examination Survey (NHANES), which contains a series of cross-sectional national health examination data conducted in mobile investigation units or health care settings such as clinics. Data were collected on indicators of the nutritional and health status of the American population in a representative manner through dietary data, biochemical tests, physical measurements, and clinical assessments. Specific components of this dataset address topics such as cardiorespiratory fitness, physical functioning, lower extremity disease, full body scans for body fat, bone density, and infections. Data also include information on chronic disease prevalence and conditions (including undiagnosed conditions) and risk factors such as obesity, smoking status, serum cholesterol levels, hypertension, diet and nutritional status, immunization status, infectious disease prevalence, health insurance, and measures of environmental exposure. Other topics addressed include hearing, vision, mental health, anemia, diabetes, cardiovascular disease, osteoporosis, oral health, mental health, pharmaceuticals and dietary supplements used, and physical fitness. The database utilized for validation purposes was the two third portions of the NHANES. This database was collected by people other than the researcher and independent of any specific hypothesis for the purposes of clinical care, follow up and billing. The chance for observer bias is consequently reduced.

NCHS used several methods to monitor the quality of the analyses performed by the NHANES to assure accuracy in questionnaire and laboratory measures. These

methods included analyzing “blind” samples collected during practice sessions. In addition, contract laboratories randomly performed repeat testing on five percent of all specimens. The participant’s lab tests were needed to clarify any identified areas of concern before data was considered valid.

The use of retrospective data is relatively efficient and inexpensive. Moreover, NHANES was gathered in a consistent and uniform way (i.e., using a cross sectional design systematically applied nationwide or in the provision of usual care), allowing for comparisons over time and between different settings. (Motheral and others 2003). In general, clinical databases, provide a convenient source of data on a large number of patients, which can expedite research involving rare diseases or events, such as MA, that are not routinely investigated. While randomized clinical trials are the gold standard to investigate causality, most of the time they are impossible to implement in practice because clinical trials of equivalent magnitude and detail to investigate the occurrence of rare conditions such as MA would be financially and logistically infeasible. Because all medical appointment data and laboratory test results are included in these databases for each patient, a broad group of covariates can be studied. The design of this study using a large database also facilitates this research because it examines health outcomes (MA) over a period of time (i.e., over years), increasing the robustness of the findings.

3.3 Database

3.3.1 NHANES (National Health and Nutrition Examination Survey)

NHANES includes clinical assessments, selected medical and laboratory tests, and self-reported data. Medical examinations and laboratory tests follow very specific protocols and are as standard as possible to ensure comparability across sites and providers. Data is collected every year from a representative sample of the civilian non-institutionalized U.S. population, newborns and older, by personal interviews and physical examinations in the mobile examination centers. The sample design is a multisite design. In the database for 2007-2008 (recently updated in some components with data from 2012), low-income individuals, people over 60, African Americans, and people with Hispanic origin were oversampled. The 2007-2008 database was selected because, at the time of initiation of this dissertation, the 2007-2008 database was the most updated database, particularly with regard to the laboratory data. That is, even though the more recent 2011-2012 database was available, many of the laboratory tests, and some clinical variables, were not yet published; therefore, this database was not utilized in this dissertation. The NHANES general content includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by qualified medical workers. NHANES are also the basis for national standards for such measurements as height, weight, and blood pressure. This database was used for the creation of the model that may be able to predict the occurrence of microalbuminuria. In total, the retrospective database contains information for

10,149 participants including 503 variables. All patients available and their information were under consideration for the modeling part of this research.

The Public Health Service Act states that the data collected by the National Center for Health Statistics (NCHS) may be used only for health statistical reporting and analysis. This act prohibits any effort to determine the identity of any patient. To comply with this law, all direct identifiers, as well as any characteristics that might lead to identification, are omitted from the NHANES dataset. Therefore, users will: (1) use the data in this dataset for statistical reporting and analysis only, and (2) make no attempt to utilize the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery, and (3) not link this dataset with individually identifiable data from other NCHS datasets. By using these data, we signify our agreement to comply with the above-stated statutorily based requirements.

The database was divided into thirds, one was utilized to create the model (model development) and the other subsequent portions were utilized for validation. The partitioning process was developed without replacement, so each observation from the dataset appeared once in the sample. A random number was generated for each observation, and the partitioning was produced by using the STATA “sample” command and dividing the database into three thirds.

3.4 Study Outcome

Proteinuria is described as an excessive elimination of any protein through urine; the principal protein found in the urine of patients with proteinuria is albumin. Under ordinary circumstances, daily albumin excretion is in the range of 5–10 mg and the urine albumin to creatinine ratio is in the range of 0–29 mg albumin/g creatinine. (Keen and Chlouverakis 1964; Khosla and others 2006; Toto 2004) Microalbuminuria (dependent variable) was defined within this research as an abnormal increase in the albumin excretion rate within the specific range of 30–300 mg of albumin/g of creatinine. (Khosla and others 2006) It is necessary to distinguish that the term microalbuminuria refers to an unusual albumin excretion rate and not the occurrence of an anomalous albumin molecule. In the clinical database from which the data were extracted the dependent variable (MA) was calculated through the utilization of calculation of the albumin/creatinine ratio (AUCR) because recently published evidence have shown that the calculation of an AUCR exhibited a higher sensitivity and specificity in the detection of microalbuminuria in different groups of patients. (Derhaschnig and others 2002) AUCR values of 30–300 mg/g were considered as positive for microalbuminuria. (Khosla and others 2006) For the purpose of developing the predictive model, the variable microalbuminuria was labeled as present in patients with AUCR levels of 30–300 mg/g, and absent in patients with values below 30 mg/g (normal albumin). Patients with levels above that range (>300 mg/g or clinical proteinuria) were excluded from the analysis.

3.5 Candidate Predictors

Candidate predictors (independent variables) were obtained from patient demographics, clinical history, physical examination, disease characteristics, test results, and any relevant variable linked with our outcome. Candidate predictors were selected following recent literature reports. These candidates were considered but not exclusively to those related with (1) patient demographics: age, gender, race, marital status, income/ socio economic status, health insurance (private, Medicare, Medicaid, other), occupation, education; (2) patient medical variables: diagnoses and comorbidities (other diseases: e.g. presence of diabetes, obesity, cardiovascular disease, liver disease, renal disease, etc.), number of visits, medication (prescribed, medication history), weight, height, body mass index, smoking status, alcohol intake, drug used (if it is available) , medical history (family disease history), clinical symptoms (description available), allergies, adverse events, diet (health behaviors); and (3) patient laboratory test / physical examination: body mass index, blood pressure, serum triglycerides, total cholesterol, blood uric acid / blood urea nitrogen, serum creatinine concentrations, fasting plasma glucose, HbA1c levels, albuminuria, hemoglobin, estimated glomerular filtration rate, albumin-creatinine ratio, creatinine clearance, and C-reactive protein. It is important to state that this research included only predictors that were available in NHANES at the time the model was created and that could be obtained in routine clinical practice. Last generation biomarkers and the variables measured through sophisticated technology/equipment were not included because they are often not available in routine clinical practice. Other researchers have stated that predictors that are

difficult to measure, or have high inter-observer variability, might not be suitable for inclusion in a predictive model because this will influence the predictive ability of the model when applied for other individuals. (Hendriksen and others 2013)

3.5.1 Variable definitions

A series of definitions were used to categorize patients in NHANES database. All of these variables were included in the model as continuous variables; however, to develop a risk score table, some variables were categorized into ordinal level data. Blood pressure was categorized into groups of normal, <120/80 mm Hg; prehypertension, 120-139/80-89 mm Hg; hypertension, $\geq$ 140/90 mm Hg or receiving blood pressure medications.(Chobanian and others 2003) Body mass index (BMI) was calculated as the ratio between weight and the squared of height and classified as follows, normal, <25; overweight, 25-29; and obese, $\geq$ 30.(1998) Fasting glucose levels were considered normal if <100 mg/dL [$<$ 5.55 mmol/L]; pre-diabetes mellitus if 100-125 mg/dL [5.55-6.94 mmol/L]; and diabetes mellitus if $\geq$ 126 mg/dL [$\geq$ 6.99 mmol/L] or if receiving insulin or oral hypoglycemic agents. (Genuth and others 2003) Total cholesterol levels normal were categorized as follows <200 mg/dL [$<$ 5.18 mmol/L]; borderline high; 200-239 mg/dL (5.18-6.19 mmol/L); and high, $\geq$ 240 mg/dL [$\geq$ 6.22 mmol/L] or receiving lipid-lowering medications. Triglyceride levels were considered normal if <150 mg/dL [$<$ 1.70 mmol/L]; borderline high if 150-199 mg/dL [1.70-2.25 mmol/L]; and high, $\geq$ 200 mg/dL [$\geq$ 2.26 mmol/L] or receiving lipid-lowering medications. High-density lipoprotein cholesterol values were considered to be normal if $\geq$ 40 mg/dL ($\geq$ 1.04

mmol/L) for men or ≥ 50 mg/dL (≥ 1.30 mmol/L) for women; low if < 40 mg/dL (< 1.04 mmol/L) for men and < 50 mg/dL (< 1.30 mmol/L) for women. (Expert Panel on Detection and Treatment of High Blood Cholesterol in 2001) Elevated levels of CRP (C-reactive protein) were classified as greater than 3 mg/L (and as a consequence dichotomized), based on the Centers for Disease Control and Prevention and the American Heart Association scientific statement. (Pearson and others 2003)

3.6 Sample size considerations

The importance of sample size depends on the context. If an analysis is performed on data that were available for other purposes, such as is the case in this research; the main question is whether the analysis of the data will produce results with sufficient statistical precision to contribute substantially to the literature, in which case sample size considerations will be informal. Under this scenario a sample size calculation may be useful when planning a new study from findings captured from the original study. Authors may consider that calculations are associated with more uncertainty than implied by the single number that is generally produced. For example, estimates of the rate of the event of interest or other assumptions to calculations are commonly imprecise, if not speculation. The precision can often not be determined beforehand because it will be reduced by inclusion of confounding variables in multivariable analyses, the degree of precision with which key variables can be measured, and the exclusion of some individuals. (Vandenbroucke and others 2007)

In addition to those considerations stated in the STROBE study, a general approach to estimate a sample size necessary to develop a predictive model using regression approaches has been proposed. In many situations a fitted regression model is likely to be reliable when the number of predictors (or candidate predictors) “p” is less than $m/20$, where “m” is the limiting sample size available in the database.(Harrell 2001) The three studies already published about predictive models and microalbuminuria have fewer than six predictors. The following table summarizes the minimum sample size necessary for this research considering evidence already published.

Table 3.1 Sample size necessary to create a predictive model (Harrel 2001)

Candidate predictors	Sample size necessary
Less than 10	500
11-20	1,000

NHANES database included 10,149 individuals of whom 7,878 had laboratory variables necessary for this research. Other predictive model studies published utilized between four and six predictors. The number of individuals satisfied the Harrell’s sample size estimations.

3.7 Analysis: Statistics and Modeling Process

There are two main methods of analyzing and summarizing data and findings in this study, a numeric method and a graphic method. The numeric method involved using descriptive statistics to summarize the main features of predictor variables in a table summary form, while the graphic method involved using several graphs to visualize the main aspects of each model and its performance in each dataset.

This univariate analysis is used mainly for descriptive and analytics purposes, and involves frequency tables and descriptive statistics. This general approach was used to identify specific information relating to each predictor variable. This information included mean, skewedness, and standard deviation. Also, comparisons were conducted between the modeling and validation sets. Categorical variables were compared using a Chi-square test. Continuous variables were compared using t-tests (Mann-Whitney U test, in case of those variables not normally distributed).

3.7.1 Study of Correlation

To satisfy the first research objective, we identified the strength and direction of association of predictor variables (independent variables) and the MA (dependent variable). The first part for this approach is considered variables in continuous scale and the utilization of Pearson correlation to measure the correlation between AUCR (mg/g) and the independent variables. The correlation “r” indicates direction and magnitude. Pearson correlation may take values from -1 to +1. A coefficient of zero indicates no association between the variables. A coefficient

of ± 1 means a perfect correlation (very strong). Pearson coefficient is affected by extreme values (e.g. outliers), which may inflate or inhibit the strength of relationship, and it is then inappropriate when the variables are not normally distributed. Because of that, another correlation measure was considered in this research, Spearman rank correlation (rho) is appropriate when we compared the correlation between AUCR and a skewed variables or ordinal (non-parametrical correlation test) and also is robust when extreme values are present, the interpretation of this test was similar to Pearson correlation. Both results were compared and reported.(Mukaka 2012) (Hinkle and others 2003)

Like a measure scale, the correlation coefficient is difficult to interpret specially in medical sciences. Labeling systems roughly categorize values where correlation coefficients (in absolute value) < 0.3 are generally considered to represent low correlations, 0.4 to 0.6 moderate correlations, 0.7 to 0.9 strong correlations and 1 perfect.(Dancey and Reidy 2004) Correlations were reported and based only on those observations that have non-missing values on all of the listed variables. Correlations were presented individually for each pair considering a significant value of $\alpha = 0.05$.

3.7.2 Variable Transformation

A total of seven variables describing the health of the subject, five describing demographic characteristics and fifteen concerning laboratory tests, were screened in univariate analyses. Variable transformation is an important component in modeling and is utilized to achieve linear relationships with the dependent variable.

Transformations were considered when clinical criteria and previous studies showed that two variables are known to be correlated and yet do not appear so. In this case, an appropriate transformation may help clarify the relationship between the variables. Also the processes of utilizing transformations create a uniform distribution of the variables. The principle is that there is (usually) nothing special about how the data are originally expressed, so we let the data suggest re-expressions that lead to effective, accurate, useful, and (if possible) theoretically justified models.

The approach utilized depending on the shape of the curve, the transformation suggested were for instance quadrant I: either variable in X or Y-axes squared.

Figure 3.1 illustrates different options of variables transformation. (Afifi and others 2012)

Figure 3.1 Transformation options according with different variable distribution

3.7.3 Multivariate logistic regression and model construction

3.7.3.1 Variable Selection Method and Model Construction

Identifying patient risk variables (independent variables) assists health professionals in optimizing patient care; however, there is no clear scientific consensus regarding the best method to select variables most likely related with the outcome and including them in the final model. There is also no consensus regarding how to maintain a balance between including too many variables and making the model parsimonious. (Austin and Tu 2004)

The goal of using logistic regression is to accurately predict the category of outcome for an individual using the most parsimonious model. To accomplish this goal, the model has to include all predictor variables that are useful in predicting the presence of MA. Several methods are available during model construction, keeping in mind that omitting important predictor variables may result in a systematically incorrect estimation of regression coefficients and as a consequence the OR. However, including too many predictors results in not accurate regression coefficients estimation.

We wanted to build a model that would handle the data in the simplest way, considering redundant predictors must be removed. Under this premise, when considering several plausible explanations for a phenomenon, the simplest is best. When applied to regression analysis, this reasoning indicates that the most parsimonious model that fits the data is preferable. Unnecessary predictors increase coefficient estimations noise. Collinearity could be caused by using too many variables to explain the same phenomenon. The descriptive statistics previously

mentioned were used to also identify outliers and influential measurements. Variable transformation was completed when appropriate to satisfy logistic regression assumptions.

Variables were entered into the model in the order specified by the researcher testing the fit of the model after each variable is added or deleted. Automated variable selection methods have been developed to facilitate or help in this process: (1) backward elimination, which begins with a full model including all candidate predictor variables, after which variables are sequentially eliminated from the model until a pre-specified rule is satisfied (e.g., significance level $p < 0.30$); (2) forward selection which starts with the empty model, then each variable is added following a sequential order until the stopping rule is satisfied. At the end of the selection process, those variables whose addition would result in the greatest increase in the summary measure are retained in the model; (3) on stepwise selection in which each step of the variable selection process, after a variable has been added to the model, predictors are permitted to be eliminated or retained in the model. For example, if the significance of a given predictor is over a pre-specified limit, it is eliminated from the model. The iterative process is ended when the stopping rule is satisfied. For this study, stepwise selection was used, as this technique places more control over the information and is especially useful for selecting over large numbers of potential independent variables.

The procedure used selects independent variables based first on, substantive and theoretical relevance, and second, on $p\text{-values} \leq 0.2$ (which the procedure later modifies in view of more complete sets of variables and the modeling of

nonlinearities). The process began by testing potential explanatory variables with the dependent variable. The significance level of $p\text{-value} < 0.2$ was selected as alpha priori. Even though this study utilized this semi-automatized technique to build the model, clinical judgment prevailed as a final criterion for variable selection. The process utilized was the following:

Command Syntax: `sw logistic depvarname indepvar, options`

`sw`: is the command for stepwise selection

`reg`: request logistic regression

`depvarname`: is the name of the dependent variable

`indepvar`: is the list of independent variables

Options `pr(#)` - specify the p-value for removal (e.g. `pr(0.2)`)

`pe(#)` - specify the p-value for entry forward - specify forward selection

Since stepwise included a combination of forward and backward selection, the program allowed the specification for separate of each removal or entry probability as follow:

`pr(#)`: only performed backward elimination

`pr(#) hier`: backward elimination in hierarchical manner

`pr(#) pe(#)` - stepwise (starting with a full model). It was the prefer method.

`pe(#)` - forward selection

`pe(#) hier` - forward selection in hierarchical manner

`pe(#) pr(#) forward` - stepwise (starting with a empty model)

The process of constructing a predictive model produces several models than happen to be equally acceptable. To satisfy the study's second objective of developing a patient data driven predictive model, a logistic regression approach was selected. Logistic regression is becoming universal because it provides a method for modeling a binary response outcome, such as presence of MA (dependent variable), taking into consideration a set of covariates (independent variables), which can be quantitative, qualitative, continuous or categorical. Binary data occur frequently in healthcare and following this rationale, this technique is useful when it may not be possible to accurately predict a continuous response with the covariates available but a prediction including categorized variables may be more feasible.

Logistic regression has a wide range of applications in medical and biomedical research, mainly to formulate models organizing the factors that might determine whether or not an outcome occurs. The distinguishing feature of logistic regression models is that the outcome variable is binary or dichotomous, in this case the presence or absence of MA.

Patient data analyses from the univariate assessment were used to establish which variables are influential in predicting the given outcome. These variables can then be measured for a new patient by entering this patient's characteristics into the logistic regression model to calculate the probability of given outcome called (MA). The logistic model used in this investigation applies the logit link to formulate the relationship between a binary outcome (MA) and its associated predictors,

following the general equation, where π_i is the mean response $E(y_i)$ or the probability $\Pr(y_i = 1)$:

Where: $\pi(x)$ = probability of disease (Microalbuminuria)

X_1 = first independent variable

X_2 = second independent variable

X_k = kth independent variable

α = intercept

β_1 = slope corresponding to the first independent variable

β_2 = slope corresponding to the second independent variable

β_k = slope corresponding to the kth independent variable

Then the logit transformation is used, because the logistic function is not linear in the coefficients

$$\log\text{-odds}(\text{disease}) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_k \cdot x_k$$

3.7.3.2 Predicted Probability

Because the dependent variable MA is not a continuous one, the goal of using logistic regression in this study is different because the intention was to predict the likelihood that Y is equal to 1 (rather than 0) given certain values of X . That is, if X and Y have a positive linear relationship, the probability that a person will have a score of $Y = 1$ (MA present) will increase as values of X increase. Therefore, in this study we used predictive probabilities rather than scores for the dependent variable

(MA). Then, the probability of having the outcome can be calculated based upon the following equation, where the odds ratio is equal to e^β :

$$\hat{p} = \frac{\exp(b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p)}{1 + \exp(b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p)}$$

$$\ln\left(\frac{\hat{p}}{(1 - \hat{p})}\right) = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

3.7.3.3 Odds ratio

The odds ratio (OR) represents the odds that an outcome will occur given a certain exposure, compared to the odds of the outcome occurring in the absence of that exposure.

The interpretation of this logistic model involves the concept of OR. The odds in favor of the event $Y = 1$ is defined as $\pi / (1 - \pi)$, where $\pi = \Pr(Y=1)$. The odds ratio then is formulated in terms of conditional probabilities where:

$Y = 1$ if individual i is found with microalbuminuria, and

$Y = 0$ otherwise

The odds ratio is also a very popular measure of the effect size, describing both the direction and the strength of association between two variables. As mentioned previously, odds ratio play an important role in interpreting the coefficient of each independent variable and its relation with the outcome (MA).

Back in the general model for this research, the slope parameter β_k corresponds to

the logarithm of the odds ratio in favor of $Y = 1$ with a one unit increase in predictor X_k , holding other predictors fixed. (Bland and Altman 2000; Szumilas 2010)

The final model's proposed logistic regression uses the Maximum Likelihood Estimation (MLE) technique, which is the way to find the smallest possible deviance between the observed and predicted values (finding the best fitting line). In other words, estimates the odds that the dependent variable values can be predicted using the independent variable values. This estimation is accomplished by starting out with a random set of coefficients, and then iteratively improving them based on improvements to the log likelihood measure. After a few iterations, the process stops when further improvement is insignificant, based on predetermined criteria. (Bewick and others 2005)

3.7.3.4 Risk score calculation

The multivariable model generated to predict the occurrence of microalbuminuria contains regression coefficients that represent weights of the variables that are significantly associated with the occurrence of MA and are available in clinical practice.

To make the model easy to use in clinical practice and to minimize confusion in the mathematical calculation a risk score was calculated. The general step for developing a point system was:

1. Estimate parameters of the multivariable model: coefficients estimated from the regression model that were identified as being statistically related with MA were considered.

2. Risk factors were organized into categories and reference values were determined: If a risk factor is continuous such as age a range between 18-80 was considered. This was divided into categories such as 18-28, 29-39, 40-49. In order to determine points for each category, a reference value was specified for each category, mid points were considered as generally acceptable reference value (e.g., 23, 34, 45). When a risk factor was included as a dummy variable (coded as 0=absent or 1=present) then the reference value was 0 or 1.
3. Determine the referent risk factor: we determine the appropriate category for each risk factor to serve as the base category. This base category was assigned with zero points in the scoring system.
4. Determine how far each category is from the base category (using regression units): For each risk factor calculate how far each category is from the base category multiplying by that value the corresponding regression coefficient.
5. Choosing a constant: The constant generated was utilized to calculate a final normalized score. The smaller coefficient was chosen as constant.
6. Determine the number of points: it was obtained by dividing the value in point number four above by the number obtained in point five above and rounding to the closest 0.5 number. The reference category had a value of zero.
7. Determine the risk associated with total points: the final step was assigning a risk score with each point total. The table summarizing each variable and its risk score was generated.

There are some limitations with the point system. Because simplicity of use was the objective for the scoring system, some information was lost and is only captured by using the equation directly. These missing information are more pronounced when the risk factors are modeled as continuous variables because the points system is based on categories. (Sullivan and others 2004)

3.7.4 Assessing the Quality of the Model: overall performance, discrimination and calibration

It is not sufficient to base a developed model's performance on the development sample only simply because some models show overly positive scores. It is essential to confirm that a developed model is predictive with similar but different individuals outside the development set. The more these other situations differ from the development study, the stronger the test of generalizability of the model. Internal validation does not make use of data other than the development data, and therefore will not provide the degree of heterogeneity that will be encountered in real-life applications of the model.(Moons and others 2012a) A prediction model usually performs better in the database or sample used to construct the model than in other samples, even though the same population from which the model was created is tested. The third objective of this research involves the validation of the performance of the predictive model to identify patients at risk of microalbuminuria. There are several methods to assess the performance of a prediction model. The general approach for evaluating the overall performance is calculating the pseudo R^2 . The traditional approach for calibration quantifies how

close prediction is to the real outcome (MA), using measures such as the Hosmer-Lemeshow “goodness-of-fit” test. Also, discrimination is essential considering the following premise: do patients with the outcome have more risk predictors than those without? Discrimination can be quantified with measures such as sensitivity, specificity, and the area under the receiver-operating characteristic ROC curve.

3.7.4.1 Overall performance

During the process of developing a predictive model it is necessary to consider a measure of overall assessment. In linear regression the coefficient of determination R^2 is the most current statistic that summarize the overall strength of a given model, with zero indicating a model with no predictive value and one indicating a perfect fit. Researchers have searched for a corresponding indicator for models with binary outcome. Many different R^2 statistics have been recommended. Cox and Snell's R^2 are based on the log likelihood for the model compared to the log likelihood for a baseline model. However, with categorical outcomes, it has a theoretical maximum value of less than 1, even for a "perfect" model. Nagelkerke's R^2 is an adjusted version of the Cox & Snell R -square that adjusts the scale of the statistic to cover the full range from 0 to 1. Both measures were computed in the assessment of the best model.

3.7.4.2 Goodness of Fit Test (GOF)

Calibration refers to the agreement between observed outcomes and predictions. For instance, if we predict a 20% MA occurrence for a urologic patient,

the observed frequency of MA should be approximately 20 of 100 patients. For calibration, one approach was utilized in this research to assess the model's adequacy, was the Hosmer-Lemeshow test, applicable to binary regression model. Hosmer-Lemeshow suggested grouping cases together according to their predicted values from the logistic regression model. The predicted values are ordered from the lowest to the highest, and then separated into several groups of approximately equal size. Ten groups is the standard recommendation. For each group, the observed number of events and non-events was calculated, as well as the expected number of events and non-events. (Hosmer and Lemeshow 1980) These calculations were obtained as follows: the expected number of events is the sum of the predicted probabilities over the individuals in the group. And the expected number of non-events is the group size minus the expected number of events. Pearson's chi-square was applied to compare observed counts with expected counts. The degrees of freedom (df) were the number of groups minus two. As with the classic GOF tests, the underlying hypothesis is that the estimated and observed frequencies agree. If the test is significant ($p\text{-value} < 0.05$), there is lack of fit.

3.7.4.3 Receiver Operator Characteristic Curve Analysis (ROC)

When the accuracy of a predictive model is assessed, it is important to assess its discrimination considering both sensitivity and specificity and also that as sensitivity increases the specificity decreases and vice versa. An ROC curve studies the trade-off between these two measurements. To build this curve, it is necessary to calculate sensitivity and specificity. Sensitivity (or true positive) is plotted on the

y-axis and 1-specificity (false positive) on the x-axis. The most parsimonious model is the one that generates the point on the ROC curve closest to the upper left corner. A model is considered useless when it has a ROC curve that is close to the 45-degree diagonal. The area under the ROC curve is often termed the c-statistic, corresponding to the likelihood that the outcome studied (MA) will have a higher predicted $\Pr(Y=1)$ than a non-event (ranging from zero to one), which accounts for the model's ability to discriminate between those patients who experience MA versus those who do not. A rule of thumb for interpreting this value in this research is $AUC = 0.5$ equal to no discrimination; $0.7 \leq AUC \leq 0.8$ equal to acceptable discrimination; $0.8 \leq AUC \leq 0.9$ equal to excellent discrimination and $0.9 \leq AUC$ equal to outstanding performance. (Hosmer and others 2000) Below are three hypothetical ROC curves representing the accuracy of three different models: gold standard (lines A; $AUC=1$) on the upper and left axes, a typical model's ROC curve (curve B; $AUC=0.85$), and a diagonal line corresponding to random chance (line C; $AUC=0.5$). As validation model accuracy improves, the ROC curve moves toward A, and the AUC approaches 1. (Zou and others 2007) Figure 3.2 shows the three hypothetical receiver operator curve (ROC) described before.

Figure 3.2 Hypothetical receiver operator curve (ROC)

3.7.4.4 Validation

Validation studies are particularly important if a prediction model is to be used with individuals who were not represented in the development sample. Application of a predictive model requires a clear definition of predictors and reproducible measurements using methods available in clinical practice. Cross validation is a statistical technique for assessing how the results of a model will generalize to an independent data set. It is mostly used in research where the goal is prediction and where one desires to evaluate how precisely a predictive model will perform in practice. In this research a predictive model was developed using a portion of the NHANES dataset, and a portion of the dataset was used for validation

purposes against which our model was tested. The goal of validation is to test the model in order to limit problems like over fitting, and understand how the model will generalize to an independent data set.

Researchers indicate that the most common technique of cross validation is the split approach, which splits one data set into two. Splitting one dataset into a development and validation set may be ineffective because it assumes heterogeneity between the two parts of the data set, when there is actually little difference between them. Despite this shortcoming, this technique is relatively easy to implement and is generally accepted by most researchers. (Steyerberg 2009) In this study, however, a true internal validation was performed. The approaches used evaluate the predictive performance comparing observed outcomes to expected values. Because the validation set is not sampled under exactly the same circumstances as NHANES, this procedure may do a better job of forecasting the utility of the prediction model for practical use. The NHANES database was split in two portions, one of each was utilized in the creation process and two-thirds were used for validation purposes.

3.7.5 Statistical Package and Statistical Test

All statistical analyses were performed using STATA statistical software, version 11.2 (StataCorp, College Station, Texas, USA). Differences between the variables contained in the derivation database and validation database were analyzed with an unpaired Student's t-test for continuous variables and the χ^2 -test for categorical variables. Univariate and multivariate logistic regression analyses

were used to assess the correlation between variables and to select the final variables for the logistic regression equation model, to predict the probability of microalbuminuria. Probabilities generated from the regression equation were recorded and used as new input variables for receiver operating characteristic (ROC) curve analysis. The accuracy of the microalbuminuria predictive logistic regression equation was evaluated by constructing ROC curves. The calibration was assessed using Pseudo- R^2 and Homer-Lemeshow test.

CHAPTER 4

RESULTS

4.1 Descriptive Statistics

4.1.1. NHANES dataset

From the 10,149 participants of the NHANES (National Health and Nutrition Examination Survey) 2,271 participants were excluded because data (i.e., urine albumin, urine creatinine) needed to calculate their albumin creatinine ratios (AUCR) were not available. Data were available for 7,878 participants; these participants are hereafter referred as subjects.

The dataset was divided in three equal portions. A random number was assigned to all subjects and the command “sample” was utilized to divide the dataset. One-third of the dataset was utilized for model development, and two-thirds were used for validation purposes.

Table 4.1 shows the clinical characteristics of participants considered in this investigation. Three variables (i.e., total protein, blood nitrogen urea and serum albumin) were significantly different in the development and validation dataset. No other differences were found when examining demographic and clinical variables.

Table 4.1. Demographic and clinical characteristics for the overall, development and validation datasets

Variable	All (7,878)	Development Dataset (2,500)	Validation Dataset (5,378)	p-value
	N (%)	N (%)	N (%)	
Sex				0.692
<i>Male</i>	3,929 (49.8)	1,255 (50.2)	2,674 (49.7)	
<i>Female</i>	3,949 (50.2)	1,245 (49.8)	2,704 (50.3)	
Race/ Ethnicity				
<i>Mexican-American</i>	1,533 (19.5)	483 (19.3)	1,070 (20.0)	0.455
<i>Other Hispanic</i>	940 (12.0)	294 (11.8)	646 (12.0)	
<i>White</i>	3,283 (41.8)	1,039 (41.6)	2,244 (41.7)	
<i>Black</i>	1,769 (22.5)	563 (22.5)	1,206 (22.4)	
<i>Others</i>	333 (4.2)	121 (4.8)	212 (3.9)	
Marital status				
<i>Married</i>	2,952 (37.5)	921 (36.8)	2,031 (37.8)	0.885
<i>Widowed</i>	477 (6.0)	154 (6.2)	323 (6.0)	
<i>Divorced</i>	806 (10.2)	250 (10.0)	556 (10.3)	
<i>Single</i>	1,287 (16.3)	412 (16.5)	875 (16.3)	
<i>Other</i>	2,356 (30.0)	763 (30.5)	1,593 (29.6)	
Health Insurance				
<i>Yes</i>	6,155 (78.1)	1,930 (77.2)	4,225 (78.6)	0.174
<i>No</i>	1,723 (21.9)	570 (22.8)	1,153 (21.4)	
Alcohol consumption				
<i>Yes</i>	3,609 (70.5)	1,148 (70.7)	2,461 (70.4)	0.877
<i>No</i>	1,509 (29.5)	476 (29.3)	1,033 (29.6)	
Smoking				
<i>Yes</i>	2,614 (47.4)	816 (46.9)	1,798 (47.6)	0.887
<i>No</i>	2,905 (52.6)	922 (53.1)	1,983 (52.5)	

Table 4.1. Demographic and clinical characteristics for the overall, development and validation datasets (continued)

Variable	All (7,878)	Development Dataset (2,500)	Validation Dataset (5,378)	p-value
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	
Age (years)	38.9 $\pm$ 23.1	39.0 $\pm$ 23.3	38.8 $\pm$ 23.0	0.7205
Body mass index (kg/m ²)	26.7 $\pm$ 7.3	26.6 $\pm$ 7.2	26.7 $\pm$ 7.2	0.5661
Systolic blood pressure (mm Hg)	121.0 $\pm$ 19.4	121.8 $\pm$ 19.3	121.0 $\pm$ 19.5	0.0891
Diastolic blood pressure (mm Hg)	67.0 $\pm$ 14.4	67.1 $\pm$ 14.5	66.9 $\pm$ 14.2	0.5633
Mean arterial pressure (mm Hg)	85.0 $\pm$ 13.5	85.1 $\pm$ 13.6	84.9 $\pm$ 13.4	0.5395
Hemoglobin A1c (%)	5.7 $\pm$ 1.0	5.6 $\pm$ 0.9	5.6 $\pm$ 1.1	1
Glucose serum (mg/dL)	100.2 $\pm$ 38.4	99.9 $\pm$ 37.7	100.3 $\pm$ 38.8	0.6674
High density lipoprotein (mg/dL)	52.0 $\pm$ 15.1	52.1 $\pm$ 15.4	52.0 $\pm$ 14.9	0.7839
Cholesterol (mg/dL)	186.7 $\pm$ 42.1	185.4 $\pm$ 41.4	187.3 $\pm$ 42.5	0.064
Triglycerides (mg/dL)	153.1 $\pm$ 132.4	151.7 $\pm$ 124.6	153.7 $\pm$ 135.8	0.5325
Serum Creatinine (mg/dL)	0.85 $\pm$ 0.31	0.86 $\pm$ 0.32	0.85 $\pm$ 0.31	0.1872
C-reactive protein (mg/dL)	3.90 $\pm$ 8.01	4.08 $\pm$ 8.59	3.82 $\pm$ 7.72	0.1798
Alb/Cr (mg/g)	25.9 $\pm$ 71.2	27.2 $\pm$ 76.6	25.2 $\pm$ 69.2	0.2488
ln(Alb/Cr)(mg/g)	2.3 $\pm$ 1.2	2.3 $\pm$ 1.2	2.3 $\pm$ 1.1	0.5512
Creatinine Urine (mg/dL)	124.3 $\pm$ 77.8	124.7 $\pm$ 77.8	124.1 $\pm$ 77.8	0.7906
Albumin Urine (mg/dL)	48.2 $\pm$ 412.7	48.8 $\pm$ 310.8	47.9 $\pm$ 452.3	0.9282
Globulin (g/dL)	2.9 $\pm$ 0.4	2.9 $\pm$ 0.5	2.9 $\pm$ 0.4	1
Uric Acid (mg/dL)	5.5 $\pm$ 1.4	5.5 $\pm$ 1.5	5.4 $\pm$ 1.4	1
Total Protein (g/dL)	7.2 $\pm$ 0.5	7.2 $\pm$ 0.5	7.1 $\pm$ 0.4	<0.001
Blood Urea Nitrogen (mg/dL)	12.7 $\pm$ 5.7	12.9 $\pm$ 5.9	12.6 $\pm$ 5.5	0.0277
eGFR (mL/min/1.73m ²)	99.3 $\pm$ 27.3	98.3 $\pm$ 27.5	99.8 $\pm$ 27.7	0.073
Albumin Serum (mg/dL)	4.3 $\pm$ 0.3	4.3 $\pm$ 0.4	4.2 $\pm$ 0.3	<0.001
Lactate dehydrogenase (U/L)	133.5 $\pm$ 29.3	133.9 $\pm$ 28.7	133.3 $\pm$ 29.6	0.3979

Alb/Cr: urine albumin/ urine creatinine ratio; ln (Alb/Cr): natural logarithm urine albumin/ urine albumin ratio; eGFR: estimated glomerular filtration rate

Data from NHANES for the years 2007-2008 showed that the total prevalence of microalbuminuria was 12.6%, with a considerably higher incidence in female participants (14%) than males (11%) (p-value <0.001). This sex difference was preserved when considering participants included in the development and validation datasets (Tables 4.2, 4.3 and 4.4).

Table 4.2 Albuminuria prevalence in included subjects: whole dataset

	All		Women		Men	
Albuminuria Group	N	%	N	%	N	%
Normoalbuminuria	6,728	85.4	3,312	83.9	3,416	86.9
Microalbuminuria	989	12.6	551	14.0	438	11.2
Macroalbuminuria	161	2.0	86	2.1	75	1.9
Total	7,878	100	3,949	100	3,929	100

Pearson $\chi^2(2) = 15.2195$ Pr < 0.001

Table 4.3 Albuminuria prevalence in included subjects: model development dataset

	All		Women		Men	
Albuminuria Group	N	%	N	%	N	%
Normoalbuminuria	2,145	85.8	1,041	83.6	1,104	88.0
Microalbuminuria	291	11.6	169	13.6	122	9.7
Macroalbuminuria	45	2.6	35	2.8	29	2.3
*Total	2,500	100	1,245	100	1,255	100

*Pearson $\chi^2(2) = 9.9641$ Pr = 0.007

Table 4.4 Albuminuria prevalence in included subjects: model validation dataset

	All		Women		Men	
Albuminuria Group	N	%	N	%	N	%
Normoalbuminuria	4,583	85.2	2,271	84.0	2,312	86.5
Microalbuminuria	698	13.0	382	14.1	316	11.8
Macroalbuminuria	97	1.8	51	1.9	46	1.7
**Total	5,378	100	2,704	100	2,704	100

**Pearson $\chi^2(2) = 6.6981$ Pr = 0.035

Considering subjects included in the model development dataset (Table 4.5), it was observed that those with micro and macroalbuminuria were older than those with normal albuminuria values.

Those subjects classified as having microalbuminuria and macroalbuminuria (clinical albuminuria) are older than those with normoalbuminuria. Females more frequently presented micro and macroalbuminuria. When analyzing results for the differences between systolic and diastolic blood pressure, it was found that only systolic blood pressure was significantly different (higher) in subjects with microalbuminuria ($AUCR > 30\text{mg/g}$ and $<300\text{mg/g}$). Other variables with significantly higher values in subjects classified with microalbuminuria were: glycohemoglobin (HbA1c); serum glucose; triglycerides; creatinine in serum and urine; C-reactive protein; uric acid; blood urea nitrogen; albumin in serum and urine; and LDH.

Table 4.5 Predictor variables in the model development dataset

Variable	Normoalbuminuria (N=2,145)	Microalbuminuria (N=291)	Macroalbuminuria (N=64)	p-value
	N (%)	N (%)	N (%)	
Sex (male/female)				
<i>Male</i>	1,104 (51.5)	122 (41.9)	29 (45.3)	0.676
<i>Female</i>	1,041 (48.5)	169 (58.1)	35 (54.7)	
Race/ Ethnicity				
<i>Mexican-American</i>	411 (19.2)	60 (20.6)	12 (18.8)	0.863
<i>Other Hispanic</i>	246 (11.5)	41 (14.1)	7 (10.9)	
<i>White</i>	907 (42.3)	108 (37.1)	24 (37.5)	
<i>Black</i>	475 (22.1)	69 (23.7)	19 (29.7)	
<i>Others</i>	106 (4.9)	13 (4.5)	2 (3.1)	
Marital status				
<i>Married</i>	798 (53.4)	102 (54.6)	21 (37.5)	0.005
<i>Widowed</i>	111 (7.4)	32 (17.1)	11 (19.6)	
<i>Divorced</i>	215 (14.4)	27 (14.4)	8 (14.3)	
<i>Single</i>	370 (24.7)	19 (10.2)	15 (26.8)	
<i>Other</i>	1 (0.07)	7 (3.7)	1 (1.8)	
Health Insurance				
<i>Yes</i>	1,637 (76.3)	240 (82.5)	53 (82.8)	0.035
<i>No</i>	508 (23.7)	51 (17.5)	11 (17.2)	
Alcohol consumption				
<i>Yes</i>	1,018 (72.7)	98 (57.6)	32 (60.4)	<0.001
<i>No</i>	383 (27.3)	72 (42.4)	21 (39.6)	
Smoking				
<i>Yes</i>	692 (46.3)	88 (47.1)	36 (64.3)	0.03
<i>No</i>	803 (53.7)	99 (52.9)	20 (35.7)	

Table 4.5 Predictor variables in the model development dataset (continued)

Variable	Normoalbuminuria (N=2,145)	Microalbuminuria (N=291)	Macroalbuminuria (N=64)	p-value
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	
Age (years)	38.1 $\pm$ 22.6	42.8 $\pm$ 26.6	53.4 $\pm$ 22.1	<0.001
Body mass index (kg/m ²)	26.6 $\pm$ 7.0	26.3 $\pm$ 7.9	28.9 $\pm$ 9.4	0.029
Systolic blood pressure (mm Hg)	119.7 $\pm$ 17.7	127.4 $\pm$ 24.7	137.1 $\pm$ 27.8	<0.001
Diastolic blood pressure (mm Hg)	66.9 $\pm$ 14.1	68.5 $\pm$ 16.3	67.4 $\pm$ 18.1	0.271
Mean arterial pressure (mm Hg)	84.5 $\pm$ 13.0	88.1 $\pm$ 15.6	90.6 $\pm$ 17.9	<0.001
Hemoglobin A1c (%)	5.6 $\pm$ 0.8	6.1 $\pm$ 1.4	6.8 $\pm$ 2.0	<0.001
Glucose serum (mg/dL)	96.4 $\pm$ 29.4	114.6 $\pm$ 53.7	150.6 $\pm$ 95.8	<0.001
High density lipoprotein (mg/dL)	52.2 $\pm$ 15.3	50.7 $\pm$ 15.8	53.5 $\pm$ 19.6	0.549
Cholesterol (mg/dL)	185.1 $\pm$ 39.9	184.9 $\pm$ 44.6	197.3 $\pm$ 68.7	0.093
Triglycerides (mg/dL)	147.1 $\pm$ 115.1	173.8 $\pm$ 146.9	210.9 $\pm$ 240.1	<0.001
Serum Creatinine (mg/dL)	0.8 $\pm$ 0.2	0.9 $\pm$ 0.4	1.3 $\pm$ 1.1	<0.001
C-reactive protein (mg/dL)	3.7 $\pm$ 6.4	6.1 $\pm$ 16.9	8.3 $\pm$ 14.9	<0.005
Alb/Cr (mg/g)	8.5 $\pm$ 5.6	79.4 $\pm$ 57.2	1325.3 $\pm$ 1436.1	<0.001
ln(Alb/Cr)(mg/g)	1.9 $\pm$ 0.6	4.2 $\pm$ 0.6	6.8 $\pm$ 0.9	<0.001
Creatinine Urine (mg/dL)	126.5 $\pm$ 76.9	115.7 $\pm$ 82.4	104.2 $\pm$ 80.7	0.009
Albumin Urine (mg/dL)	10.2 $\pm$ 9.5	91.8 $\pm$ 102.0	1143.8 $\pm$ 1582.5	<0.001
Globulin (g/dL)	2.9 $\pm$ 0.4	3.1 $\pm$ 0.5	3.2 $\pm$ 0.6	<0.001
Uric Acid (mg/dL)	5.4 $\pm$ 1.4	5.7 $\pm$ 1.7	6.2 $\pm$ 2.1	<0.001
Total Protein (g/dL)	7.2 $\pm$ 0.4	7.3 $\pm$ 0.5	7.1 $\pm$ 0.8	<0.001
Blood Urea Nitrogen (mg/dL)	12.4 $\pm$ 4.9	14.7 $\pm$ 7.0	20.9 $\pm$ 14.8	<0.001
eGFR (mL/min/1.73m ²)	99.4 $\pm$ 26.4	93.0 $\pm$ 31.7	79.5 $\pm$ 37.2	0.0055
Albumin Serum (mg/dL)	4.3 $\pm$ 0.3	4.2 $\pm$ 0.4	3.9 $\pm$ 0.6	<0.001
Lactate dehydrogenase (U/L)	132.9 $\pm$ 28.0	139.6 $\pm$ 30.8	141.5 $\pm$ 36.3	<0.001

Alb/Cr: urine albumin/ urine creatinine ratio; ln(Alb/Cr): natural logarithm urine albumin/ urine albumin ratio; eGFR: estimated glomerular filtration rate

4.2. Inferential Statistics

4.2.1 Study of correlation and univariate analysis (Objective 1, hypothesis 1.1 to 1.17)

The calculation of Pearson correlation coefficient for the comparison of $\ln(\text{Alb/Cr})$ (natural logarithm urine albumin/ urine albumin ratio) and continuous variables (Table 4.6) showed a statistically significant correlation ($p\text{-value} < 0.05$) between albuminuria and age, systolic blood pressure, median arterial pressure (MAP), glycohemoglobin, glucose serum, triglycerides, creatinine in serum and urine, C-reactive protein, globulin, blood urea nitrogen, albumin in serum and urine and LDH. Conversely, HDL, total cholesterol, diastolic blood pressure, BMI, uric acid, and smoking status were not correlated with microalbuminuria. Spearman correlation was utilized to calculate the correlation between $\ln(\text{Alb/Cr})$ and the nominal variables. Female sex, presence of health insurance and alcohol consumption were significantly correlated ($p < 0.05$) with the occurrence of albuminuria. Univariate logistic regression analysis showed the same relationship between these variables and the presence of microalbuminuria.

Considering the results of these correlations, it was possible to make conclusions to the hypotheses for Objective One. These conclusions are presented in Table 4.7.

Table 4.6 Correlation and univariate logistic regression relationships of albuminuria and predictors in model development dataset

Variable	Pearson correlation (r)	Spearman correlation (rho)	Univariate logistic regression
	ln(Alb/Cr)	ln(Alb/Cr)	OR (95% CI)
Sex	0.1427*	0.2118*	1.43 (1.14-1.80)**
Race/ Ethnicity	-0.0386	-0.0679	0.98 (0.88-1.08)
Hispanic	0.0257	0.029	1.12 (0.89-1.41)
White	-0.0349	-0.0302	0.95 (0.80-1.14)
Black	-0.0064	-0.0364	1.15 (0.95-1.39)
Others	-0.011	-0.012	0.79 (0.55-1.14)
Marital status	-0.047	-0.028	0.93 (0.86-1.00)
Health Insurance	-0.0405*	-0.0649*	0.68 (0.51-0.91)*
Type 2 diabetes mellitus diagnosis	0.2004	0.2145	2.21 (1.75-2.78)*
Alcohol Consumption	-0.1423*	-0.1694*	0.53 (0.39-0.70)*
Smoking	0.0308*	0.0325	1.20 (0.92-1.59)

ln(Alb/Cr): natural logarithm albumin/creatinine; OR: Odds ratio; CI: confidence interval.

Table 4.6 Correlation and univariate logistic regression relationships of albuminuria and predictors in model development dataset (continued)

Variable	Pearson correlation (r)	Spearman correlation (rho)	Univariate logistic regression
	ln(Alb/Cr)	ln(Alb/Cr)	OR (95% CI)
Age (years)	0.1281*	0.0704*	1.01 (1.01-1.02)*
Body mass index (kg/m ²)	-0.0018	0.0535	1.00 (0.99-1.02)
Systolic blood pressure (mm Hg)	0.2092*	0.2217*	1.02 (1.02-1.03)*
Diastolic blood pressure (mm Hg)	0.0084	0.0326	1.01 (0.99-1.02)
Mean arterial pressure (mm Hg)	0.1049*	0.1475*	1.02 (1.01-1.03)*
Hemoglobin A1c (%)	0.3073*	0.2227*	1.66 (1.49-1.86)*
Glucose serum (mg/dL)	0.2579*	0.2726*	1.01 (1.00-1.02)*
High density lipoprotein (mg/dL)	0.026	0.0016	1.00 (0.99-1.01)
Cholesterol Total (mg/dL)	0.0231	0.0355	0.99 (0.99-1.00)
Triglycerides (mg/dL)	0.0986*	0.0491*	1.00 (1.00-1.01)*
Serum Creatinine (mg/dL)	0.1793*	0.1335*	4.86 (3.14-7.50)*
C-reactive protein (mg/dL)	0.1209*	0.1187*	1.17 (1.06-1.28)*
Creatinine Urine (mg/dL)	-0.1068*	-0.1357*	0.99 (0.99-1.00)*
Albumin Urine (mg/dL)	0.4673*	0.7288*	1.10 (1.09-1.11)*
Globulin (g/dL)	0.1792*	0.1708*	2.50 (1.90-3.29)*
Uric Acid (mg/dL)	0.0377	0.0483	1.19 (1.09-1.30)*
Blood Urea Nitrogen (mg/dL)	0.2567*	0.0803*	1.08 (1.06-1.11)*
eGFR (mL/min/1.73m ²)	-0.1022*	-0.0195	0.98 (0.98-0.99)*
Albumin Serum (mg/dL)	-0.1809*	-0.1058*	0.42 (0.30-0.60)*
Lactate dehydrogenase (U/L)	0.1086*	0.1049*	1.01 (1.00-1.01)*

ln(Alb/Cr): natural logarithm albumin/creatinine; OR: Odds ratio; CI: confidence interval; eGFR: estimated glomerular filtration rate; ** female

Results from correlation analysis showed that variables positive and significantly correlated with albuminuria were: sex, age, systolic blood pressure, mean arterial pressure, glycosylated hemoglobin, glucose serum, triglycerides, serum creatinine, C-reactive protein, albumin urine, globulin, blood urea nitrogen, and lactate dehydrogenase. Variables negatively and significantly correlated were: lack of health insurance, alcohol consumption, creatinine urine, estimated glomerular filtration rate, and albumin serum.

4.2.1.1 Conclusions to Hypotheses for objective one

Table 4.7 Summary of the hypotheses and conclusions for objective one

Hypothesis	Hypothesis statement	Conclusion
1.1	Systolic blood pressure is not correlated with microalbuminuria	Reject
1.2	Diastolic blood pressure is not correlated with microalbuminuria	Fail to reject
1.3	Glycosylated hemoglobin is not correlated with microalbuminuria	Reject
1.4	Age is not correlated with microalbuminuria	Reject
1.5	Male sex is not correlated with microalbuminuria	Fail to reject
1.6	Type 2 diabetes mellitus is not correlated with microalbuminuria	Reject
1.7	Lack of health care access is not correlated with microalbuminuria	Reject
1.8	Race is not correlated with microalbuminuria	Fail to reject
1.9	Body mass index is not correlated with microalbuminuria	Fail to reject
1.10	Total cholesterol level is not correlated with microalbuminuria	Fail to reject
1.11	Triglyceride level is not correlated with microalbuminuria	Reject
1.12	Fasting glucose level is not correlated with microalbuminuria	Reject
1.13	Smoking is not correlated with microalbuminuria	Fail to reject
1.14	Alcohol consumption is not correlated with microalbuminuria	Reject
1.15	C-reactive protein is not correlated with microalbuminuria	Reject
1.16	Estimated glomerular filtration rate (eGFR) is not correlated with microalbuminuria	Reject
1.17	Systolic blood pressure, HbA1c, age and male sex are not the strongest predictors of microalbuminuria	Fail to reject

4.2.2. Model development (objective 2, hypothesis 2.1-2.18)

Possible predictors for the occurrence of microalbuminuria were used in the modeling process. Specifically, those variables that showed a significant correlation in the univariate analysis were included. Variables included in the modeling process included all of the following: age; gender; BMI; systolic blood pressure; median arterial blood pressure (MAP); HbA1c; serum glucose; triglyceride; C-reactive protein; urine creatinine; globulin; blood urea nitrogen; alcohol and smoking status and lactate dehydrogenate (LDH). A multivariate logistic model was built using stepwise selection. Only those variables having a p-value less than 0.2 in the univariate analysis were included. A variable was considered to be statistically significantly associated with the presence of microalbuminuria when p-value <0.05. For analyses of MA occurrence, four models were estimated using the best subsets: (1) model 1, which included systolic blood pressure, glyated hemoglobin, blood urea nitrogen, uric acid, smoking status, fasting glucose and age as covariates. This model showed a pseudo R squared of 0.1455, an area under the receiver operator characteristic (ROC) curve of 0.70, and a Hosmer and Lemeshow test with a p-value= 0.040, meaning that this model does not fit the data well; (2) model 2, which included age, sex, systolic blood pressure, glyated hemoglobin, uric acid, total cholesterol. This model showed a pseudo R squared of 0.111, an area under the ROC curve of 0.72, and a Hosmer and Lemeshow test of 0.078; (3) model 3, which considered age, sex, systolic blood pressure, fasting glucose, triglycerides, alcohol consumption, blood urea nitrogen, uric acid, c-reactive protein, globulin, lactate dehydrogenase, smoking as covariates. The stepwise regression of this model,

including the aforementioned variables, resulted in model A, which included as final variables glucose level, C-reactive protein, systolic blood pressure, blood urea nitrogen, and alcohol consumption. The variables included in model A are explained below. This model was selected to predict microalbuminuria because of its capacity to fit the data, its area under the curve, and its potential for easy utilization in common clinical practices. (4) In the generation of model 4, the stepwise regression explained for model three was repeated considering also the presence of type 2 diabetes mellitus, health insurance, and glycated hemoglobin instead of fasting glucose. The result of this subset was model B, which included as final variables systolic blood pressure, hemoglobin A1c, blood urea nitrogen, smoking status, and alcohol consumption.

From the models created, two were selected considering the following aspects: applicability in real life situations based upon clinical criteria and novelty in variables related with the occurrence of microalbuminuria. Table 4.8.1 includes beta coefficients, odds ratios, standard errors, and Wald statistics for variables that, after backward selection, contributed to model A with a p-value <0.05 . In model A glucose serum, c-reactive protein, systolic blood pressure, blood urea nitrogen and alcohol consumption were statistically related with the occurrence of microalbuminuria. The interpretation for the odds ratio for each variable included in model A, is as follows when all other variables are held constant:

- The odds of having microalbuminuria are 1.01 times greater for each unit of increase in glucose level (mg/dL);

- The odds of having microalbuminuria are 1.35 times greater for each unit of increase in C-reactive protein (mg/dL);
- The odds of having microalbuminuria are 1.03 times greater for each unit increase in systolic blood pressure (mmHg);
- The odds of having microalbuminuria are 1.10 times greater for each unit of increase in blood urea nitrogen (mg/dL); and
- The odds of having microalbuminuria are 0.60 times lower with alcohol consumption.

Table 4.8.1 Results of the multivariate logistic regression analysis exploring patient characteristics associated with microalbuminuria: Model A.

						Pseudo R2: 0.2075	
Microalbuminuria	β - coefficient	Odds ratio	Std.Err.	Wald Statistic	p-value	Odds ratio (95% CI)	
Glucose (fasting)	0.008	1.01	0.002	3.58	0.001	1	1.01
C-reactive protein	0.297	1.35	0.13	2.14	0.033	1.02	1.77
Systolic blood pressure	0.032	1.03	0.01	5.74	0.001	1.02	1.04
Blood urea nitrogen	0.099	1.10	0.02	5.09	0.001	1.06	1.15
Alcohol consumption	-0.515	0.60	0.25	-2.02	0.044	0.36	0.99
_cons	-8.28		1.04	-9.34	0.001		

Std. Err: standard error; β - coefficient: regression coefficient; pseudo R2: pseudo R squared; _cons: model constant

Equation one is the prediction rule based on the coefficients of the multivariate logistic regression model A. In equation one glucose (fasting) was entered as mg/dL, C-reactive protein as mg/dL, systolic blood pressure as mmHg,

blood urea nitrogen as mg/dL and moderate alcohol consumption as yes (=1) or no (=0).

$$(1) P = \frac{1}{1 + e^{-[-8.28 + 0.008 (\text{glucose-fast}) + 0.297 (\text{crp}) + 0.032 (\text{SBP}) + 0.099 (\text{BUN}) - 0.515 (\text{alcohol-consumption})]}}$$

Where: crp= C-reactive protein, SBP= systolic blood pressure, BUN= blood urea nitrogen.

Table 4.8.2 includes beta coefficient, odds ratio, standard error, and Wald statistics for variables that, after backward selection, contributed to the model B with a p-value <0.05. In model B, systolic blood pressure, hemoglobin A1c, blood urea nitrogen, smoking status, alcohol consumption were statistically related with the occurrence of microalbuminuria.

Table 4.8.2 Results of the multivariate logistic regression analysis exploring patient characteristics associated with microalbuminuria: Model B.

						Pseudo R2: 0.1806	
Microalbuminuria	β - coefficient	Odds ratio	Std.Err.	Wald Statistic	p-value	Odds ratio(95% CI)	
Systolic blood pressure	0.026	1.03	0.006	4.28	0.001	1.01	1.04
Hemoglobin A1c	0.362	1.44	0.09	3.99	0.033	1.2	1.72
Blood urea nitrogen	0.098	1.10	0.02	4.25	0.001	1.05	1.15
Smoking status	0.661	1.94	0.27	2.41	0.001	1.13	3.32
Alcohol consumption	-0.607	0.54	0.28	-2.16	0.044	0.31	0.95
_cons	-8.85		1.04	-8.63	0.001		

Std. Err: standard error; β - coefficient: regression coefficient; pseudo R2: pseudo R squared; _cons: model constant

The interpretation for the odds ration for each variable included in model B, is as follows when all other variables are held constant:

- The odds of having microalbuminuria are 1.03 times greater for each unit increase in systolic blood pressure (mmHg);
- The odds of having microalbuminuria are 1.44 times greater for each unit of increase in HbA1c;
- The odds of having microalbuminuria are 1.10 times greater for each unit of increase in blood urea nitrogen (mg/dL);
- The odds of having microalbuminuria are 1.94 times greater in those smokers participants; and
- The odds of having microalbuminuria are 0.54 times lower with alcohol consumption.

Equation two is the prediction rule based on the coefficients of the multivariate logistic regression model B. Equation two uses systolic blood pressure in mmHg, HbA1c as a percentage, blood urea nitrogen in mg/dL, smoking status as (yes) or zero (no) and moderate alcohol consumption as yes (=1) or no (=0).

$$(2) P = \frac{1}{1 + e^{-[-8.85 + 0.026 (SBP) + 0.362 (HbA1c) + 0.098 (BUN) + 0.661 (Smoking) - 0.607 (alcohol-consumption)]}}$$

Where: SBP= systolic blood pressure, HbA1c= glycated hemoglobin BUN= blood urea nitrogen.

4.2.3 Risk Score

To facilitate the calculation of a patient's risk of microalbuminuria in clinical practice, a risk score analysis was performed. Model A, rather than model B, was chosen to create the risk scores because it showed the best fit and performance in both the development and validation sets. Also this model has more continuous variables that are currently measured in common clinical health care settings. Model B, on the other hand, has two categorical variables (smoking and alcohol) that are not particularly informative because they are simple yes/no responses. Alcohol consumption was also excluded from model A for the risk score calculation because its lack of specificity. For example, a subject who answers yes theoretically may only drink once a week, while another subject to answer yes may consume several drinks daily. This lack of specificity makes assigning a probability to alcohol consumption difficult. Leaving the four other model A variables to comprise the risk score assessment.

Assignment of score points was based on organizing each variable into clinically meaningful intervals to determine a reference value for each category. While each interval had a reference value, the reference value that was assigned zero points and that was used to calculate scores for the other ranges, was chosen to be the middle value of the range considered most normal in the general population. For example, SBP of 125 mmHg was chosen as SBP's zero-score-assigned reference value because it is the middle of the most normal clinical range, 120-129 mmHg.

The intervals in which the zero-score-assigned reference values lie were also assigned zero points in the score system. For example, if a subject presents with a

SPB anywhere between 120 and 129, they would be assigned zero points for SBP. For the other intervals, a middle value was also selected as that interval's (non-zero-score-assigned) reference value. If the lowest range did not include a lower bound (e.g., <120 mmHg), the middle value between the 1-percentile for that variable in the dataset and the interval's upper limit was chosen. For example, the lowest range for SBP was <120 mmHg; the 1-percentile value in the dataset for that variable was 88 mmHg. Therefore, the reference value for that interval was 104 mmHg (halfway between 88 and 120). If the highest range for a variable did not include an upper bound (e.g., >160 mmHg), the halfway point between the lower bound of that interval and the 99-percentile value in the dataset was selected as that interval's reference value. For example, the highest range for SBP was >160 mmHg; the 99-percentile value in the dataset for that variable was 180 mmHg. Therefore, the reference value chosen for that variable was 170 mmHg.

Once reference values were chosen for each interval, a beta-coefficient for each interval was calculated by subtracting the zero-score-assigned reference value from each interval's reference value. This number was then multiplied by the beta coefficient of the variable in model A. These values are listed under the beta coefficient column in Table 4.9. For example, the reference value for the clinical range of SBP of 130-139 mmHg was 135 (the middle value of that range). The interval that was assigned the 0 score (i.e., the most normal interval) was 120-129 mmHg, and the zero-score-assigned reference value was therefore 125. Thus, 125 was subtracted from 135, resulting in a value of 10, which was then multiplied by

the beta coefficient for SBP in model A, which was 0.032. The resulting value was 0.32 (listed in Table 4.9 under the beta coefficient column for the 130-139 range).

The beta coefficients for each interval were calculated using the method described above. The beta coefficient with the lowest absolute value, excluded those with a value of 0 (from the zero-score-assigned intervals), was then chosen as the number used to standardize the other beta coefficients to calculate a final score for each interval. The beta coefficient with the lowest absolute value from all the intervals of all four variables was 0.224, for the glucose fasting reference value of 117 mg/dL (in the 108-126 mg/dL interval). To standardize the other beta coefficients, the beta coefficients were divided by this 0.224 value. The closest integer or 0.5 value of this resulting quotient was then chosen to be the score for that interval. For example, the SBP range of 130-139 mmHg includes the reference value of 135. The beta coefficient for this reference value is 0.32. 0.32 was then divided by 0.224 (the beta coefficient with the lowest absolute value) to obtain a value of 1.43. This was then rounded to 1.5 (the closest integer or 0.5 value). Scores for each of the other intervals for each variable were calculated using the same method. The sum of all of these scores was 18. However, some intervals (below the most normal zero-assigned range) were assigned negative values. Therefore, a subject could obtain a final score anywhere between -5 and 18.(Mongkolsomlit 2012; Moons and others 2002; Sullivan and others 2004) Table 4.9 shows the range of values and the corresponding score for each variable.

Table 4.9 Risk score for model A

Glucose fast	Range value	Reference value	β coefficient	Score
mg/dL	70-107	89	ref	0
	108-126	117	0.224	1
	>126	216	1.016	4.5
Systolic blood pressure				
mmHg	<120	104	-0.672	-3
	120-129	125	ref	0
	130-139	135	0.32	1.5
	140-159	150	0.8	3.5
	>160	170	1.44	6.5
C-reactive protein				
mg/dL	<1	0.5	-0.446	-2
	1-2.9	2	ref	0
	>3.0	3.3	0.386	1.5
Blood urea nitrogen				
mg/dL	10-20	15	ref	0
	>21	28	1.287	5.5
Score range: 0-18				

ref: reference category; alcohol consumption was not considered in score calculation

Each patient's final risk score is determined by summing the scores for each of that patient's variables. For example, a patient with a fasting glucose value of 120, SBP of 142, C-reactive protein of 2.5, and BUN of 23, would, considering the intervals in which these values lie, be assigned the following score values: 1 for fasting glucose, 3.5 for SBP, 0 for C-reactive protein, and 5.5 for BUN. The sum of these scores is 10. This final risk score corresponds to a probability of developing microalbuminuria, (see Table 4.10). The probabilities were obtained by plugging median dataset values for each variable into the model A equation, excluding alcohol

consumption and including the equation's constant. For each score, the result of this equation was then added to that score multiplied by 0.224. For example, the final risk score of 10 (for the example patient above) would be multiplied by 0.224, resulting in 2.24. This number would then be added to the model A equation (utilizing median dataset values, excluding alcohol consumption, and including the equation constant). The resulting value (0.446) is the probability of developing microalbuminuria for a patient with a final risk score of 10. Therefore, this patient has a 44.6% probability of developing microalbuminuria. Scores below zero were not included in Table 4.10; however, it should be understood that if a patient has a final risk score less than zero, then that patient's predicted probability of developing microalbuminuria is less than 7.9%.

Table 4.10 Predicted probabilities for each score

Score	Predicted Probability
0	0.0789
1	0.0968
2	0.1182
3	0.1436
4	0.1734
5	0.2079
6	0.2472
7	0.2912
8	0.3395
9	0.3914
10	0.4458
11	0.5016
12	0.5574
13	0.6117
14	0.6634
15	0.7115
16	0.7552
17	0.7942
18	0.8284

4.2.3.1 Conclusions to Hypotheses for Objective Two

Considering the results of model development, it was possible to make conclusions to the hypotheses for objective two. These conclusions are presented in Table 4.11.1 for model A and 4.11.2 for model B.

Table 4.11.1 Summary of the hypotheses for objective two for model A and conclusions

Hypothesis	Hypothesis statement	Conclusion
2.1	Systolic blood pressure does not predict the occurrence of microalbuminuria (MA)	Reject
2.2	Diastolic blood pressure does not predict the occurrence of MA	Fail to reject
2.3	Glycosylated hemoglobin does not predict the occurrence of MA	Fail to reject
2.4	Age does not predict the occurrence of MA	Fail to reject
2.5	Male sex does not predict the occurrence of MA	Fail to reject
2.6	Type 2 diabetes does not predict the occurrence of MA	Fail to reject
2.7	Lack of health care access does not predict the occurrence of MA	Fail to reject
2.8	Race does not predict the occurrence of MA	Fail to reject
2.9	Body mass index does not predict the occurrence of MA	Fail to reject
2.10	Total cholesterol level does not predict the occurrence of MA	Fail to reject
2.11	Blood urea nitrogen does not predict the occurrence of MA	Reject
2.12	Triglyceride level does not predict the occurrence of MA	Fail to reject
2.13	Fasting glucose level does not predict the occurrence of MA	Reject
2.14	Smoking does not predict the occurrence of MA	Fail to reject
2.15	Alcohol consumption does not predict the occurrence of MA	Reject
2.16	C-reactive protein does not predict the occurrence of MA	Reject
2.17	Estimated glomerular filtration does not predict the occurrence of MA	Fail to reject
2.18	A model consisting of SBP, HbA1c, age and male sex does not predict development of MA	Fail to reject

Table 4.11.2 Summary of the hypotheses for objective two for model B and conclusions

Hypothesis	Hypothesis statement	Conclusion
2.1	Systolic blood pressure does not predict the occurrence of microalbuminuria (MA)	Reject
2.2	Diastolic blood pressure does not predict the occurrence of MA	Fail to reject
2.3	Glycosylated hemoglobin does not predict the occurrence of MA	Reject
2.4	Age does not predict the occurrence of MA	Fail to reject
2.5	Male sex does not predict the occurrence of MA	Fail to reject
2.6	Type 2 diabetes does not predict the occurrence of MA	Fail to reject
2.7	Lack of health care access does not predict the occurrence of MA	Fail to reject
2.8	Race does not predict the occurrence of MA	Fail to reject
2.9	Body mass index does not predict the occurrence of MA	Fail to reject
2.10	Total cholesterol level does not predict the occurrence of MA	Fail to reject
2.11	Blood urea nitrogen does not predict the occurrence of MA	Reject
2.12	Triglyceride level does not predict the occurrence of MA	Fail to reject
2.13	Fasting glucose level does not predict the occurrence of MA	Fail to reject
2.14	Smoking does not predict the occurrence of MA	Reject
2.15	Alcohol consumption does not predict the occurrence of MA	Reject
2.16	C-reactive protein does not predict the occurrence of MA	Fail to reject
2.17	Estimated glomerular filtration does not predict the occurrence of MA	Fail to reject
2.18	A model consisting of SBP, HbA1c, age and male sex does not predict development of MA	Fail to reject

4.3 Model Validation (objective 3, hypothesis 3.1- 3.3)

The performance of models A and B created with the development dataset were evaluated by the analysis of the area under the curve of the receiver operator characteristic curve (ROC). Figure 4.1 represents the ROC curve for model A.

Figure 4.1 Area under the curve. Model A

ROC: receiver operator curve

The area under the curve for model A was 0.78 [95%CI: 0.73 – 0.84], indicating that the discrimination of the model was high.

The performance of model B is showed in figure 4.2.

Figure 4.2 Area under the curve. Model B

ROC: receiver operator curve

The area under the curve for the model B was 0.77 [95%CI: 0.70 – 0.83], indicating that the discrimination of the model was high.

The next step was to evaluate the performance of these two models. So the internal validity was evaluated in the validation portion of the NHANES dataset (2/3 parts). To do this, the coefficients generated by each model were inputted in a vector of coefficients with the beta option and tested as follows:

Model A:

matrix input b=(0.008, 0.297, 0.032, 0.099, -0.515, -8.28)

matrix colnames b= glucose (fast), c-reactive protein, systolic blood pressure, blood urea nitrogen, alcohol consumption, _cons.

Where colnames represent the name of each variable.

The result of this validation were for model A an area under the curve of 0.74 [95%CI: 0.70 – 0.78]. Figure 4.3 shows the area under the curve for model A in both the development (blue) and validation datasets (red).

Figure 4.3 Performance of model A in both datasets.

ROC: receiver operator curve

Table 4.12 provides the results for the comparison in the performance of model A considering the subjects included in each dataset portion.

Table 4.12 Comparisons between areas under the curve for model A in development and validation datasets

Models	N	ROC area	SE	95% Confidence Interval	
Model A development dataset	684	0.78	0.029	0.73	0.84
Model A validation dataset	1,523	0.74	0.021	0.70	0.78
				Chi2 (1)= 1.61, p-value=0.2051	

ROC: receiver operator curve; SE: standard error; Chi2: chi-squared test

Differences between the area under the curve for model A in the development and validation datasets were not significant (p-value:0.2051).

Model B:

matrix input b=(0.026, 0.362, 0.098, 0.661, -0.607, -8.85)

matrix colnames b= Systolic blood pressure, hemoglobin A1c, blood urea nitrogen, smoking status, alcohol consumption, _cons.

Where colnames represent the name of each variables is the statistical package. The result of this validation were for model B an area under the curve of 0.74 [95%CI: 0.69 – 0.79]. Figure 4.4 shows the area under the curve for model B in both the development (blue) and validation datasets (red).

Figure 4.4 Performance of model B in both datasets

ROC: receiver operator curve

Table 4.13 provides the results for the comparison in the performance of model B considering the subjects included in each dataset portion.

Table 4.13 Comparisons between area under the curve for model B in development and validation datasets

Models	N	ROC area	SE	95% Confidence Interval	
Model B development dataset	637	0.77	0.033	0.70	0.83
Model B validation dataset	1,428	0.74	0.024	0.69	0.79
				Chi2 (1)= 0.46, p-value=0.4976	

ROC: receiver operator curve; SE: standard error; Chi2: chi-squared test

Differences between the area under the curve for model B in the development and validation datasets were not significant (p-value:0.4976).

4.3.1 Model diagnostic assessments

4.3.1.1 Goodness of fit test (Hosmer & Lemeshow)

Model adequacy was measured by the Hosmer & Lemeshow goodness of fit test. Model fit compared the predicted outcome probability frequency to the observed outcome probability frequency. The underlying hypothesis was that the estimated and observed frequencies agree and as a consequence a good fit yields a large p-value (>0.05). In table 4.14 the results of the goodness of fit test are shown. Model A obtained a p-value of 0.667 and model B obtained a p-value of 0.400. Both models showed good fit to the data.

Table 4.14 Goodness of fit tests (Hosmer & Lemeshow) for models A and B

Model A	
Number of observations	712
Number of covariates	712
Pearson chi2(706)	689.25
p-value	0.667
Model B	
Number of observations	660
Number of covariates	654
Pearson chi2(648)	656.45
p-value	0.400

4.3.1.2 Specification errors

Specification error tested if the model was properly specified, it should not find any additional predictors to include in either model that are statistically significant except by chance. The variable “_hat” was statistically significant for model A and B, meaning that the model is properly specified. The variable “_hatsq” was not statistically significant in both models meaning that relevant variables in both models have not been omitted. Table 4.15 shows the analysis of measuring specification errors in both models.

Table 4.15 Specification errors for models A and B

Model A						
Microalbuminuria	Coefficient	SE	Wald test	p-value	95%CI	
_hat	0.8538	0.1534	5.56	<0.001	0.55	1.15
_hatsq	-0.0626	0.0539	-1.16	0.245	-0.16	0.04
_cons	0.0061	0.1998	0.03	0.976	-0.38	0.39
Model B						
Microalbuminuria	Coefficient	SE	Wald test	p-value	95%CI	
_hat	0.998	0.2538	3.93	<0.001	0.5	1.49
_hatsq	-0.0007	0.0783	-0.01	0.993	-0.15	0.15
_cons	-0.0005	0.2497	0	0.998	-0.49	0.49

hat: linear predicted value; _hatsq: linear predictor value squared; _cons: constant;
SE: standard error; CI: confidence interval

4.3.1.3 Pseudo R squared

All of the R-squared reported here agree that model A better fits the outcome data than the model B. While these pseudo R-squared cannot be interpreted independently or compared across datasets, they are valid and useful in evaluating models predicting the same outcome on the same dataset. In other words, a pseudo R-squared statistic without context has little meaning. A pseudo R-squared only has meaning when compared to another pseudo R-squared of the same type, on the same data, predicting the same outcome. In this situation, the higher pseudo R-squared indicates which model better predicts the outcome. Table 4.16 summarized the pseudo R squared calculated and commonly reported in literature.

Table 4.16 Measures of fit, pseudo R squared by model.

Model A	Pseudo R squared	Model B	Pseudo R squared
McFadden's R2	0.208	McFadden's R2	0.181
Cox-Snell R2	0.153	Cox Snell R2	0.125
Efron's R2	0.206	Efron's R2	0.173
Pseudo R2	0.2075	Pseudo R2	0.1806

R2: r-squared

In linear regression the R-squared explains variability. The higher the R-squared, the more variability explained and the better the model. However, when analyzing models produced using a logistic regression approach, an equivalent statistic to R-squared does not exist. The estimates produced by a logistic regression are maximum likelihood estimates calculated through an iterative process. They are not calculated to minimize variance, so the OLS approach used in linear regression does not apply. However, to evaluate the goodness-of-fit of logistic

models, several pseudo R-squared statistics have been developed; the most popular (shown in Table 4.16) were calculated here. The interpretation is only valid when we compare different models created utilizing the same dataset for the same outcome. A pseudo R-squared from a model created with another dataset, even when the same outcome is studied, does not provide any meaningful information in terms of comparing goodness of fit.

4.3.1.4 Collinearity

Collinearity means that within the set of independent variables, some of the independent variables are nearly or totally predicted by the other independent variables. The variance inflation factor (VIF) quantifies the level of collinearity for each predictor, as a rule of thumb, If all of the variables are orthogonal to each other, in other words, completely uncorrelated with each other, both the tolerance and VIF are 1. A variable whose VIF is greater than 10 might worth further investigation for collinearity. Table 4.17 shows the study of collinearity for both models considering the VIF value.

Table 4.17 Study of collinearity in both models

Model A		Model B	
Variable	VIF	Variable	VIF
Glucose (fast)	1.06	Systolic blood pressure	1.05
C-reactive protein	1.02	Glycated hemoglobin	1.07
Systolic blood pressure	1.07	Blood urea nitrogen	1.05
Blood Urea nitrogen	1.06	Smoking status	1.05
Alcohol consumption	1.03	Alcohol consumption	1.06
Mean VIF	1.05	Mean VIF	1.06

VIF: variance inflation factor

Considering the results of model validation it was possible to make conclusions to the hypotheses for objective three. These conclusions are presented in Table 4.18.1 for model A and Table 4.18.2 for model B.

Table 4.18.1 Summary of the hypotheses and conclusions for objective three for model A

Hypothesis	Hypothesis statement	Conclusion
3.1	An overall measure of model accuracy using Hosmer-Lemeshow test does not show a good model fit	Reject
3.2	The predictive power of the model assessed through receiving operating characteristic curves (ROC) does not show a higher area under the curve (AUC), or a good discriminative performance of the model	Reject
3.3	ROC curves generated from the model development and model validating portions of the NHANES are not significantly different	Reject

Table 4.18.2 Summary of the hypotheses for objective three and conclusions for model B

Hypothesis	Hypothesis statement	Conclusion
3.1	An overall measure of model accuracy using Hosmer-Lemeshow test does not show a good model fit	Reject
3.2	The predictive power of the model assessed through receiving operating characteristic curves (ROC) does not show a higher area under the curve (AUC), or a good discriminative performance of the model	Reject
3.3	ROC curves generated from the model development and model validating portions of the NHANES are not significantly different	Reject

CHAPTER 5

DISCUSSION

5.1 Prevalence of microalbuminuria

Microalbuminuria (MA) is linked with adverse health consequences, especially in diabetic and hypertensive populations. The prevalence and clinical significance of MA in non-diabetic subjects is less clear. A study published in 2002 using data from NHANES (1988-1994) showed that overall MA prevalence was almost 8% but considerably lower in males (6.1%) compared with females (9.7%). Also, MA prevalence was higher for non-Hispanic blacks and Mexican-Americans than for non-Hispanic whites.

This dissertation, using data from NHANES for the years 2007-2008, found a higher overall MA prevalence of nearly 13% and a prevalence of 14% for female participants and 11% for males. Considering race and ethnicity, MA prevalence was highest for whites and Hispanics. The variables total protein, blood urea nitrogen and albumin in serum were statistically different when they were compared; however, the deviation may be too small to be of any clinical interest.

5.2 Correlation of predictors with microalbuminuria

Albuminuria has been identified as a risk factor and predictor for cardiovascular disease. Its prevalence varies by population, ranging from around 7% in the general population to 40% or more in subjects with hypertension and diabetes mellitus.(de Zeeuw and others 2006) Likelihood of developing albuminuria increases due to a series of factors such as diabetic nephropathy, renal

atherosclerosis or a dysfunction of vascular endothelia.(O'Seaghdha and others 2010)

To characterize microalbuminuria occurrence, this dissertation studied the correlation between microalbuminuria and several variables, some of which were identified in other studies. A univariate logistic regression was conducted to measure the relationship between microalbuminuria and each predictor.

Correlation showed whether two variables had some association. Hypothetically, if the number of subjects in a dataset approaches infinity, all correlations will be significant even when correlation coefficients tend to zero. Accordingly, several researchers have chosen to perform univariate regression analysis to identify relationships between variables. (Dancey and Reidy 2004; Hinkle and others 2003) In this dissertation, Pearson and Spearman correlation coefficients were calculated and, to help corroborate findings, univariate logistic regression was performed.

In this dissertation, several variables were identified as having a weak positive or negative relationship with the presence of albuminuria. The magnitude and direction of these relationships varied. The variables correlated with the presence of microalbuminuria included female sex, alcohol consumption, age, systolic blood pressure, glycated hemoglobin, glucose serum, triglycerides, creatinine in urine and serum, c-reactive protein, globulin, blood urea nitrogen, estimated glomerular filtration rate, albumin serum, and lactate dehydrogenase. The following paragraphs provide a discussion of each variable and its relationship with microalbuminuria.

5.2.1. Blood Pressure

Cardiovascular disease is one of the principal causes of global mortality, and high blood pressure is one of the most studied and prevalent risk factors for the occurrence of cardiovascular disease. (Murai and others 2014) Several studies have reported that the prevalence of microalbuminuria in subjects with high blood pressure ranges from 6% to 36%, significantly higher than in the general population. (Robles and others 2012; Schrader and others 2006)

In this dissertation, 17.1% of subjects classified as hypertensive also had microalbuminuria. In addition, systolic blood pressure rather than diastolic blood pressure was found to be related with the occurrence of microalbuminuria; this finding corroborated the results in other studies. (Chen and others 2011; Murai and others 2014; Robles and others 2012)

This dissertation used a Urine Albumin to Creatinine Ratio (UACR) threshold between 30 and 300 mg/g. This range has been utilized in several investigations and is the range proposed by the National Institute of Health. (Keane and Eknoyan 1999; Mattix and others 2002a) However, recent studies have documented that there are no data suggesting that this threshold is generalizable to subjects with hypertension. (Redon and Williams 2002) Redon and Williams have suggested that the urinary albumin level required to define increased risk in persons with hypertension is below the current UACR threshold and that this threshold should therefore be revised downward, especially considering that most hypertensive subjects have other diseases (e.g., diabetes mellitus). (Redon and Williams 2002)

The mechanisms that produce microalbuminuria in hypertensive subjects vary according to the severity of the condition and the presence of other risk factors. In uncomplicated hypertension, the hemodynamic mechanism appears to be the most probable source of microalbuminuria, but in moderate and complicated hypertension, especially if correlated with other risk factors, glomerular injury is probably the cause. (Palatini 2003) The pathophysiological mechanism explored in the HOPE (Heart Outcomes Prevention Evaluation) Study explains that the presence of microalbuminuria is a signal of increased cardiovascular risk due to altered vascular responses caused by high blood pressure. High blood pressure produces glomerular injury and general endothelial impairment. (Sleight 2000)

An increase in the plasma enzyme metalloproteinase-9, produced by endothelial cells, has been identified as responsible for the development of microalbuminuria. (Ebihara and others 1998) An association of microalbuminuria with high levels of fibrinogen and thrombin-antithrombin III suggests that microalbuminuria is a marker of generalized endothelial dysfunction. Despite these findings, the main determinant of microalbuminuria in hypertension is hemodynamic overload. That a reduction of microalbuminuria in hypertensive subjects can be attained with the use of antihypertensive drugs gives further support to this view. (Khosla and others 2006; Palatini 2003)

Oliveras et al. followed hypertensive subjects using a 24-hour register of blood pressure and found that elevated nighttime systolic blood pressure is associated independently with a higher UACR, suggesting a need for more vigorous treatment of hypertension. (Oliveras and others 2011)

In Spain, Pascual et al. studied a population with low cardiovascular risk and untreated mild hypertension at baseline. After a mean follow-up of 30 months, the researchers found that persistently high systolic blood pressure preceded the development of microalbuminuria. An upward trend in glucose values was also observed in this Spanish study. The researchers stated that the potential for developing microalbuminuria was very low for subjects who had normal blood pressure or for those who maintained normal fasting glucose values. (Pascual and others 2005)

Furthermore, there is no doubt that high blood pressure causes kidney dysfunction; therefore, an increase in systemic blood pressure may cause an augmented intraglomerular pressure and, as a consequence, may increase urinary excretion of albumin. (Khosla and others 2006)

A clinical trial revealed that antihypertensive drugs targeting the renin-angiotensin system reduce urinary excretion of albumin. (Asselbergs and others 2004) This action is attributed to the in-situ renal effect of these classes of drugs (angiotensin-converting enzyme inhibitors and AT1 receptor antagonists) that selectively decrease intraglomerular pressure. (Mogensen and others 2000) Results from the Irbesartan Diabetic Nephropathy Trial found that reducing SBP but not DBP decreases the development of microalbuminuria. (Pohl and others 2005) Additionally, the use of renin-angiotensin system-blocking medications, through 24-hour blood pressure control, reduces microalbuminuria independently of BP control. As a result, microalbuminuria reduction has been shown to be greatest with

intensive antihypertensive regimens that take advantage of this mechanism.

(Karalliedde and Viberti 2005)

Other scientists support the idea that regardless of antihypertensive mechanism, reducing high blood pressure has a benefit on urinary albumin presence. Furthermore, treatments that prevent hypertensive subjects from developing atherosclerosis could be useful for reducing microalbuminuria. (Murai and others 2014) Ramipril, felodipine, doxazosin, and atenolol have demonstrated an effect in lowering urinary albumin, but the main determinant of this reduction is simply the final reduction in blood pressure. (Erley and others 1993) Despite the notion that lowering blood pressure may decrease the occurrence of microalbuminuria, a detrimental effect of calcium channel blockers on microalbuminuria compared to other antihypertensive drug classes has been identified. (Monster and others 2002)

Holtkamp et al. showed that systolic blood pressure is related with the presence of microalbuminuria; however, reduction or control of microalbuminuria is not possible in hypertensive subjects even with pharmacological treatment, which could corroborate the multifactorial nature of this condition. (Holtkamp and others 2011)

Based on this dissertation's findings and the evidence provided by other researchers, the idea of concentrating attention on diastolic pressure more than on systolic blood pressure (Cirillo and others 2000) is inappropriate in terms of offering protection from microalbuminuria. Microalbuminuria should be studied yearly in all hypertensive subjects (Palatini 2003) to estimate the impact of

antihypertensive therapy on microalbuminuria regression/progression. The true prevalence of microalbuminuria in subjects with hypertension has not been defined and could be higher than estimated, particularly in subjects with poor blood pressure control. (Volpe 2008)

5.2.2 Sex

In this dissertation, female sex was correlated with the occurrence of microalbuminuria. Prevalence of microalbuminuria was higher in females than in males. This relationship was preserved when stratification by hypertension was considered.

Current literature also indicates a relationship between female sex and the occurrence of microalbuminuria.(Poudel and others 2012; Sánchez Azcona and Quaglia) Results published by Kweon et al. in a recent cross-sectional study found a higher overall prevalence of microalbuminuria in females than in males. Microalbuminuria was present in 13.3% of males and 18.5% of females who did not have hypertension or type 2 diabetes mellitus and in 32.7% of males and 38.9% of females who had hypertension or type 2 diabetes mellitus. However, the authors cautioned that because the study response rate was relatively low, selection bias might have limited the generalizability of the findings. Nevertheless, the age and sex distributions of subjects in the study were comparable to those of the general population. (Kweon and others 2012)

Another study that included 202 consecutively recruited subjects with stage 1 or 2 hypertension corroborates the high prevalence of microalbuminuria in

females compared to males. (Podzolkov and others 2011) Data from the 1999-2000 National Health and Nutrition Examination Survey (NHANES) indicated that 9% of adults had microalbuminuria. Older age, female sex, and non-Hispanic black ethnicity were factors associated with this prevalence. (Jacobs and others 2002; Weir 2007)

In this dissertation, when only sex and the occurrence of microalbuminuria were considered, females had a higher frequency of microalbuminuria. Hidden confounders, such as lifestyle differences or the presence of other risk factors between males and females that were not incorporated in the study because they were not available in the dataset, may account for the disparity.

Despite these consistent findings, it is also necessary to mention that some studies showed a high prevalence of microalbuminuria in males—though the subjects in these studies had other risk factors (e.g., diabetes mellitus, renal impairment, cardiovascular disease, dyslipidemia). (Gould and others 1993; Jacobs and others 2002; Verhave and others 2003)

It is common for researchers to use the albumin/creatinine ratio (AUCR) to identify subjects who have microalbuminuria, but it is necessary to mention that there are differences in urine creatinine concentrations between men and women and among ethnic groups and among people with different muscle mass. Therefore, using a standardized AUCR cut-off for microalbuminuria may underestimate microalbuminuria in subjects with higher muscle mass or overestimate it in females or those with lower muscle mass. Several authors have suggested that future studies that use the AUCR to define microalbuminuria should use sex- and race-

specific cut-offs to avoid such misclassifications. (Karalliedde and Viberti 2004; Mattix and others 2002a)

5.2.3. Age

A Spanish investigation published in 2013 found the occurrence of microalbuminuria to be close to 11% in a group of randomly selected subjects who ranged from 40 to 59 years of age. The researchers, who expected the percentage of microalbuminuria to double for each 20-year age increment, observed an age-related increase in microalbuminuria independent of sex. They also found high microalbuminuria prevalence in the elderly, particularly those with high blood pressure. (Robles and others 2013)

Results from the Third National Health and Nutrition Examination Survey (NHANES), which collected data on American individuals in 2002, revealed that the variables of older age, minority race, type 2 diabetes mellitus, high blood pressure, and elevated serum creatinine level were significantly and independently correlated with albuminuria. (Jones and others 2002a)

It has also been established by researchers that incidence of microalbuminuria increases with age in subjects with high blood pressure and other comorbidities; however, this could be due to the presence of other risk factors, such as type 2 diabetes mellitus. (Thakkar* and others 2011) There is evidence that age has a relationship with the occurrence of microalbuminuria and some authors suggest the use of an age-specific scale for AUCR; however, there is no a clear consensus on this topic. The gradual reduction in muscle mass with increased age

could reduce urinary creatinine elimination, causing older subjects to have higher AUCRs and thus false-positive results for microalbuminuria. (Prujm and others 2008) Therefore, the overestimation of microalbuminuria in older people due to the reduction in creatinine elimination cannot be excluded from any analysis.

When the dataset to generate the predictive model for this dissertation was studied, it showed that those subjects classified as positive for microalbuminuria tended to be older than those subjects with normoalbuminuria. Also, age was significantly related with MA in both the Pearson correlation and the univariate logistic regression employed in this dissertation.

5.2.4. Diabetes Mellitus

The number of subjects with type 2 diabetes mellitus is increasing globally, and the epidemic is affecting both developed and developing countries. The World Health Organization (WHO) is projecting that diabetes prevalence in the world will escalate from 4% in 1995 to 5.4% in 2025. This will be equivalent to more than 300 million people having diabetes mellitus in 2025. The combined effects of an aging population, an increase in the prevalence of obesity, and increasingly sedentary lifestyles are fueling this pandemic. (King and others 1998) Mortality attributable to diabetes mellitus accounts for around 3% of deaths in developing countries and around 8% of deaths in the U.S., Canada, and the Middle East. In people between the ages of 35 and 64, 6 to 27% of deaths were related to diabetes mellitus. (van Dieren and others 2010) Deferrari et al. found that the prevalence of microalbuminuria in subjects with diabetes mellitus appears to be related with disease duration, and

evidence suggests that the prevalence of microalbuminuria ranges from 10% to 40% in diabetic subjects. For subjects diagnosed with diabetes mellitus, the prevalence is around 15%, increasing to 25% ten years post diagnosis. (Deferrari and others 1998)

Hemoglobin A1C (HbA1c) is a measure of glycated hemoglobin that evaluates the mean blood glucose concentration over the lifetime of a red blood cell. This measure is used mostly to assess long-term glycemic control in the diabetic population. An International Expert Committee (IEC) published a suggestion that a HbA1C level greater than 6.5% be utilized to diagnose diabetes mellitus. (International Expert 2009)

Albuminuria in diabetes mellitus may be secondary to factors unrelated to diabetes mellitus, such as hypertension, congestive heart failure, prostate disease, or infection. Microalbuminuria in diabetes mellitus is a risk factor for cardiovascular disease.(Afkhami-Ardekani and others 2008) A central concern in diabetic subjects is the high possibility of progressive renal damage, which can be measured through urinary protein excretion. When microalbuminuria is persistent, it is considered to be a risk factor for diabetic nephropathy. Because of this, diabetes mellitus subjects with microalbuminuria have higher long-term mortality and an increased risk of developing macroalbuminuria.(Newman and others 2005)

Microalbuminuria is produced by the increased passage of albumin through the glomerular filtration barrier. This involves structural changes rather than modifications in glomerular pressure. The vascular endothelium acts as an interface between blood and the vascular wall and is sensitive to certain stimuli (e.g.,

mechanical, hormonal and other vasoactive substances). In response to these stimuli, the vascular endothelium releases several substances including vasodilator substances, such as nitric oxide, prostacyclin and vasoconstricting substances, such as angiotensin II and reactive oxygen species. Dysfunction of this structure contributes to several cardiovascular diseases (e.g., hypertension, chronic heart failure). (Singh and Satchell 2011)

Dysfunction of the vascular endothelium is an important variable related to microalbuminuria in both types of diabetes. Glomerular permeability is compromised and can be measured in early diabetic nephropathy. The loss of a systemic endothelial protein in diabetes implies that damage to this layer is the main reason for the occurrence of microalbuminuria. The epidemiology of microalbuminuria exposes an association with systemic endothelial dysfunction and vascular disease. (Satchell and Tooke 2008) After 5–10 years of diabetes mellitus, more than a third of subjects start to eliminate small quantities of albumin in the urine. The occurrence of microalbuminuria in diabetes mellitus is a very important predictor of progression to macro albuminuria (>300 mg/d). Once this threshold is reached, more than half of cases progress to end-stage renal disease in less than ten years. (Afkhami-Ardekani and others 2008)

In the Strong Heart Study not only blood pressure but also fasting blood glucose and diabetes mellitus were independent variables correlated with the occurrence of future microalbuminuria, suggesting that the duration of diabetes could produce glomerular damage that may be responsible for the occurrence of microalbuminuria. (Xu and others 2008) The Framingham Offspring Study also

linked diabetes mellitus at baseline with the occurrence of microalbuminuria.

(O'Seaghdha and others 2010)

In this dissertation, both HbA1c and fasting glucose were positively and significantly correlated with microalbuminuria; microalbuminuria prevalence was 17% in diabetic subjects in the database utilized to generate the model and 19% when the entire NHANES dataset was considered. Also HbA1c (%) and glucose serum (mg/dL) were significantly higher in subjects with microalbuminuria. These findings support the relationship between the presence of diabetes mellitus and the occurrence of microalbuminuria, already evidenced in other published studies.

A study by Chowta and others found a positive correlation between duration of diabetes ($r = 0.839$, $P < 0.0001$) and a higher prevalence of microalbuminuria (37%). The study also found that incidence of microalbuminuria increases with age. The investigators reported that the findings may be related with the characteristics of the subjects included (i.e., poor glycemic control overtime and treatment adherence issues) and may have been due to the small sample size, which can overestimate correlation coefficients. (Chowta and others 2009)

The presence of diabetes acts like “adding gasoline to a fire” because the glycated state of the proteins, especially albumin, acts like an antigenic molecule associated with free radicals, which produces direct damage in the epithelial cells of the glomerulus. The inability of the glomerulus to filter protein is translated into an increase in albumin excretion. This increase in excretion is associated with the presence of a high systolic blood pressure, and a late diminution in glomerular filtration leads to subsequent renal failure. (Khosla and others 2006) This

dissertation also found that 22% of subjects with MA in the full NHANES dataset also had hypertension and diabetes mellitus; this percentage decreased to 19% when only the development dataset was considered.

There is evidence that glycemic control and protein restriction have an impact on renal function; therefore, achieving blood pressure control is critical in diabetic subjects. There is also increasing evidence to suggest that the use of antihypertensive agents like an angiotensin-converting enzyme or angiotensin receptor blockers could delay the progression of renal disease and provide cardiovascular protection in subjects with type 2 diabetes mellitus and microalbuminuria. (Jerums and MacIsaac 2002) Frequent screenings for microalbuminuria in subjects with type 2 diabetes mellitus and intervention in the case of positive results are necessary to prevent the progression of renal and cardiovascular disease and should be considered as part of standard clinical practice. A discovery of microalbuminuria should trigger an intensified intervention to treat common risk factors like hyperglycemia, hypertension and dyslipidemia.

5.2.5 Triglycerides

When higher levels of circulating lipids are present, these lipids, particularly triglycerides, can become trapped by extracellular components and undergo an oxidation process that produces reactive oxygen species. Oxidative stress, with increased reactive oxygen species generation, contributes to a chronic degenerative process. The result of this process is a significant alteration of vascular function, producing other significant vascular and renal consequences. Macrophages

phagocytize oxidized lipids and change into foam cells. This mechanism may provide the conditions necessary for the progression of atherosclerosis and kidney disease.(Scheuer and others 2000)

It also is possible that some of the deleterious effects of lipids on the kidneys are mediated by other mechanisms that are responsible for the adverse lipid profile that is present in subjects who are susceptible to renal damage, such as those with concomitant diseases like diabetes mellitus. A study conducted by Franciosi et al. identified a close relationship between the risk for microalbuminuria and lipid pattern changes. The adjusted risk of having microalbuminuria almost doubles with the presence of high levels of triglycerides (≥ 150 mg/dl), which is a typical component of the metabolic syndrome. (Franciosi and others 2007)

In this dissertation, we found a weak but significant correlation between triglyceride level and the occurrence of albuminuria. This finding follows similar findings in other investigations. Trevisan et al. mentioned in a study published in 2006 that, according to epidemiological and experimental clinical studies, there is a correlation between the progression of renal disease and the corresponding occurrence of albuminuria and dyslipidemia. High triglyceride levels have been demonstrated to be independent risk factors for renal disease. (Trevisan and others 2006) To that effect, Ravid et al. demonstrated in a prospective study including more than 500 subjects with type 2 diabetes mellitus that a high level of triglycerides, even in subjects with normal renal function, was associated with a higher incidence of microalbuminuria and as a consequence more cardiovascular events. (Ravid and others 1998) Winocour et al. in his book about microalbuminuria

and clinical practice pointed out that both urinary albumin concentration and AUCR were correlated directly with serum cholesterol and triglycerides and correlated inversely with HDL levels. It was also described that in a multivariate analysis the association with triglycerides was stronger, although it was inferred that combined hyperlipidemia was always a component of microalbuminuria. (Winocour and Marshall 1998)

5.2.6 C-reactive protein (CRP)

C-reactive protein has been described as a risk marker for cardiovascular disease because it interferes in the generation of nitric oxide in the endothelial cells. This phenomenon is directly related with vasoconstriction and triggers thrombosis. It has also been reported that CRP interferes with the renin-angiotensin system, increasing the propensity for hypertension. (Chen and others 2013; Sesso and others 2003) Mijahedi et al. found a correlation between serum CRP levels and microalbuminuria in diabetic subjects and in the general population. These reports also show that low but persistent inflammation as identified by high CRP levels could play a role in the induction of microalbuminuria. (Mojahedi and others 2009)

Researchers have used a CRP level of 2 mg/L and higher to serve as a surrogate marker for cardiovascular risk. Considering this threshold, 52 percent of the adult population in the United States is at high risk. (Woloshin and Schwartz 2005) A study by Gumbinger et al. on the presence of microalbuminuria and its potential as a prognostic marker for acute stroke found that MA and CRP are linked in the inflammatory response that occurs in cardiovascular diseases and stroke. The

mechanism for this response is not clear though it seems to be related with endothelial dysfunction.(Gumbinger and others 2012)

In this dissertation, the results also indicate that CRP is positively correlated with the presence of microalbuminuria, supporting the findings of the investigations cited above. A univariate analysis indicated 17% increase in the presence of microalbuminuria for each unit of increase in level of CRP. 214 of 291 subjects with microalbuminuria had an average CRP of 6.06 mg/L, higher than the 2mg/L suggested by Woloshin et al. as the threshold for cardiovascular risk. When comparing subjects with normo and microalbuminuria, there are significant differences between both groups.

A study by Pannacciulli et al that included overweight non-diabetic women shows that there is a strong relationship ($r=0.238$, $p\text{-value}<0.05$) between AUCR and CRP concentrations after controlling for age and other metabolic variables. These two variables are considered to participate concurrently in the atherogenic process, especially in the population included.(Pannacciulli and others 2001)

5.2.7 Blood Urea Nitrogen (BUN)

BUN level, as determined by a blood test, is directly related to the metabolic function of the liver and the excretory function of the kidneys. Blood urea nitrogen is a simple clinical variable that provides useful prognostic information for decompensated cardiovascular disease. Elevated blood urea nitrogen levels probably reflect the cumulative effects of hemodynamic and neurohormonal alterations.(Aronson and others 2004) Blood urea nitrogen could have

atherosclerotic effects, as uremia has been related to an increased burden of oxidative stress. High BUN levels are related to an increase in the activity of the renin-angiotensin system, which is also related with a detrimental effect in the glomerular system, and as a consequence may contribute to microalbuminuria among other conditions. Also, activation of renin-angiotensin system produces an increase in BUN reabsorption in the renal tubules and results in an increase in its atherosclerotic effect. (Kirtane and others 2005)

A weak correlation between UACR level and BUN was found in this dissertation. 17% of subjects with microalbuminuria had elevated BUN values, and 17% of subjects with microalbuminuria also had diabetes mellitus and high levels of BUN (mg/dL). In the univariate analysis, there was an 8% increase in microalbuminuria incidence for each unit of increase in BUN (mg/dL). BUN was significantly higher in those subjects with microalbuminuria. When stratified by sex, 24% of diabetic subjects with higher levels of BUN and microalbuminuria were female versus 8% of males with the same characteristics. These findings corroborate other studies published in 2013 that investigated the common variables suspected of being related with the occurrence of microalbuminuria that found that BUN is related with MA, especially when there is stratification by diabetes mellitus.

Zakkerkish et al. also showed that in women with microalbuminuria, there are high levels of BUN and that this finding was not present in men. A multivariable regression analysis developed for the same study indicated that increased urinary albumin excretion was connected with increased blood urea nitrogen, HbA1c and longer duration of diabetes mellitus.(Zakkerkish and others 2013)

5.2.8 Uric Acid

Uric acid is a residual of purine metabolism. Elevated serum uric acid concentration could be the consequence of increased generation, decreased elimination, or ingestion of purine-rich foods or alcohol. Elevated uric acid levels can also be the result of reduced renal function and can be detrimental to renal function, inciting renin expression and inhibiting the amount of nitric oxide in the juxtaglomerular cells. (Tseng 2005)

Epidemiologic studies have reported that elevated serum uric acid concentration is associated with the occurrence of cardiovascular diseases (Fang and Alderman 2000) and reflects concomitant risk factors, such as hypertension, insulin resistance, and dyslipidemia. Hyperuricemia has also been reported to be associated with increased renal impairment. (Culleton and others 1999; Niskanen and others 2004) Other studies have shown an association between hyperuricemia and microalbuminuria in hypertensive subjects (Viazzi and others 2005) and in subjects with type 2 diabetes mellitus. (Fukui and others 2008; Tseng 2005)

In this dissertation, 26% of subjects with microalbuminuria also had elevated levels of uric acid. Those subjects with microalbuminuria had significantly higher blood uric acid levels. The univariate regression results showed a significant relationship between uric acid level and the occurrence of albuminuria. However, correlations values were not significant.

A mathematical model developed by Chen et al. using overweight adult male Chinese subjects to estimate the risk of microalbuminuria included uric acid as one of the variables that showed an independent association with microalbuminuria.

(Chen and others 2011) A study conducted to determine the relationship between serum uric acid, metabolic syndrome, and albuminuria in subjects with type 2 diabetes mellitus found that subjects with elevated serum uric acid levels had a significantly increased risk of albuminuria after adjusting for age, gender and other conventional risk factors. The authors suggested that regular measurements of uric acid levels could provide information for predicting the occurrence of albuminuria. (Kim and others 2011)

Recent research has revealed that uric acid plays an important role in the progression of nephropathy in diabetic subjects. As already discussed, diabetes mellitus is strongly associated with cardiovascular morbidity and mortality, and the presence of albuminuria in diabetic subjects suggests a more generalized vascular process that involves endothelium impairment. Serum uric acid could be a candidate for a link between endothelial disorders in this type of patient. Hyperuricemia also produces glomerular hypertension and a decrease in renal perfusion due to stimulation of the afferent arteriolar in the juxtaglomerular system. (Iseki and others 2004) In a recent review by Hovind et al, which followed more than 300 subjects for six years, researchers concluded that serum uric acid was a strong predictor of the occurrence of albuminuria in subjects with type 1 diabetes. The risk of microalbuminuria increased 1.8-fold per 1 mg/dL increase in serum uric acid. (Hovind and others 2011)

In another study, Culleton et al. concluded that blood uric acid only reflects other connected risk factors, such as hypertension, insulin resistance, or dyslipidemia. (Culleton and others 1999) Tseng et al. showed an association

between hyperuricemia and microalbuminuria in hypertensive subjects and in subjects with type 2 diabetes mellitus. (Tseng 2005) An investigation by Bonakdaran et al. that included more than 1,000 type 2 diabetic mellitus subjects indicated that high serum uric acid is associated with albuminuria and that hyperuricemia is linked with a high serum concentration of triglycerides and elevated fasting blood glucose and HbA1c levels. (Bonakdaran and others 2011)

An observational study by Nashar and Fried found an association of uric acid levels with reduced eGFR and the presence of microalbuminuria. However, the authors explain that it is difficult to interpret these studies because, as the studies are observational, it is not possible to determine the direction of the relationship, and whether the presence of chronic kidney disease is the cause of elevated uric acid levels or vice versa. However, some models and animal studies suggest that uric acid plays a causative role in the development of renal disease risk factors, such as hypertension, metabolic syndrome, and microalbuminuria. (Nashar and Fried 2012)

A prospective cohort study conducted in Taiwan by Chang that recruited participants of 40 years of age and older found that after adjusting for confounders, each 1mg/dL increase in uric acid was associated with a 1.42-fold increased risk of microalbuminuria in all subjects, being more pronounced in females. These researchers stated that hyperuricemia could induce among other outcomes endothelial dysfunction, glomerular hypertrophy, and vascular smooth muscle hypertrophy, as well as a reduction in endothelial nitric oxide production, thus triggering the renin-angiotensin system. The researchers also concluded that hyperuricemia could predict the presence of microalbuminuria in this particular

cohort and offered new evidence of a common pathogenic mechanism of hyperuricemia and microalbuminuria. Accordingly, hyperuricemia may increase cardiovascular risk through the same pathogenic mechanism as microalbuminuria.(Chang and others 2013)

5.2.9 Creatinine

Serum values of creatinine have been correlated with kidney function. Creatinine concentration is the most widely used measure of renal function in clinical medicine. Because it does not undergo any metabolism in the kidneys, creatinine is considered the perfect molecule to measure kidney filtration.(Perrone and others 1992) Furthermore, there are several formulas to calculate glomerular filtration that include other variables (e.g., sex, age, weight) in addition to creatinine. However, none of these estimates provide the true filtration value but rather an approximation. (Bostom and others 2002)

Under normal circumstances creatinine is eliminated only by the kidneys. Creatine is the creatinine precursor formed in the liver and captured by muscle mass. As a result, muscle mass is the most important reservoir of creatine and thus creatinine production. Age and sex are significant determinants of creatinine production, though protein deficiency due to poor diet can also influence creatinine production. The use of some drugs, such as corticosteroids, may decrease creatinine production, and when there is failure in kidney filtration, creatinine levels are higher, along with other nitrogenous waste products.

One study by de Jong et al. linked the occurrence of microalbuminuria to the diabetic population. In this study, ten years after diabetes diagnosis, the glomerular filtration rate of participants had increased while the albumin elimination remained normal; however, the urinary albumin excretion was classified as microalbuminuria and the glomerular filtration rate started to decrease until eventually a stage of chronic renal disease was reached. Therefore, the presence of microalbuminuria and an increase in serum creatinine preceded the development of glomerulosclerosis. (de Jong and others 2003)

A study by Scheven et al. found that progressive renal function loss measured by estimated glomerular filtration rate (eGFR) was the most important predictor, but systolic blood pressure, urinary albumin excretion, and C-reactive protein were also significantly associated with progressive albuminuria. (Scheven and others 2013a)

In this dissertation, serum creatinine in subjects in the NHANES database was positively correlated with microalbuminuria. While this finding supports the majority of microalbuminuria research, there are studies that have shown a negative correlation between serum creatinine levels and AUCR. The investigators of these studies argue that this may reflect some other mechanism of renal damage that resulted in a decrease of the glomerular filtration rate before the occurrence of microalbuminuria. (Kohler and others 2000) An investigation by Chaiken et al. indicated that subjects with hyper filtration and consequently lower levels of creatinine were more likely to have normal albumin levels than subjects without hyper filtration. (Chaiken and others 1998)

5.2.10 Estimated Glomerular Filtration Rate (eGFR)

The most known function of the kidneys is related with their capacity to filter the blood. This function is assessed measuring the glomerular filtration rate (GFR). This GFR is understood as the clearance of a substance from the plasma by the kidneys in a period of time, generally 24 hours. This method is widely used and no other method is routinely available to directly measure filtration across the glomerular membrane. (Simon and others 2011) To easily estimate GFR, researchers have developed a mathematical equation that estimates GFR. This equation purposed by Cockcroft and Gault takes into account factors such as age, sex, and ethnicity. (Cockcroft and Gault 1976)

In this dissertation, this equation was utilized to calculate an estimated glomerular filtration rate value. There was a negative correlation between this value and the occurrence of microalbuminuria, meaning the occurrence of microalbuminuria is likely to occur with the presence of abnormal renal filtration. Drury et al. published that eGFR and albuminuria separately are predictors of cardiovascular disease; there is a strong positive relationship between lower GFR and death from all causes, particularly CVD-related ones. (Drury and others 2011) Veriava published that preclinical renal disease in diabetic and hypertensive subjects can be detected by the combination of microalbuminuria and a lowered GFR. Instead of more complex methods of measuring the GFR, the utilization of this formula to obtain eGFR combined with microalbuminuria assessment could be a more efficient indicator of renal damage. (Baber and others 2009) Campos et al. found that MA was significantly associated with a moderate decrease in eGFR. In a

multiple logistic regression analysis MA was independently related with age, male gender, BMI, systolic BP, diabetes mellitus and eGFR. However, when subjects with poor eGFR (<60 ml/min/1.73 m²) were excluded, the association with MA was no longer present. (Carlos Campo 2004) Saha et al also found that there is a correlation between eGFR (<60 mL/min/1.73 m²) and microalbuminuria. They recognized that both parameters provide a complimentary benefit in the management of subjects with chronic kidney disease.(Saha and others 2014)

5.2.11 Globulins

There are two abundant groups of serum protein in the blood; one is albumin and the other is globulin. Albumin accounts for more than 50 percent of all serum proteins. A principal function of proteins is to sustain the right oncotic pressure. The majority of oncotic pressure in capillaries is generated by the presence of high quantities of albumin, which are responsible for approximately 80% of the total oncotic pressure on interstitial fluid. The other function is to bind and carry substances that are poorly soluble in water.

Globulins are divided into alpha, beta and gamma globulins. Alpha and beta globulins also transport substances, while gamma globulins are known as immunoglobulins or antibodies. Testing for protein in the urine can include all the different proteins or albumin only. No articles were found when medical literature was consulted to assess the role of globulins and their link with microalbuminuria. This shows that when researchers refer to proteinuria, they are referring mostly to the presence of albumin. In a 1989 article published by Pun et al. the authors

conclude that to sustain the oncotic pressure, there is an increase in the production of globulins when the synthesis of albumin is decreased in the liver. (Pun and others 1989)

In this dissertation, levels of globulins were positively correlated with the presence of albuminuria; however, the explanation of this finding is not straightforward, considering that globulin is part of the protein pool measured in current clinical practice. A study from Stankeviciute et al. that proposes to predict albuminuria using routine total urine protein that includes globulin assessment found that this method to predict microalbuminuria was not reliable due to the high frequency of false positive values. (Stankeviciute and others 1998) A study published in 2008 by McIntyre and Taal suggests that despite some worries about lack of precision and standardization, total urine protein measurements should be used for the assessment and monitoring of subjects with non-diabetic chronic kidney disease. The utilization of AUCR measurements should be utilized for detecting microalbuminuria and monitoring subjects with diabetic nephropathy. More research is needed to investigate the relationship between total urine protein and albumin in renal diseases and to compare the predictive significance of total urine protein with that of albumin. (McIntyre and Taal 2008)

5.2.12 Ethanol

A cross-sectional study by Klein et al. that assessed the occurrence of microalbuminuria and several risk factors studied individuals with a diagnosis of diabetes mellitus with or without insulin requirements. The authors found a higher

prevalence of microalbuminuria in subjects with insulin requirements (29%) versus those who did not use insulin (22%). The study also found that microalbuminuria was significantly associated with male sex, older age, systolic blood pressure and alcohol consumption with OR: 1.26, $p\text{-value} < 0.001$. (Klein and others 1993)

In this dissertation, a protective effect from the consumption of alcohol on the presence of microalbuminuria was found. The odds ratio for the univariate analysis was 0.52 and the correlation coefficient showed a significant and negative value. However, the alcohol variable in the dataset was a dichotomous variable without a specification about the amount of alcohol that would produce this effect. Nevertheless, this research supports some previous literature.

For example, a cross-sectional study by Orchard et al. produced results in the same direction, indicating that regular consumption of alcohol in diabetic subjects is related with fewer complications, including microalbuminuria, among others. The authors explained that this effect occurs because alcohol regulates glycemic levels and exerts vasodilator and other cardio-protective effects. (Orchard and others 1990) The same study measured alcohol consumption, cigarette smoking, and exercise and their relation with albuminuria and found that only slight albuminuria occurred in this study due to concomitant presence of hyperlipidemia. Ethanol increased plasma lipid concentrations, but the authors suggested a more rigorous study design in order to confirm this relationship. They nevertheless concluded that microalbuminuria is associated with heavy alcohol consumption and that other variables may be involved in the relationship. (Orchard and others 1990)

A recent study by White et al. added more evidence to this relationship after

studying more than six thousand subjects. This study found that those labeled as non-moderate to heavy alcohol consumers had an elevated risk of microalbuminuria independent of sex. The authors recommend moderate alcohol consumption as a measure to control MA; however, the way alcohol induces kidney damage and is related to markers of renal function in the general population requires further research. (White and others 2009)

5.2.13 Lactate dehydrogenase (LDH)

Lactate dehydrogenase is an enzyme in charge of catalyzing the oxidation or reduction of nicotinamide adenine dinucleotide (NAD). LDH is important in this aspect because it plays a critical role in glycolytic metabolism, allowing organisms to generate a temporary oxygen debt in the form of accumulated lactate to be later discharged by the re-oxidation of lactate to pyruvate when oxygen becomes available. This process regulates the amount of oxygen available in different tissues. (Markert 1984) The correlation of this enzyme with the presence of microalbuminuria has been studied particularly in subjects suffering sickle nephropathy, a common complication of sickle cell disease (SCD), characterized by glomerular hypertrophy. This condition produces changes in kidney filtration capacity (hyper-filtration), which can manifest itself clinically as proteinuria.

A study published by Gurkan et al. found that after recruiting exclusively children with SCD, microalbuminuria was positively correlated with levels of LDH and microalbuminuria presence was preserved even after separating subjects by disease stage and complexity. (Gurkan and others 2010) Day et al. found in a

retrospective study conducted in an adult cohort of SCD subjects that LDH was one of the stronger predictors of the occurrence of microalbuminuria in these subjects. (Day and others 2012)

This dissertation found a positive correlation between LDH levels and the occurrence of microalbuminuria and is the first study to report this association while considering a broad spectrum of subjects not limited only to subjects suffering SCD.

5.3 Predictive modeling and microalbuminuria

The term microalbuminuria first emerged in medical literature in the 1980s to describe the presence of albumin in urine below the level that can be detected by a standard dipstick. (Glasscock 2010) At that level, the presence of albumin is linked with the development of nephropathy, especially in diabetic subjects. After this point, numerous studies placed attention on biological risk markers, and their presence was linked with the development of cardiovascular disease. (Ruggenenti and Remuzzi 2006) The presence of microalbuminuria involves the dysfunction of the glomerular filtration barrier. This impairment is the consequence of hemodynamic and functional-structural injury of the glomerular barrier. For example, microalbuminuria in subjects with essential hypertension is the result of an increase in the passage of this protein because of an increase in pressure rather than a reabsorption failure. (Cerasola and others 2010) A threshold was proposed to classify subjects as having microalbuminuria. This threshold using albumin urine creatinine ratio (AUCR: 30–300 mg albumin/g creatinine) was used to categorize

subjects in NHANES database, though researchers are now suggesting that, like blood pressure, the concept of a threshold is inconsistent with epidemiological data. (Forman and Brenner 2006) Albuminuria is a marker of glomerular function and its filtration capacity. It is also associated with the presence of metabolic syndrome, which can explain why it is linked with the occurrence of renal and cardiovascular impairment. From this perspective, every increment of albuminuria represents a significant presence of risk. (Ruggenenti and Remuzzi 2006)

Few models have been proposed to identify subjects at risk of suffering microalbuminuria and no models have used general population data or a retrospective database. As a consequence, the models that do exist have limited application and are restricted to subjects with defined characteristics. These models were constructed using narrow inclusion criteria and include some but not all traditional demographic, cardiovascular, metabolic and other potential risk factors. As a consequence, they focus on a very specific portion of the population.

A study conducted by Mongkolsomlit et al. in Thailand explains there are two main methods to assess the presence of microalbuminuria. Both methods require at least three samples of urine, which are not easily obtained in the Thai health care system, and the utilization of radioimmunoassay to process those urine samples is expensive. The aim of the Thai study was to determine the predictive value of a combination of the risk factors for microalbuminuria. The development of this particular model involved the recruitment of diabetic subjects and measurement of subjects variables (e.g., age, gender, body mass index, glycemic control, blood pressure, dyslipidemia) through a case-control study design. The model was

generated using a logistic regression. The model identified duration of diabetes, systolic blood pressure, creatinine and alcohol consumption as independent predictors of microalbuminuria. Despite these findings, the authors identify several limitations. Those limitations included small sample size and that, because data were captured in one hospital, any findings may not be generalizable to other clinical settings (e.g., primary care units), and only Thai type 2 diabetic subjects were included. (Mongkolsomlit 2012)

Another study conducted by Scheven et al. looked to identify the predictors of albuminuria progression in the general population. This study is one of the most robust in the sense that it followed a cohort for nine years. Since it was a prospective cohort, the authors' interest was to develop a predictive model to estimate the risk of developing progressive urinary albumin excretion. Around 5,000 subjects constituted the study cohort and went to a control appointment once every three years. During each control, anthropometrical and clinical measures were performed. After the study was completed, the investigators identified age, male gender, higher BMI and albuminuria at baseline as predictors of progressive albuminuria but not systolic blood pressure or higher plasma glucose. When the authors took out of the model baseline albuminuria, then systolic blood pressure and plasma glucose were predictors of progressive albuminuria.

In the Framingham offspring cohort baseline albuminuria, age, male gender, diabetes mellitus, cholesterol and smoking were associated with the occurrence of microalbuminuria. (O'Seaghdha and others 2010)

Most studies already available in this topic investigated factors associated with the occurrence of microalbuminuria in type 1 or 2 diabetes mellitus subjects. In this dissertation two models were developed and proposed to measure the probability of microalbuminuria occurrence. These models used a cross sectional dataset with a more representative American population. Model A included glucose serum, systolic blood pressure, c-reactive protein, blood urea nitrogen and alcohol consumption. Model B included systolic blood pressure, glycohemoglobin A1c (HbA1c), blood urea nitrogen, smoking status and alcohol consumption. The most applicable model may differ from situation to situation, considering the patient characteristics; clinicians should use the model based on the most current evidence, which was developed in a population that appears as close as similar to his/her subjects as possible. For example in the case of glycohemoglobin A1c, in many clinical practices it is not a variable measured for a patient who is coming to the clinic the first time since HbA1c is used to check the long-term control of blood glucose levels in subjects with diabetes.

The pie chart described in the theoretical framework section provides an illustration of the possible classes of sufficient risk factors for MA, with some of those pieces representing unknown component causes. This sufficient component causes theory provides a proper framework to suggest in this dissertation that multiple conditions, which may interact with one another, contribute to the development of MA. In the case of models proposed in this dissertation, both captured variables that, for this dataset, represent the best approximation of variables that acting together can predict the occurrence of microalbuminuria;

however, because the nature of the data, is not possible to establish a causal relationship. Furthermore, both models do not account for 100% of microalbuminuria occurrence.

Predictive modeling has received increased consideration recently; with literature suggesting it may improve patient outcomes due to individualized disease prediction and therapeutic approaches.(Duncan 2011) The addition of different and more sophisticated risk markers can reclassify individuals who are close to the threshold for intervention. In the intermediate risk category adding supplementary laboratory tests or biomarkers to detect subclinical disease stages, may improve risk assessment results in persons with higher risk. The results from more sophisticated tests can re-classify subjects into other risk categories thus influencing clinical decisions. (De Backer 2014) However, this could restrict the development and applicability of predictive modeling because some health settings might not have these novel markers and as a result would not be able to utilize such a model. These models use advances in information technology that rely on demographics and laboratory data to improve clinical practice and individual decisions. Despite these benefits, these models are difficult to implement because of their complexity and the fact that they do not use generalized biomarkers. The lack of external validation models has delayed the widespread use of models in medicine. (Tangri and others 2011)

Beginning in the 1980's, there have been reports of albuminuria and its biological implications in the development of cardiovascular disease.(Kannel and others 1984; Kozan and others 2011; Lane 2004) These studies have usually

included other risk factors (i.e. low eGFR, high blood pressure) and albuminuria; however, albuminuria can occur separately and be independently associated with cardiovascular risk. (Charytan 2011) Microalbuminuria and its measurement has been a popular research focus in recent years, and a method that can identify the factors or predictors involved in the development of this condition may provide a better understanding of albuminuria and particularly microalbuminuria, the stage that precedes progressive renal disease, and the variables involved in its development. Basically, all cardiovascular risk factors are in some way related to the occurrence of microalbuminuria. However, the strength and variance of this relationship are unknown.

In this way, this research offers valuable information that may lead to a better understanding of the variables involved in the development of microalbuminuria, a well-established indicator of cardiovascular and renal disease.

5.3.1 Variables included in model A and B

5.3.1.1 Systolic Blood Pressure

Both predictive models included systolic blood pressure as predictor of the occurrence of microalbuminuria. A major difference between the models proposed in this investigation and other models available is mainly the data source (retrospective dataset) and the inclusion of a broad population. The effect of systolic blood pressure describes a magnitude of microalbuminuria similar to that described in models utilizing a prospective cohort.

Microalbuminuria may be more frequent in subjects with moderate to severe hypertension, and less frequent in subjects with uncomplicated hypertension. It is necessary to mention that other confounding variables could affect the protein excretion and therefore influence the occurrence of microalbuminuria in certain populations. This is one of the limitations to consider within our study, because it was not possible to obtain more exhaustive information on the population included beyond those variables available (e.g. number of years being hypertensive, pharmacological treatment if any) Exercise is known to increase albuminuria, as well as urinary tract infections in normal individuals and other factors besides magnitude of hypertension can affect the prevalence of microalbuminuria.

A study by Pascual et al assessed the development of microalbuminuria in a population with low cardiovascular risk and untreated mild hypertension. The results demonstrated that elevated systolic blood pressure precedes the occurrence of microalbuminuria, suggesting that proper control of blood pressure and maintaining fasting glucose level within a normal range could prevent the occurrence of microalbuminuria. (Pascual and others 2005) A study by Oliveras et al included a cohort of subjects with resistant hypertension aimed to determine the relationship between microalbuminuria and blood pressure measured during appointments and using 24-hour ambulatory monitors. This study found that MA is more highly associated with increased nighttime systolic blood pressure. (Oliveras and others 2011)

These studies reinforced the findings of this dissertation and support the hypothesis that systolic blood pressure is a main predictor of microalbuminuria occurrence.

Screening for microalbuminuria in hypertensive subjects is an easy procedure that can inform treatment decisions. Antihypertensive treatment, for example, can be intensified and blood pressure followed closely since subjects with hypertension are at a greater risk of worsening microalbuminuria. In mild hypertensive subjects, the development of microalbuminuria is linked to insufficient blood pressure control and to a progressive increase in glucose values. Therefore, the existence of those risk factors could permit interventions that can be started at an earlier stage of microalbuminuria.

5.3.1.2 Fasting glucose

Microalbuminuria is usually absent at diagnosis of type 1 diabetes but could be present in subjects with diagnosis of type 2 diabetes, because often the time the diagnosis of type 2 diabetes mellitus is delayed.(Mogensen 1984) Hyperglycemia in the diabetic range can cause microalbuminuria. Risk of occurrence of microalbuminuria is directly related with the level and duration of hyperglycemia, with a higher proportion in those subjects with poor control and more than five years since diagnoses of diabetes mellitus, because aggressive glycemic control prevents albuminuria onset and progression, demonstrating that hyperglycemia over the course of several years is a cause of microalbuminuria.(Meigs and others 2002)

Ritz et al. also demonstrated that intensive lowering of blood glucose in subjects with diabetes prevents the onset and progression of microalbuminuria.(Ritz and others 2010) Mechanisms of glucose-related albuminuria had been discussed in this research but involved glomerular hyperperfusion and hyperfiltration because lack of selectivity in glomerular membrane.(Satchell and Tooke 2008)

This dissertation confirms that fasting plasma glucose is associated with abnormal urinary albumin excretion and it was included in model A. Redon et al. found a positive relationship and prediction of values of fasting glucose and microalbuminuria.(Redon and others 2002) Also Meigs et al. found that elevations in baseline and 24-year time fasting glucose were strongly associated with microalbuminuria; meaning that subjects in whom type 2 diabetes was developed over 24 years were also those most probable to have microalbuminuria. This association was independent of age and elevated systolic blood pressure.(Meigs and others 2002) The mathematical predictive model created by Chen et al. considered the inclusion of 825 overweight male subjects. The predictive model created showed that microalbuminuria was significantly associated with body mass index, systolic blood pressure, fasting glucose, triglycerides, total cholesterol, blood uric acid and serum creatinine. However, other variables were not considered as predictors such as blood urea nitrogen and high and low density lipoprotein.(Chen and others 2011)

Those models used to predict albuminuria already published corroborate the results in this dissertation. However, is clear that depending on inclusion and

exclusion criteria the variables that are included in the final model varies. Though the repetition of certain variables is showing a trend about which diseases are consistently related with the presence of albuminuria.

5.3.1.3 Hemoglobin glycosylated (HbA1c)

Measuring HbA1c, physicians can quantify a patient's average glycaemia over the previous three months and assess treatment efficacy and adherence. Also it is recommended that this test should be performed routinely in all subjects with diabetes mellitus, to tailor whatever necessary intervention. Since the HbA1c test reveals mean glycaemia for the previous 3 months, its measure every 3 months is necessary to determine if a diabetes patient is on right metabolic control. (American Diabetes 2003)

The FinnDianne (Finnish Diabetic Nephropathy) study showed that HbA1c is predictive of the incidence of microalbuminuria and renal impairment. (Waden and others 2009) Another prospective study by Hsu et al. included more than 800 normoalbuminuric subjects who were followed for more than six years and showed that HbA1c levels was associated with the development of microalbuminuria. This association was preserved even considering glycohemoglobin values and their variability from two-year records. The authors emphasize the concept that sustaining glycaemia control at an early stage of diabetes diagnosis is important to avoid disease progression. Sustained levels of glucose over time cause deregulation of mediators, including an increase in reactive oxygen species that directly damage

the glomerular endothelium, thus leading to microalbuminuria. (Hsu and others 2012)

This research identified HbA1c in model B as one of the variables involved in the prediction of the occurrence of microalbuminuria. As previously mentioned, subjects with diabetes mellitus are at an increased risk of developing renal insufficiency and as result, have a high risk of mortality from cardiovascular disease.(Abbott and Bakris 2003) Screening diabetic subjects for microalbuminuria could lead to an early identification of those at higher risk. Thus, a particular treatment can then be initiated or a treatment already in progress can be augmented or tailored according with specific necessities. An advantage of early identification and intervention could avoid morbidity and mortality from renal or cardiovascular causes. The goal of treatment in subjects with microalbuminuria is to avoid the progression of the condition.

Groos et al. suggested some treatment strategies including the use of angiotensin receptor blockers or angiotensin converting enzyme inhibitors. These agents act to reverse endothelium injury and also help control high blood pressure if it exists (<130/80 mmHg). Strict glycaemia control (HbA1c<7%) is also necessary, as are smoking cessation, statins to control LDL cholesterol (<100mg/dL), and the use of acetyl salicylic acid if thrombosis prevention is required. (Gross and others 2005)

5.3.1.4 BUN (blood urea nitrogen)

The determination of serum blood urea nitrogen is the most widely used screening test to evaluate kidney function. Urea is the final product of product of protein degradation. The ammonia produced in this process is transformed to urea in the liver. This is the most important catabolic pathway for eliminating excess nitrogen from the body. Increased blood urea nitrogen (BUN) may be due to prerenal causes (cardiac deficiency, water depletion due to decreased intake and excessive loss, increased protein catabolism, and high protein diet), renal causes (glomerulonephritis, nephritis, polycystic kidney disease, and tubular necrosis) and postrenal causes (e.g., all types of obstruction of the urinary tract, such as stones, enlarged prostate gland or tumors).(Burtis and others 2006)

Eghan et al. in a cross sectional study including diabetes subjects found that BUN among other variables was a predictor of microalbuminuria occurrence.(Eghan and others 2007) Zakerdish et al. found in a study that albuminuria in diabetes subjects is correlated with blood urea nitrogen level, HbA1c and duration of diabetes mellitus; however, no model was proposed to calculate the probability of subjects to develop microalbuminuria considering these variables.(Ganji 2013)

However, considering the result from those formally predictive model proposed by Chen et al. blood urea nitrogen did not show a predictive capacity of microalbuminuria.(Chen and others 2011) The other two models proposed by Scheven et al. and by Mongkolsomlit et al. did not consider the inclusion of BUN measures in the development of their corresponding models.(Mongkolsomlit 2012; Scheven and others 2013a)

The implications of these facts and the differences between both models proposed in this dissertation are mainly due to inclusion criteria, characteristic of the subjects and its differences and availability of these variables in the corresponding studies or databases, but as was discussed before, other studies found that BUN has a relationship with the presence of microalbuminuria principally an indirect mechanism; elevated levels of BUN are related with deficiencies in renal function which has been discussed as a primary cause of microalbuminuria.

5.3.1.5 C-reactive protein (CRP)

C-reactive protein is one of the markers of sub-clinical inflammation and is thought to represent chronic low-grade inflammation of the arterial walls.(Ross 1999) C-reactive protein has been identified as a novel cardiovascular risk factor and its presence is related with injury at the endothelial level. Microalbuminuria has been identified not only as an indicator for early kidney impairment (National Kidney 2002), but also as a risk factor for progressive renal disease and cardiovascular morbidity.(Keane 2000) Studies have indicated that C-reactive protein is related with microalbuminuria in diabetic (Navarro and others 2003) and hypertensive subjects.(Stuveling and others 2004) A study conducted by Sabanayagam et al. found that when combining data from two Asian cohorts, CRP level was significantly associated with microalbuminuria and that association was independent of age, sex, alcohol consumption, smoking status, blood pressure. Furthermore, the relationship was preserved when CRP was analyzed as either a categorical or continuous variable and when the two cohorts were analyzed

separately.(Sabanayagam and others 2010)

Nakamura et al. examined the presence of CRP and microalbuminuria in a healthy Japanese population and their findings were that in a logistic model for the prediction of microalbuminuria, obesity, hypertension, diabetes, age and CRP levels were independently associated with the presence of microalbuminuria independent of gender. Their conclusion included that low grade inflammation indicated by CRP levels may be related to the presence of microalbuminuria in the general population.(Nakamura and others 2004) In Mojahedi et al. CRP was independently related to the occurrence of microalbuminuria. In type 2 diabetic subjects, microalbuminuria is accompanied by high CRP levels, suggesting activation of inflammatory pathways in progression of renal and cardiovascular atherosclerotic disease.(Mojahedi and others 2009)

In model A proposed by this dissertation, CRP is significantly related with microalbuminuria. This is the first model, considering those already published that included CRP as a predictor of microalbuminuria and a variable utilized to calculate the probability of microalbuminuria. CRP has not been included in a formal predictive model but it has been associated with the presence of microalbuminuria in many univariate studies as those presented above.

5.3.1.6 Smoking

Cigarette smoking has been linked with the occurrence of multiple health disorders predominantly affecting both respiratory and cardiovascular systems with several degrees of complexity. (Pinto-Sietsma and others 2000) Pinto et al.

included more than 7,000 non-diabetic subjects in a cross sectional study found a positive association between currently smoking and the occurrence of microalbuminuria when compared to not smoking. This association was also dose dependent in the sense that it was stronger in subjects consuming more than 20 cigarettes per day (relative risk: 1.92, CI: 1.54-2,53) and found that stopping smoking is protective toward developing albuminuria. (Pinto-Sietsma and others 2000)

The mechanism involved in the association between smoking and microalbuminuria is related with the glycation of several molecules, such as plasma proteins, lipids, and nucleic acids. These glycated products are known to produce an increase in vascular permeability and accelerate the occurrence of this phenomenon in diabetic population. (Chuahirun and others 2004; Koga and others 2009)

Smoking also induces renal damage through an increase in insulin resistance, which is related to albuminuria and a decrease in renal function. The mechanisms that create endothelial dysfunction by inducing an imbalance between the constricting and relaxing elements produced by the endothelium are more likely to generate microalbuminuria, according to findings from an Indian cross-sectional community study. (Gupta and others 2014) Further, an Australian cohort study by Tapp et al. showed that current smokers have twice the chance of developing albuminuria compared to those who have never smoked; in this same cohort the risk of albuminuria was equal for ex-smokers and those who had never smoked.(Tapp and others 2004)

Researchers suggest that smoking increases blood pressure, which leads to direct damage to glomerular vascularity. Blood pressure increases during and after each cigarette smoked, and it has been reported that hypertensive smokers are likely to develop more complicated hypertension than those subjects who have never smoked. Smoking causes a nicotine-induced stimulation of the sympathetic nervous system (i.e., adrenaline and noradrenaline release) that acutely contributes to increased arterial pressure and heart rate. (Halimi and Mimran 2000; Tapp and others 2004) The Findings from the Framingham offspring cohort associate smoking and the development of persistent microalbuminuria and include evidence of a dose-dependent relationship, while suggesting that further studies are needed to be conclusive.

On the other hand, the Prevention of Renal and Vascular End-stage Disease (PREVEND) study did not find any association between smoking and the occurrence of microalbuminuria and between smoking cessation and the regression of MA. (O'Seaghdha and others 2010) A study by Kohler et al. showed a relationship between smoking, race, and the occurrence of microalbuminuria, particularly in Caucasian diabetic smokers. (Kohler and others 2002) In the dataset utilized to create the models in this dissertation, Caucasian smokers were also more likely to have microalbuminuria; however, these findings were not significantly different compared to the other races included.

In this dissertation, the univariate analysis found a positive relationship between smoking and the occurrence of microalbuminuria (OR: 1.27, CI: 1.06-1.53). However, it was not possible to establish a dose-dependent effect toward

microalbuminuria due to the dichotomous nature of the smoking status variable (yes/no) captured in the dataset. The presence of a smoking habit and microalbuminuria was also preserved in our predictive model. To our knowledge, this is the first model that includes smoking as one of the explanatory variables related with the occurrence of microalbuminuria. This finding could recommend that some action could be taken to control this habit with the aim of preventing and regressing microalbuminuria. However, models in this dissertation included other variables that modulate the effect of smoking on MA. For this reason, it is possible that the smoking effect is modified by other variables not considered, including gender, presence of cardiovascular disease, or use of medications.

5.3.1.7 Ethanol consumption

Alcohol has been consumed throughout history either for pleasure or as a part of rituals. Several epidemiological studies have attempted to determine the effect of its consumption on the occurrence of different conditions. General findings suggest that excessive or heavy consumption is related with the occurrence of metabolic and degenerative diseases (e.g. diabetes mellitus, hepatic impairment, cardiovascular diseases, neurodegenerative disease). However, more recently evidence showed that moderate consumption has been linked with reduced cardiovascular and metabolic risk. (Poli and others 2013)

According to American definitions, a standard drink is equal to 14 gr (0.6 ounces) of pure alcohol. This amount of pure alcohol is found in 12-ounces of beer (5% alcohol content), 5-ounces of wine (12% alcohol content), 1.5-ounces or a “shot”

(40% alcohol content) of distilled spirits or liquor (e.g., gin, rum, vodka, whiskey).

(Prevention 2014)

A relationship between alcohol consumption and occurrence of cardiovascular complications, especially in diabetic population, has been studied in several cohorts. One cohort study in Europe by Beulens et al. that included type 1 diabetes mellitus subjects from 16 countries found that alcohol consumption and microalbuminuria follows a “J” shape effect. Subjects who drank moderately, defined as 30-70 g alcohol per week had a lower risk of developing vascular complications (OR: 0.36, (0.18-1.71)) for microalbuminuria; this relationship was not present when non-drinkers and heavy drinkers were assessed. The protective effect was most evident with the consumption of wine. The study also found that drinking spirits was more likely to increase the risk of microvascular complications in these subjects and consequently microalbuminuria. (Beulens and others 2008) Findings in the same direction were presented in a meta-analysis of observational studies to link the moderate consumption of alcohol with the occurrence of type 2 diabetes and nephropathy impairment. The protective effect was also observed in men and women with low or high BMI who consumed alcohol moderately but not in those who consumed alcohol heavily. (Koppes and others 2005)

More recent studies with large sample sizes have shown a protective effect of moderate alcohol consumption on loss of renal function, particularly in male elderly individuals. Proteinuria was also lower in those with light to moderate alcohol consumption than in abstainers and heavy drinkers. In this study, the authors suggest that a daily alcohol intake of 11-27 g in women and 11 to 40 g in men was

optimal. The mechanism underlying this effect is related with less hyalinization of arterioles in renal tissue; the polyphenols present in red wine in particular can act as antioxidants directly or increase the function of antioxidant enzymes already present in the tissue. (Schaeffner and Ritz 2012)

Moderate alcohol use impacts accepted risk factors for cardiovascular disease, such as HDL plasma levels and C-reactive protein. These effects may justify the association between alcohol consumption and cardiovascular risk reduction. Available data suggest that ethanol rather than other non-alcoholic components exerts the proposed protective effects, even though an additive beneficial effect of some red wine components like polyphenols is probable. Moreover, the protective effects of alcohol only appear if its consumption is moderate and regular over time. The effect of drinking only on weekends and in large quantities is not associated with any positive effect and should be strongly discouraged. (Poli and others 2013)

Both models developed in this dissertation to predict microalbuminuria were the first to include the consumption of alcohol as a protective variable to avoid the occurrence of microalbuminuria. Studies of alcohol consumption and its impact on health are methodologically difficult to execute due to the difficulty involved in measuring alcohol consumption, selecting an appropriate control group, and correctly identifying confounding factors. Assessment of alcohol consumption is usually made by questionnaire, as in the database used in our research, where frequency and amount of consumption are not clearly reported. This issue undervalues true consumption, especially in subjects with heavy consumption, who tend to report less use. (Dawson 2003)

Studies that seek to describe the effect of alcohol consumption on the occurrence of microalbuminuria in general include a small number of subjects and exclusively subjects with type 2 diabetic. These studies also do not specify the amount of alcohol consumed, and as previously mentioned, this is important because there is no linear relationship between prevention of microalbuminuria and alcohol consumption. Our study also has this limitation because it was not possible to obtain the amount of alcohol consumed. However, unlike prior studies, our study included a broad variety of patients and not only type 2 diabetes mellitus subjects. The influence of consuming alcohol and its combination with a healthy diet (e.g., Mediterranean) on the occurrence or protection of microalbuminuria remains a topic of research.

5.4 Prediction of an individual patient's risk.

Computations using either of the models proposed in this dissertation could be tedious and time consuming. Each patient's variables would have to be plugged into the equation, which may lead to computation errors. To avoid these tedious calculations, the point system simplifies the computation of the probability of microalbuminuria. This is achieved by assigning integer points to ranges of each risk factor. These ranges represent clinically meaningful intervals, and the points for each range were assigned using the strategy detailed in Sullivan et al.(Sullivan and others 2004) Using this system, a clinician can easily assign a point for a specific risk factor and then sum these points to obtain a final score. The risk estimate is then determined from a reference table that provides risk estimates for each point total.

These risk estimates convey the probability of a patient developing MA. The variables included in the risk score system correspond to those used in Model A (SBP, CRP, BUN, and fasting glucose) because this model has a better fit than Model B. This score system can be utilized for any subject for whom these variables are available. It is not necessary for a patient to already be diagnosed with a condition associated with microalbuminuria, such as diabetes, because this score is based upon the result of clinical and laboratory tests and not based upon diagnosis.

Alcohol consumption was excluded from the risk score system because it was a dichotomous variable and failed to distinguish between heavy and light consumers. Along the same line, the calculation of risk score in this research excluded the assessment of alcohol consumption because: (1) the ethanol consumption variable was dichotomous and did not provide information about the amount of alcohol consumed and the period of time of that consumption; and (2) the assessment and evaluation of drinking is largely dependent on subjects' self-reports, which introduces validity and reliability issues. The measure of alcohol consumption is considered subjective and complicated in clinical practice because it is a sensitive topic for many subjects and trends in other research show an underestimation in ethanol consumption when a patient is directly asked or an under- or over-estimation due to what is considered moderate alcohol consumption. There are several instruments available to evaluate ethanol consumption, such as The Alcohol Timeline Follow back (TLFB), which is a daily drinking estimation method that provides a detailed picture of a person's drinking over a designated time period. (Sobell and Sobell 1992) If it is necessary to consider ethanol

consumption to estimate the risk of microalbuminuria, the use of the equation provided by model A is the best approach.

Because this score provides an indication of the risk of developing microalbuminuria, it also could indicate who is most likely to benefit from prevention. For this reason, health care providers could use the microalbuminuria risk score as a valuable tool to determine who should be offered preventive treatment (e.g., medications to lower blood pressure or glycaemia levels).

The methodology proposed in this dissertation is easy to follow with statistical analyses that could be run in different and popular statistical packages. Developing and using predictive models for health care based upon the approach presented in this research should include the following steps: (1) choose a disease or condition, considering the population necessary to address, the prevalence of the disease or condition, and the cost-saving interventions available; (2) the model created should identify those patients at high risk but also those subjects at “high opportunity” to receive a tailored intervention, perhaps at an appropriate cost; and (3) plan the intervention, as result of the identification process generated by the use of predictive modeling; several general approaches are available such, as educating subjects about their condition, advocating changes in lifestyle, and increasing adherence to indicated treatment. A good workflow should be followed after identifying subjects for a specific intervention that triggers follow-up activities and deploys communications to subjects and providers as part of a cost-effective care management program.(Duncan 2011)

5.5 Comparing predictive models for microalbuminuria

In Table 5.1, there is a summary of all the predictive models for microalbuminuria available until now. There include several study designs as well as different inclusion/exclusion criteria. Despite these differences, there are some similarities with regard to the variables included. Systolic blood pressure and fasting glucose are the predictors more repeated in the existing studies. The study presented by Fox et al. investigated the use of more sophisticated biological markers, most of them not available in routine clinical practice. In all of the studies there is also no reported external validation, and all of the studies mentioned the lack of generalizability because of the population included. This dissertation, utilizing the NHANES dataset, employed a broad inclusion criteria, which helps in terms of the applicability of findings. It also proposed the utilization of ethanol consumption and smoking status as valid predictors of the occurrence of microalbuminuria. Also, the inclusion of C-reactive protein (CRP) in model A is fresh evidence of utilization of an inflammatory marker easy to measure in clinical practice. CRP is a protein that acts as a marker of inflammatory process; it is known that its presence in apparently healthy men and women is highly predictive of future risk of heart attack, stroke, sudden cardiac death, and the development of peripheral arterial disease.(Ridker 2003) For future research, it is necessary to validate our findings using a longitudinal sample of subjects.

Table 5.1 Summary of models available to predict microalbuminuria

Study	Characteristics	Variables Measured	Variables in final model	Comments
Fox, 2010	1,822 subjects. Included only non-MA subjects. Only Caucasian.	CRP; aldosterone; renin; BNP; plasminogen-activator inhibitor type 1; fibrinogen and homocysteine.	<u>Model 1</u> : Aldosterone, BNP, and homocysteine.	Excluded older, likely to be smokers, higher rates of DM and hypertension. No validation.
Chen, 2011	1,179 subjects. Only Chinese overweight male.	BMI; SBP; DBP; FPG; TG; TC; HDL-C; LDL-C; BUN; BUA; sCr and AUCR.	<u>Model 1</u> : BMI; SBP; FPG and BUN.	Validation is required. Results only applicable to Chinese obese male.
Mongkolsomlit, 2012	116 type 2 DM with MA and 116 type 2 DM without MA. Only Thai subjects.	Age; sex; occupation; education; duration of type 2 DM; Onset DM; UA; history of diabetes; SBP; DBP; BMI; history of hypertension; alcohol and smoking consumption; lipid profile; sCr; FBG and HbA1c.	<u>Model 1</u> : Duration of diabetes; LDL; Onset DM; SBP; sCr and alcohol consumption.	Recall bias (case-control study). The score created needs to be validated in other cohorts. The size of the sample should be increased.
Scheven, 2013	5,825 subjects. Median follow-up of 9.3 year. Excluded type 1 DM. Examination each three years.	Sex; Age; Smoking; history of CVD; FBG; BMI; SBP; DBP; Cholesterol; CRP; sCr; eGFR and UA.	<u>Model 1</u> : sex; age; BMI; UA. <u>Model 2</u> : Sex; Age; BMI; SBP; Hyperlipidemia; eGFR. <u>Model 3</u> : Sex; Age; BMI; UA; Change in SBP. <u>Model 4</u> : Sex; Age; BMI; SBP; Know hypertension; eGFR; Change in glucose; Change in SBP.	Model 1 was chosen to predict MA, UA is considered as predictor. Study only includes Caucasian subjects. Mostly self-reported data. No external validation.
Villa, 2014	7,878 subjects. No exclusion criteria. NHANES is a U.S. representative dataset.	Sex; race; marital status; health insurance; type 2 DM; age; BMI; SBP; DBP; HbA1c; FBG; HDL; TC; TG; sCr; CRP; uCr; UA; globulin; uric acid; BUN and LDH.	<u>Model 1</u> : FBG; SBP; BUN; CRP; Alcohol consumption. <u>Model 2</u> : SBP, HbA1c; BUN; Smoking; Alcohol consumption.	Risk score based on model A. First predictive model including alcohol consumption and smoking status. Better generalizability. Necessary future external validation.

AUCR: urine albumin/creatinine ratio; BNP: B-type natriuretic peptide; BUA: blood uric acid; BUN: blood urea nitrogen; CVD: cardiovascular disease; DBP: diastolic blood pressure; DM: diabetes mellitus; FPG: fasting plasma glucose; HbA1c: glycated hemoglobin; HDL-c: high density lipoprotein cholesterol; LDL-c: low density lipoprotein cholesterol; MA: microalbuminuria; SBP: systolic blood pressure; sCr: serum creatinine; TC: total cholesterol; TG: triglycerides; UA: urinary albumin; uCr: urinary creatinine.

5.6 Publication process

Publishing this dissertation in a peer-reviewed journal is the natural next step after the external model validation is completed. Publication plays a central role in the dissemination of the methodological and clinical evidence developed in this dissertation. The targeted journals are in the area of primary health care (e.g., American Journal of Public Health, Journal of Family Practice, and Preventing Chronic Disease). Readers of these journals seek new scientific and clinical data, in terms of new strategies to identify and prevent the occurrence of chronic conditions. This allows the findings of this dissertation (predictive models and risk score) to be directed at specific target audiences, such as family physicians and primary health care providers. Chances of publications are high because: (1) there is a lack of American evidence related with predictive modeling in microalbuminuria; (2) the methodology is simple and avoids complicated statistical approaches; and (3) this research has the opportunity to stimulate the use of retrospective datasets to identify other predictors for other conditions.

5.7 Summary of findings

- This is the first study using a large retrospective dataset to develop a model to predict the occurrence of microalbuminuria.
- Our model is based on cross-sectional data collected with epidemiological purposes to study the American population.
- In this study, seven predictors of microalbuminuria were identified. These variables were: systolic blood pressure, fasting glucose, glycosylated

hemoglobin, blood urea nitrogen, C-reactive protein, alcohol consumption, and smoking status. These data can be easily obtained by physical examinations and laboratory tests. Therefore, this approach is an inexpensive tool for health care professionals and perhaps epidemiologists to identify subjects at risk of microalbuminuria.

- The probability of microalbuminuria can be predicted by the utilization of each formula generated with model A or B or the utilization of the risk score generated for model A, considering values of fasting glucose, systolic blood pressure, C-reactive protein and blood urea nitrogen, all of which can be easily obtained in clinical practice.
- It is recommended to offer to hypertensive and diabetic subjects a primary screening for microalbuminuria by applying these models routinely. This information can assist clinicians to stratify their patients by risk, facilitate appropriate treatment of patients regarding prognosis, and possibly influence the intensity of follow-up care and medical management.

5.8 Limitations

This dissertation has some limitations that are necessary to explain. First, the dataset utilized (NHANES) was divided into thirds; one of the thirds was used to develop the model, and the other two were used to conduct a validation of the model. The data contained in NHANES were collected utilizing a weighted procedure. This procedure could not be preserved during the random splitting necessary to perform the creation of the three portions. Second, the retrospective nature of this dataset could potentially contain several misclassifications or errors

in the reporting of results; furthermore, some of the variables utilized were self-reported which may introduce some errors. Third, subjects without data on albuminuria were eliminated from the sample, which may have led to selection bias. The use of a single Alb/Cr ratio may have misclassified the albuminuria status of some subjects since this ratio is known to vary from day to day and could be increased by fever, urinary tract infection, exercise, etc. Fourth, because this study utilized observational data (i.e., cross-sectional), it is not possible to determine causality even with multivariate analyzes that concurrently control for several variables to estimate the independent effect of each variable on the presence of microalbuminuria. Fifth, for smoking and alcohol consumption, it was not possible to estimate the exact effect because data were dichotomous. Sixth, when developing predictive models, it is necessary to make an initial selection of variables to be tested. However, there is always the possibility that a variable that was not measured or selected could have had a critical relationship with the outcome.

5.9 Lessons learned and future research.

Medical data are an invaluable source for developing predictive models that can be used to improve patient evaluation and generate new evidence to inform clinical practice. Unfortunately, due to data regulations (i.e., HIPAA requirements that necessitate data use agreements that have requirements on which few people involved in the process agree), data pulls by third parties without invested interest in research, and limitations concerning data storage (i.e., requirements that limited data sets have their own computer, be in a locked room where only the investigators

have access, and be that they be maintained for a period of years) provide space, time, administrative, and logistical barriers to creating new clinical tools and new knowledge capable of helping clinicians improve patient care. Each one of these barriers and lessons learned will be discussed briefly below.

Data regulations and other administrative barriers: Some obstacles faced in this dissertation involved the multiple review cycles, Institutional Review Board requirements, agreements with medical centers, and requirements for securing protected patient data. Even though each of these issues represented an obstacle, they are also necessary and provided opportunities to learn and prepare for future research. For example, data necessary for this dissertation considered socioeconomic, biomedical, and health behavior variables. These data were stored in a variety of electronic health records. These highly fragmented clinical data across the healthcare system presented a challenge but also a significant opportunity to learn and coordinate with professionals with experience in data resources. It will be important and necessary to explore technologies to enhance data standardization and access for future investigations. However, in the future, the process could be streamlined. One of the primary difficulties was that we received different answers and advice from almost all parties involved on how to proceed with the data approval process. This difficulty presented an idea: in the future key people should be consulted and consensus reached prior to proceeding with data acquisition. The parties to include in the process include the HSPP (Human Subjects Protection Program), David Nix at the IRB for the College of Pharmacy; Dick Haney or Lewis Barbieri at ORCA (Office of Research and Contract

Analysis); Andrew Mahler at HIPAA (Health Insurance Portability and Accountability Act); and Timothy Wunz in Information Technology and Data Security at the College of Pharmacy. Consulting all of these people and reaching a consensus prior to proceeding with data acquisition could have significantly streamlined the data approval process. This process has begun.

Data pulls by third parties without invested interest in research: One of the most frustrating difficulties during this dissertation process involved a paid third party pulling data from the Arizona Urology Specialists dataset (i.e., the medical record company). Valuable time was used waiting for this third party, based in Colorado and not invested in the research outcomes, to extract the data necessary for analysis. This party also made the assurance that desired data, such as clinical variables (e.g., systolic blood pressure, diastolic blood pressure, patient weight and height) and laboratory test results (e.g., albumin, creatinine, blood glucose, glycated hemoglobin, etc.) would be pulled from the dataset. However, these data were not extracted. By the time the data was finally accessed, it was discovered to be of little use. Without these variables, it was not possible to test the model created with NHANES. Furthermore, because of the time constraints caused by difficulties in the data administration and approval process, there was not sufficient time to request that this third party make a second attempt to pull the data. In addition, when the data were pulled none of the subject identifiers matched those used in other data files. In the future, third parties could be held more accountable to provide a useful service. A request will be made to re-pull these data but it will be after the time line for this dissertation.

Limitations concerning data storage: The university's HSPP requested that the data be stored and accessed on a specific computer in a specific room used solely for this purpose. Access to this computer had to be restricted to Dr. Warholak and myself. The College of Pharmacy also had to provide a secure connection to the data. Meeting these requirements was time consuming and involved a number of parties. In addition, the necessity for physical space and resources seems wasteful and ignores the possibility of other technological solutions to data security. In the future, this process could have been streamlined if all parties involved had reached been consulted and consensus reached prior to acquiring the data. These issues are also being addressed.

The use of predictive models in healthcare will benefit from the integration of different data sources, a trend that just started in some developing countries. The more we know about patient or population characteristics, the better and more precise interventions can be. With more data, models can be fitted to a specific patient or group of subjects, which can lead to more accurate and efficient treatments that are destined to improve the overall efficacy of the healthcare system while at the same time reducing costs. This future research should consider clinical data combined with socio-cultural knowledge to better inform and create sustainable interventions, policies and health care programs.

I have also learned that while other predictive models for MA have been developed, no other has used the same methodology as this dissertation. There is an urgent need for more research studies of this type especially in developing countries to complement and expand the current knowledge about microalbuminuria and

other chronic conditions to help optimize interventions and their related costs, especially in healthcare provided by government and financed with taxpayer money.

Future Research

A number of lessons were learned from the difficulties encountered in this dissertation. Most importantly, I learned that it is always necessary to have a contingency plan in case of problems with data. For example, we were ultimately unable to use the data from the Arizona Urology Specialists dataset. Initially, the plan was to use only the Urology dataset for this dissertation. Luckily, I had anticipated problems and identified a free, readily available dataset (NHANES). In the future, it would be interesting to incorporate the data from Arizona Urology Specialists, and I intend to do so. These data could then be used to either create a new model or validate the model created using the NHANES database. Another approach, considering the publication process, will be the validation of both models using the 2011-2012 NHANES dataset.

In addition to validating the microalbuminuria model with data from other sources, there is the potential to follow a similar methodology to produce models for other diseases and conditions. Furthermore, future studies could be conducted to evaluate whether the risk score assessment represents an accurate probability of development microalbuminuria. For example, new patients could be followed after receiving a risk score assessment to determine how many developed microalbuminuria and whether the assessment was accurate.

5.10 Conclusions

Predictive models represent a valuable tool that is intended to consistently and accurately provide more evidence-based estimates to identify the occurrence of certain outcome. The intention of this dissertation was to create, validate and present a predictive model based on data obtained from a nationally representative dataset. Several patient factors can help to discriminate between subjects with favorable and unfavorable albuminuria risk probabilities. Based on the multivariate analysis of the studied variables used to identify the occurrence of microalbuminuria, systolic blood pressure, c-reactive protein, blood urea nitrogen, blood glycaemia levels (or glycated hemoglobin), ethanol consumption, and smoking have been proposed as predictors in both models developed in this dissertation. Despite their limitations, both models provide fresh evidence about variables related with albuminuria, which affirms the usefulness of predictive models. Both models have a reasonable discriminative capacity and are calibrated, but they will require external validation before they can be applied to other populations.

5.11 References

1998. Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. *Arch Intern Med* 158(17):1855-67.
- Abbott KC, Bakris GL. 2003. Cardiology patient page. Kidney failure and cardiovascular disease. *Circulation* 108(16):e114-5.
- Abougambou SSI, Abougambou AS. 2013. Prevalence and risk factors of microalbuminuria in type 2 diabetes mellitus outpatients at University Sains Malaysia Hospital. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews* 7(2):64-67.
- Actuaries TAAo. Risk Assessment and Risk Adjustment [Internet]. Available from: http://www.actuary.org/pdf/health/Risk_Adjustment_Issue_Brief_Final_5-26-10.pdf
- Afifi AA, May S, Clark V, Afifi AA. 2012. Practical multivariate analysis. Boca Raton: Taylor & Francis.
- Afkhami-Ardekani M, Modarresi M, Amirchaghmaghi E. 2008. Prevalence of microalbuminuria and its risk factors in type 2 diabetic patients. *Indian J Nephrol* 18(3):112-7.
- Agruss J, Bakris G, Gruskovak-Hernandez T, Siomos M. 2007. An Evidence-Based Practice Protocol for the Diagnosis and Management of Microalbuminuria in the Diabetic Patient. *The Journal for Nurse Practitioners* 3(3):172-177.
- American Diabetes A. 2003. Standards of medical care for patients with diabetes mellitus. *Diabetes Care* 26 Suppl 1:S33-50.
- Anderson GF, Hussey PS, Frogner BK, Waters HR. 2005. Health spending in the United States and the rest of the industrialized world. *Health Aff (Millwood)* 24(4):903-14.
- Arnlov J, Evans JC, Meigs JB, Wang TJ, Fox CS, Levy D, Benjamin EJ, D'Agostino RB, Vasan RS. 2005. Low-grade albuminuria and incidence of cardiovascular disease events in nonhypertensive and nondiabetic individuals: the Framingham Heart Study. *Circulation* 112(7):969-75.
- Aronson D, Mittleman MA, Burger AJ. 2004. Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure. *Am J Med* 116(7):466-73.
- Asselbergs FW, Diercks GF, Hillege HL, van Boven AJ, Janssen WM, Voors AA, de Zeeuw D, de Jong PE, van Veldhuisen DJ, van Gilst WH et al. . 2004. Effects of fosinopril and pravastatin on cardiovascular events in subjects with microalbuminuria. *Circulation* 110(18):2809-16.
- Austin PC, Tu JV. 2004. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol* 57(11):1138-46.
- Baber U, Mann D, Shimbo D, Woodward M, Olin JW, Muntner P. 2009. Combined role of reduced estimated glomerular filtration rate and microalbuminuria on the prevalence of peripheral arterial disease. *Am J Cardiol* 104(10):1446-51.
- Bakris GL. 1996. Microalbuminuria: prognostic implications. *Curr Opin Nephrol Hypertens* 5(3):219-23.
- Beulens JW, Kruidhof JS, Grobbee DE, Chaturvedi N, Fuller JH, Soedamah-Muthu SS. 2008. Alcohol consumption and risk of microvascular complications in type 1 diabetes patients: the EURODIAB Prospective Complications Study. *Diabetologia* 51(9):1631-8.
- Bewick V, Cheek L, Ball J. 2005. Statistics review 14: Logistic regression. *Crit Care* 9(1):112-8.
- Bland JM, Altman DG. 2000. The odds ratio.
- Bodenheimer T, Chen E, Bennett HD. 2009. Confronting the growing burden of chronic disease: can the U.S. health care workforce do the job? *Health Aff (Millwood)* 28(1):64-74.
- Bonakdaran S, Hami M, Shakeri MT. 2011. Hyperuricemia and albuminuria in patients with type 2 diabetes mellitus. *Iran J Kidney Dis* 5(1):21-4.
- Bostom AG, Kronenberg F, Ritz E. 2002. Predictive performance of renal function equations for patients with chronic kidney disease and normal serum creatinine levels. *J Am Soc Nephrol* 13(8):2140-4.
- Burtis CA, Ashwood ER, Bruns DE. 2006. Tietz textbook of clinical chemistry and molecular diagnostics. St. Louis, Mo.: Elsevier Saunders.

- Carlin BP, Xia H. 1999. Assessing environmental justice using Bayesian hierarchical models: two case studies. *J Expo Anal Environ Epidemiol* 9(1):66-78.
- Carlos Campo JS, Cecilia Roldan, Luisa Fernandez, Lucia Guerrero, Luis M. Ruilope. 2004. Correlation between estimated glomerular filtration rate and microalbuminuria in treated essential hypertensive patients. *AMERICAN JOURNAL OF HYPERTENSION* 17(5):89A-89A.
- Cerasola G, Cottone S, Mule G. 2010. The progressive pathway of microalbuminuria: from early marker of renal damage to strong cardiovascular risk predictor. *J Hypertens* 28(12):2357-69.
- Cestari A. 2013. Predictive models in Urology. *Urologia* 80(1):42-5.
- Chaiken RL, Eckert-Norton M, Bard M, Banerji MA, Palmisano J, Sachimechi I, Lebovitz HE. 1998. Hyperfiltration in African-American patients with type 2 diabetes. Cross-sectional and longitudinal data. *Diabetes Care* 21(12):2129-34.
- Chang HY, Lee PH, Lei CC, Tung CW, Hsu YC, Huang TJ, Lu LC, Lin CL. 2013. Hyperuricemia is an independent risk factor for new onset micro-albuminuria in a middle-aged and elderly population: a prospective cohort study in taiwan. *PLoS One* 8(4):e61450.
- Charytan DM. 2011. Albuminuria and cardiovascular risk: time for a new direction? *Clin J Am Soc Nephrol* 6(6):1235-7.
- Chen B, Xu W, Chen RJ, Li X, Ye B, Chen Y. 2011. A mathematical model to estimate the risk of microalbuminuria in an overweight adult male population. *J Int Med Res* 39(6):2247-55.
- Chen FQ, Wang J, Liu XB, Ma XY, Zhang XB, Huang T, Ma DW, Wang QY. 2013. Levels of inflammatory cytokines in type 2 diabetes patients with different urinary albumin excretion rates and their correlation with clinical variables. *J Diabetes Res* 2013:138969.
- Chen H, Zheng Z, Huang Y, Guo K, Lu J, Zhang L, Yu H, Bao Y, Jia W. 2012. A microalbuminuria threshold to predict the risk for the development of diabetic retinopathy in type 2 diabetes mellitus patients. *PLoS ONE* 7(5).
- Cho BH, Kim JT, Chang J, Choi KH, Nam TS, Choi SM, Lee SH, Park MS, Kim BC, Kim MK et al. . 2012. Early clinical implications of microalbuminuria in patients with acute ischaemic stroke. *Postgrad Med J* 88(1045):632-8.
- Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jr., Jones DW, Materson BJ, Oparil S, Wright JT, Jr. et al. . 2003. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 42(6):1206-52.
- Chowta NK, Pant P, Chowta MN. 2009. Microalbuminuria in diabetes mellitus: Association with age, sex, weight, and creatinine clearance. *Indian J Nephrol* 19(2):53-6.
- Chuahirun T, Simoni J, Hudson C, Seipel T, Khanna A, Harrist RB, Wesson DE. 2004. Cigarette Smoking Exacerbates and Its Cessation Ameliorates Renal Injury in Type 2 Diabetes. *The American Journal of the Medical Sciences* 327(2):57-67.
- Cirillo M, Stellato D, Laurenzi M, Panarelli W, Zanchetti A, De Santo NG. 2000. Pulse pressure and isolated systolic hypertension: association with microalbuminuria. The GUBBIO Study Collaborative Research Group. *Kidney Int* 58(3):1211-8.
- Cockcroft DW, Gault MH. 1976. Prediction of creatinine clearance from serum creatinine. *Nephron* 16(1):31-41.
- Collins GS, Moons KG. 2012. Comparing risk prediction models. *BMJ* 344:e3186.
- Culleton BF, Larson MG, Kannel WB, Levy D. 1999. Serum uric acid and risk for cardiovascular disease and death: the Framingham Heart Study. *Ann Intern Med* 131(1):7-13.
- Dancey CP, Reidy J. 2004. Statistics without maths for psychology : using SPSS for Windows. Harlow: Prentice Hall.
- Dawson DA. 2003. Methodological issues in measuring alcohol use. *Alcohol Res Health* 27(1):18-29.
- Day TG, Drasar ER, Fulford T, Sharpe CC, Thein SL. 2012. Association between hemolysis and albuminuria in adults with sickle cell anemia. *Haematologica* 97(2):201-5.
- De Backer GG. 2014. New risk markers for cardiovascular prevention. *Curr Atheroscler Rep* 16(8):427.
- de Jong PE, Curhan GC. 2006. Screening, Monitoring, and Treatment of Albuminuria: Public Health Perspectives. *Journal of the American Society of Nephrology* 17(8):2120-2126.
- de Jong PE, Hillege HL, Pinto-Sietsma SJ, de Zeeuw D. 2003. Screening for microalbuminuria in the general population: a tool to detect subjects at risk for progressive renal failure in an early phase? *Nephrol Dial Transplant* 18(1):10-3.

- de Zeeuw D, Parving HH, Henning RH. 2006. Microalbuminuria as an early marker for cardiovascular disease. *J Am Soc Nephrol* 17(8):2100-5.
- Deferrari G, Repetto M, Calvi C, Ciabattini M, Rossi C, Robaudo C. 1998. Diabetic nephropathy: from micro- to macroalbuminuria. *Nephrol Dial Transplant* 13 Suppl 8:11-5.
- Derhaschnig U, Kittler H, Woisetschlager C, Bur A, Herkner H, Hirschl MM. 2002. Microalbumin measurement alone or calculation of the albumin/creatinine ratio for the screening of hypertension patients? *Nephrol Dial Transplant* 17(1):81-5.
- Diercks GF, van Boven AJ, Hillege JL, de Jong PE, Rouleau JL, van Gilst WH. 2002. The importance of microalbuminuria as a cardiovascular risk indicator: A review. *Can J Cardiol* 18(5):525-35.
- Drury PL, Ting R, Zannino D, Ehnholm C, Flack J, Whiting M, Fassett R, Ansquer JC, Dixon P, Davis TM et al. . 2011. Estimated glomerular filtration rate and albuminuria are independent predictors of cardiovascular events and death in type 2 diabetes mellitus: the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) study. *Diabetologia* 54(1):32-43.
- Duncan I. 2011. Healthcare Risk Adjustment & Predictive Modeling. ACTEX publications, Inc.
- Ebihara I, Nakamura T, Shimada N, Koide H. 1998. Increased plasma metalloproteinase-9 concentrations precede development of microalbuminuria in non-insulin-dependent diabetes mellitus. *Am J Kidney Dis* 32(4):544-50.
- Echouffo-Tcheugui JB, Kengne AP. 2012. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med* 9(11):e1001344.
- Eghan BA, Jr., Frempong MT, Adjei-Poku M. 2007. Prevalence and predictors of microalbuminuria in patients with diabetes mellitus: a cross-sectional observational study in Kumasi, Ghana. *Ethn Dis* 17(4):726-30.
- Elyas S, Shore AC, Eastwood H, Keenan S, Stewart J, Strain WD. 2013. Microalbuminuria could improve risk prediction of stroke in patients with transient ischaemic attacks and minor strokes. *The Lancet* 381:S40.
- Erley CM, Haefele U, Heyne N, Braun N, Risler T. 1993. Microalbuminuria in essential hypertension. Reduction by different antihypertensive drugs. *Hypertension* 21(6 Pt 1):810-5.
- Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. 2001. Executive Summary of The 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). *JAMA* 285(19):2486-97.
- Fang J, Alderman MH. 2000. Serum uric acid and cardiovascular mortality the NHANES I epidemiologic follow-up study, 1971-1992. National Health and Nutrition Examination Survey. *JAMA* 283(18):2404-10.
- Flanders WD. 2006. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *Eur J Epidemiol* 21(12):847-53.
- Folsom AR. 2013. Classical and novel biomarkers for cardiovascular risk prediction in the United States. *Journal of epidemiology / Japan Epidemiological Association* 23(3):158-162.
- Forman JP, Brenner BM. 2006. 'Hypertension' and 'microalbuminuria': the bell tolls for thee. *Kidney Int* 69(1):22-8.
- Fox CS, Gona P, Larson MG, Selhub J, Tofler G, Hwang SJ, Meigs JB, Levy D, Wang TJ, Jacques PF et al. . 2010. A multi-marker approach to predict incident CKD and microalbuminuria. *J Am Soc Nephrol* 21(12):2143-9.
- Franciosi M, Pellegrini F, Sacco M, De Berardis G, Rossi MC, Strippoli GF, Belfiglio M, Tognoni G, Valentini M, Nicolucci A et al. . 2007. Identifying patients at risk for microalbuminuria via interaction of the components of the metabolic syndrome: a cross-sectional analytic study. *Clin J Am Soc Nephrol* 2(5):984-91.
- Fukui M, Tanaka M, Shiraishi E, Harusato I, Hosoda H, Asano M, Kadono M, Hasegawa G, Yoshikawa T, Nakamura N. 2008. Serum uric acid is associated with microalbuminuria and subclinical atherosclerosis in men with type 2 diabetes mellitus. *Metabolism* 57(5):625-9.
- Ganji MR. 2013. Albuminuria and management of type 2 diabetes mellitus. *Iran J Kidney Dis* 7(4):245-8.
- Gansevoort RT, Verhave JC, Hillege HL, Burgerhof JG, Bakker SJ, de Zeeuw D, de Jong PE, Group PS. 2005. The validity of screening based on spot morning urine samples to detect subjects with microalbuminuria in the general population. *Kidney Int Suppl*(94):S28-35.

- Genuth S, Alberti KG, Bennett P, Buse J, Defronzo R, Kahn R, Kitzmiller J, Knowler WC, Lebovitz H, Lernmark A et al. . 2003. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 26(11):3160-7.
- Glasscock RJ. 2010. Is the presence of microalbuminuria a relevant marker of kidney disease? *Curr Hypertens Rep* 12(5):364-8.
- Goldstein MM, Thorpe Jane H. 2010. The First Anniversary of the Health Information Technology for Economic and Clinical Health (HITECH) Act: the regulatory outlook for implementation. *Perspect Health Inf Manag* 7.
- Gorelick MH. 2006. Bias arising from missing data in predictive models. *J Clin Epidemiol* 59(10):1115-23.
- Gould MM, Mohamed-Ali V, Goubet SA, Yudkin JS, Haines AP. 1993. Microalbuminuria: associations with height and sex in non-diabetic subjects. *BMJ* 306(6872):240-242.
- Grobman WA, Stamilio DM. 2006. Methods of clinical prediction. *Am J Obstet Gynecol* 194(3):888-94.
- Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML, Zelmanovitz T. 2005. Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes Care* 28(1):164-76.
- Gruber J. 2011. The impacts of the Affordable Care Act : how reasonable are the projections? Cambridge, Mass.: National Bureau of Economic Research.
- Guillonneau B. 2007. Ceteris paribus and nomograms in medicine. *Eur Urol* 52(5):1287-9.
- Gumbinger C, Sykora M, Diedler J, Ringleb P, Rocco A. 2012. Microalbuminuria: a potential prognostic marker for acute stroke. *Nervenarzt* 83(10):1357-60.
- Gupta RK, Gupta R, Maheshwari VD, Mawliya M. 2014. Impact of smoking on microalbuminuria and urinary albumin creatinine ratio in non-diabetic normotensive smokers. *Indian J Nephrol* 24(2):92-6.
- Gurkan S, Scarponi KJ, Hotchkiss H, Savage B, Drachtman R. 2010. Lactate dehydrogenase as a predictor of kidney involvement in patients with sickle cell anemia. *Pediatr Nephrol* 25(10):2123-7.
- Halimi JM, Mimran A. 2000. Renal effects of smoking: potential mechanisms and perspectives. *Nephrology Dialysis Transplantation* 15(7):938-940.
- Harrell FE. 2001. Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. New York ; London: Springer.
- Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. 2013. Diagnostic and prognostic prediction models. *J Thromb Haemost* 11 Suppl 1:129-41.
- Hillege HL, Janssen WM, Bak AA, Diercks GF, Grobbee DE, Crijs HJ, Van Gilst WH, De Zeeuw D, De Jong PE, PREVEND Study G. 2001. Microalbuminuria is common, also in a nondiabetic, nonhypertensive population, and an independent indicator of cardiovascular risk factors and cardiovascular morbidity. *J Intern Med* 249(6):519-26.
- Hinkle DE, Wiersma W, Jurs SG. 2003. Applied statistics for the behavioral sciences. Boston: Houghton Mifflin.
- Hodgman SB. 2008. Predictive modeling & outcomes. *Prof Case Manag* 13(1):19-23; Quiz 24-5.
- Holtkamp FA, de Zeeuw D, de Graeff PA, Laverman GD, Berl T, Remuzzi G, Packham D, Lewis JB, Parving HH, Lambers Heerspink HJ. 2011. Albuminuria and blood pressure, independent targets for cardioprotective therapy in patients with diabetes and nephropathy: a post hoc analysis of the combined RENAAL and IDNT trials. *Eur Heart J* 32(12):1493-9.
- Hosmer DW, Lemeshow S. 1980. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods* 9(10):1043-1069.
- Hosmer DW, Lemeshow S, NetLibrary Inc. Applied logistic regression [Internet]. New York: Wiley; 2000. Available from:
<http://ezproxy.library.arizona.edu/login?url=http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=99027>
- Hovind P, Rossing P, Johnson RJ, Parving HH. 2011. Serum uric acid as a new player in the development of diabetic nephropathy. *J Ren Nutr* 21(1):124-7.
- Howard P. The Impact of the Affordable Care Act On the Economy, Employers, and the Workforce [Internet]. Available from:
http://edworkforce.house.gov/uploadedfiles/02.09.11_howard.pdf

- Hsu CC, Chang HY, Huang MC, Hwang SJ, Yang YC, Lee YS, Shin SJ, Tai TY. 2012. HbA1c variability is associated with microalbuminuria development in type 2 diabetes: a 7-year prospective cohort study. *Diabetologia* 55(12):3163-72.
- International Expert C. 2009. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care* 32(7):1327-34.
- Iseki K, Ikemiya Y, Inoue T, Iseki C, Kinjo K, Takishita S. 2004. Significance of hyperuricemia as a risk factor for developing ESRD in a screened cohort. *Am J Kidney Dis* 44(4):642-50.
- Jacobs DR, Jr., Murtaugh MA, Steffes M, Yu X, Roseman J, Goetz FC. 2002. Gender- and race-specific determination of albumin excretion rate using albumin-to-creatinine ratio in single, untimed urine specimens: the Coronary Artery Risk Development in Young Adults Study. *Am J Epidemiol* 155(12):1114-9.
- Jarraya F, Lakhdar R, Kammoun K, Mahfoudh H, Drissa H, Kammoun S, Abid M, Hachicha J. 2013. Microalbuminuria: A useful marker of cardiovascular disease. *Iranian Journal of Kidney Diseases* 7(3):178-186.
- Jerums G, MacIsaac RJ. 2002. Treatment of microalbuminuria in patients with type 2 diabetes mellitus. *Treat Endocrinol* 1(3):163-73.
- Jones CA, Francis ME, Eberhardt MS, Chavers B, Coresh J, Engelgau M, Kusek JW, Byrd-Holt D, Narayan KM, Herman WH et al. . 2002a. Microalbuminuria in the US population: third National Health and Nutrition Examination Survey. *Am J Kidney Dis* 39(3):445-59.
- Jones CA, Francis ME, Eberhardt MS, Chavers B, Coresh J, Engelgau M, Kusek JW, Byrd-Holt D, Narayan KM, Herman WH et al. . 2002b. Microalbuminuria in the US population: Third National Health and Nutrition Examination Survey. *American Journal of Kidney Diseases* 39(3):445-459.
- Kannel WB, Stampfer MJ, Castelli WP, Verter J. 1984. The prognostic significance of proteinuria: the Framingham study. *Am Heart J* 108(5):1347-52.
- Karalliedde J, Viberti G. 2004. Microalbuminuria and cardiovascular risk. *American Journal of Hypertension* 17(10):986-993.
- Karalliedde J, Viberti G. 2005. Hypertension and microalbuminuria: risk factors for cardiovascular disease in diabetes. *Curr Hypertens Rep* 7(1):1-2.
- Keane WF. 2000. Proteinuria: its clinical importance and role in progressive renal disease. *Am J Kidney Dis* 35(4 Suppl 1):S97-105.
- Keane WF, Eknoyan G. 1999. Proteinuria, albuminuria, risk, assessment, detection, elimination (PARADE): a position paper of the National Kidney Foundation. *Am J Kidney Dis* 33(5):1004-10.
- Keen H, Chlouverakis C. 1964. Urinary Albumin Excretion and Diabetes Mellitus. *Lancet* 2(7370):1155-6.
- Khan KN, Khan MH, Haque MZ. 2013. Correlation between microalbuminuria with complexity of coronary artery disease in diabetic patients. *Mymensingh medical journal : MMJ* 22(2):353-357.
- Khosla N, Sarafidis PA, Bakris GL. 2006. Microalbuminuria. *Clin Lab Med* 26(3):635-53, vi-vii.
- Kim ES, Kwon HS, Ahn CW, Lim DJ, Shin JA, Lee SH, Cho JH, Yoon KH, Kang MI, Cha BY et al. . 2011. Serum uric acid level is associated with metabolic syndrome and microalbuminuria in Korean patients with type 2 diabetes mellitus. *J Diabetes Complications* 25(5):309-13.
- King H, Aubert RE, Herman WH. 1998. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. *Diabetes Care* 21(9):1414-31.
- Kirtane AJ, Leder DM, Waikar SS, Chertow GM, Ray KK, Pinto DS, Karmaliotis D, Burger AJ, Murphy SA, Cannon CP et al. . 2005. Serum blood urea nitrogen as an independent marker of subsequent mortality among patients with acute coronary syndromes and normal to mildly reduced glomerular filtration rates. *J Am Coll Cardiol* 45(11):1781-6.
- Klein R, Klein BE, Moss SE. 1993. Prevalence of microalbuminuria in older-onset diabetes. *Diabetes Care* 16(10):1325-30.
- Kleinberg S, Hripcsak G. 2011. A review of causal inference for biomedical informatics. *J Biomed Inform* 44(6):1102-12.
- Koga M, Saito H, Mukai M, Otsuki M, Kasayama S. 2009. Serum glycated albumin levels are influenced by smoking status, independent of plasma glucose levels. *Acta Diabetol* 46(2):141-4.

- Kohler KA, McClellan WM, Ziemer DC, Kleinbaum DG, Boring JR. 2000. Risk factors for microalbuminuria in black americans with newly diagnosed type 2 diabetes. *Am J Kidney Dis* 36(5):903-13.
- Kohler KA, McClellan WM, Ziemer DC, Kleinbaum DG, Boring JR. 2002. Smoking and Microalbuminuria: A case-control study in African-Americans with type 2 diabetes. *Diabetes Care* 25(1):243-245.
- Konno S, Hozawa A, Miura Y, Ito S, Munakata M. 2013. High-normal diastolic blood pressure is a risk for development of microalbuminuria in the general population: the Watari study. *J Hypertens* 31(4):798-804.
- Koppes LLJ, Dekker JM, Hendriks HFJ, Bouter LM, Heine RJ. 2005. Moderate Alcohol Consumption Lowers the Risk of Type 2 Diabetes: A meta-analysis of prospective observational studies. *Diabetes Care* 28(3):719-725.
- Kozan O, Ozcan EE, Sancaktar O, Kabakci G. 2011. The prevalence of microalbuminuria and relevant cardiovascular risk factors in Turkish hypertensive patients. *Turk Kardiyol Dern Ars* 39(8):635-45.
- Kweon SS, Shin MH, Choi JS, Nam HS, Lee YH, Park KS, Lee JY, Jeong SK. 2012. Prevalence of albuminuria and associated cardiovascular risk factors: a community cohort in Namwon City, Korea. *Diabetes Res Clin Pract* 97(3):492-8.
- Lane JT. 2004. Microalbuminuria as a marker of cardiovascular and renal risk in type 2 diabetes mellitus: a temporal perspective. *Am J Physiol Renal Physiol* 286(3):F442-50.
- Lewis G, Curry, N., Bardsley, M. Choosing a predictive risk model: a guide for commissioners in England [Internet]. Available from: http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/choosing_predictiv_e_risk_model_guide_for_commissioners_nov11.pdf
- Markert CL. 1984. Lactate dehydrogenase. Biochemistry and function of lactate dehydrogenase. *Cell Biochem Funct* 2(3):131-4.
- Mattix HJ, Hsu C-y, Shaykevich S, Curhan G. 2002a. Use of the Albumin/Creatinine Ratio to Detect Microalbuminuria: Implications of Sex and Race. *Journal of the American Society of Nephrology* 13(4):1034-1039.
- Mattix HJ, Hsu CY, Shaykevich S, Curhan G. 2002b. Use of the albumin/creatinine ratio to detect microalbuminuria: implications of sex and race. *J Am Soc Nephrol* 13(4):1034-9.
- McIntyre NJ, Taal MW. 2008. How to measure proteinuria? *Curr Opin Nephrol Hypertens* 17(6):600-3.
- Meigs JB, D'Agostino RB, Sr., Nathan DM, Rifai N, Wilson PW, Framingham Offspring S. 2002. Longitudinal association of glycemia and microalbuminuria: the Framingham Offspring Study. *Diabetes Care* 25(6):977-83.
- Metcalf PA, Scragg RK. 1994. Epidemiology of microalbuminuria in the general population. *J Diabetes Complications* 8(3):157-63.
- Michael S. Cousins LMS, and John A. Bander. 2004. An Introduction to Predictive Modeling for Disease Management Risk Stratification. *Disease Management* 5(3):157-167.
- Mogensen CE. 1984. Microalbuminuria predicts clinical proteinuria and early mortality in maturity-onset diabetes. *N Engl J Med* 310(6):356-60.
- Mogensen CE, Neldam S, Tikkanen I, Oren S, Viskoper R, Watts RW, Cooper ME. 2000. Randomised controlled trial of dual blockade of renin-angiotensin system in patients with hypertension, microalbuminuria, and non-insulin dependent diabetes: the candesartan and lisinopril microalbuminuria (CALM) study. *BMJ* 321(7274):1440-4.
- Mojahedi MJ, Bonakdaran S, Hami M, Sheikhan MR, Shakeri MT, Aiatollahi H. 2009. Elevated serum C-reactive protein level and microalbuminuria in patients with type 2 diabetes mellitus. *Iran J Kidney Dis* 3(1):12-6.
- Mongkolsomlit S, Patumanond J, Tawichasril C, Komoltri C, Rawdaree P. 2012. Meta-regression of risk factors for microalbuminuria in type 2 diabetes. *Southeast Asian J Trop Med Public Health* 43(2):455-66.
- Mongkolsomlit SR, P. Komoltri, C. Tawichasri, C. Patumanond, J. 2012. The development and validation of a risk score for predicting microalbuminuria in type 2 diabetic patients. *Journal of Diabetes Mellitus* 2(2):227-233.

- Monster TB, Janssen WM, de Jong PE, de Jong-van den Berg LT, Group PS. 2002. The impact of antihypertensive drug groups on urinary albumin excretion in a non-diabetic population. *Br J Clin Pharmacol* 53(1):31-6.
- Moons KG, Harrell FE, Steyerberg EW. 2002. Should scoring rules be based on odds ratios or regression coefficients? *J Clin Epidemiol* 55(10):1054-5.
- Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. 2012a. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 98(9):691-8.
- Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. 2012b. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 98(9):683-90.
- Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. 2009. Prognosis and prognostic research: what, why, and how? *BMJ* 338:b375.
- Moses H, 3rd, Matheson DH, Dorsey ER, George BP, Sadoff D, Yoshimura S. 2013. The anatomy of health care in the United States. *JAMA* 310(18):1947-63.
- Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, Martin BC, Stang P. 2003. A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. *Value Health* 6(2):90-7.
- Mukaka MM. 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 24(3):69-71.
- Murai S, Tanaka S, Dohi Y, Kimura G, Ohte N. 2014. The prevalence, characteristics, and clinical significance of abnormal albuminuria in patients with hypertension. *Sci Rep* 4:3884.
- Nakamura M, Onoda T, Itai K, Ohsawa M, Satou K, Sakai T, Segawa T, Sasaki J, Tonari Y, Hiramori K et al. . 2004. Association between serum C-reactive protein levels and microalbuminuria: a population-based cross-sectional study in northern Iwate, Japan. *Intern Med* 43(10):919-25.
- Nashar K, Fried LF. 2012. Hyperuricemia and the progression of chronic kidney disease: is uric acid a marker or an independent risk factor? *Adv Chronic Kidney Dis* 19(6):386-91.
- Nass SJ, Levit LA, Gostin LO, Institute of Medicine (U.S.). Committee on Health Research and the Privacy of Health Information the HIPAA Privacy Rule. 2009. Beyond the HIPAA privacy rule : enhancing privacy, improving health through research. Washington, D.C.: National Academies Press.
- National Health and Nutrition Examination Survey Data. In: (NCHS) CfDCaPCNCfHS, editor.: Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
- National Kidney F. 2002. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis* 39(2 Suppl 1):S1-266.
- Navarro JF, Mora C, Maca M, Garca J. 2003. Inflammatory parameters are independently associated with urinary albumin in type 2 diabetes mellitus. *Am J Kidney Dis* 42(1):53-61.
- Newman DJ, Mattock MB, Dawney AB, Kerry S, McGuire A, Yaqoob M, Hitman GA, Hawke C. 2005. Systematic review on urine albumin testing for early detection of diabetic complications. *Health Technol Assess* 9(30):iii-vi, xiii-163.
- Niskanen LK, Laaksonen DE, Nyyssonen K, Alfthan G, Lakka HM, Lakka TA, Salonen JT. 2004. Uric acid level as a risk factor for cardiovascular and all-cause mortality in middle-aged men: a prospective cohort study. *Arch Intern Med* 164(14):1546-51.
- O'Seaghdha CM, Hwang SJ, Upadhyay A, Meigs JB, Fox CS. 2010. Predictors of incident albuminuria in the Framingham Offspring cohort. *Am J Kidney Dis* 56(5):852-60.
- OCS. Predictive Modeling—To Improve Outcomes in Patients and Home Care [Internet]. OCS-Leading Healthcare Insight; [cited 2013 12/04/2013]. Available from: http://www.ocshomecare.com/OCSHomeCare/media/OCSHomeCare/Whitepapers/RR_Predictive_Modeling_08-003.pdf
- OECD. Health at a Glance 2013. OECD Publishing.
- Oliveras A, Armario P, Martell-Claros N, Ruilope LM, de la Sierra A, Spanish Society of Hypertension-Resistant Hypertension R. 2011. Urinary albumin excretion is associated with nocturnal systolic blood pressure in resistant hypertensives. *Hypertension* 57(3):556-60.

- Olsen MH, Wachtell K, Ibsen H, Lindholm L, Kjeldsen SE, Omvik P, Nieminen MS, Dahlof B, Okin PM, Devereux RB. 2011. Changes in subclinical organ damage vs. in Framingham risk score for assessing cardiovascular risk reduction during continued antihypertensive treatment: a LIFE substudy. *J Hypertens* 29(5):997-1004.
- Orchard TJ, Dorman JS, Maser RE, Becker DJ, Ellis D, LaPorte RE, Kuller LH, Wolfson SK, Jr., Drash AL. 1990. Factors associated with avoidance of severe complications after 25 yr of IDDM. Pittsburgh Epidemiology of Diabetes Complications Study I. *Diabetes Care* 13(7):741-7.
- Palatini P. 2003. Microalbuminuria in hypertension. *Curr Hypertens Rep* 5(3):208-14.
- Pannacciulli N, Cantatore FP, Minenna A, Bellacicco M, Giorgino R, De Pergola G. 2001. Urinary albumin excretion is independently associated with C-reactive protein levels in overweight and obese nondiabetic premenopausal women. *J Intern Med* 250(6):502-7.
- Paper OW. How Providers Can Lower Costs and Improve Patient Care Using Evidence Based Medicine [Internet]. ORACLE. Available from: <http://www.oracle.com/us/industries/018896.pdf>
- Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, Walsh PC. 1993. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. *J Urol* 150(1):110-4.
- Pascual JM, Rodilla E, Gonzalez C, Perez-Hoyos S, Redon J. 2005. Long-term impact of systolic blood pressure and glycemia on the development of microalbuminuria in essential hypertension. *Hypertension* 45(6):1125-30.
- Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, 3rd, Criqui M, Fadl YY, Fortmann SP, Hong Y, Myers GL et al. . 2003. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 107(3):499-511.
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. 1996. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49(12):1373-9.
- Perrone RD, Madias NE, Levey AS. 1992. Serum creatinine as an index of renal function: new insights into old concepts. *Clin Chem* 38(10):1933-53.
- Pinto-Sietsma SJ, Mulder J, Janssen WM, Hillege HL, de Zeeuw D, de Jong PE. 2000. Smoking is related to albuminuria and abnormal renal function in nondiabetic persons. *Ann Intern Med* 133(8):585-91.
- Podzolkov VI, Bragina AE, Rodionova JN. 2011. GENDER FEATURES OF MICROALBUMINURIA AND ITS ASSOCIATION TO INTRARENAL HEMODYNAMIC AND LEPTIN LEVEL IN ESSENTIAL HYPERTENSION: PP.29.435. *Journal of Hypertension* 29:e422.
- Pohl MA, Blumenthal S, Cordonnier DJ, De Alvaro F, Deferrari G, Eisner G, Esmatjes E, Gilbert RE, Hunsicker LG, de Faria JB et al. . 2005. Independent and additive impact of blood pressure control and angiotensin II receptor blockade on renal outcomes in the irbesartan diabetic nephropathy trial: clinical implications and limitations. *J Am Soc Nephrol* 16(10):3027-37.
- Poli A, Marangoni F, Avogaro A, Barba G, Bellentani S, Bucci M, Cambieri R, Catapano AL, Costanzo S, Cricelli C et al. . 2013. Moderate alcohol use and health: a consensus document. *Nutr Metab Cardiovasc Dis* 23(6):487-504.
- Poudel B, Yadav BK, Nepal AK, Jha B, Raut KB. 2012. Prevalence and association of microalbuminuria in essential hypertensive patients. *N Am J Med Sci* 4(8):331-5.
- Prevention CfDCA. Alcohol and Public Health [Internet]. Available from: <http://www.cdc.gov/alcohol/faqs.htm>
- Prujm MT, Madeleine G, Riesen WF, Burnier M, Bovet P. 2008. Prevalence of microalbuminuria in the general population of Seychelles and strong association with diabetes and hypertension independent of renal markers. *J Hypertens* 26(5):871-7.
- Pun KK, Ho PW, Varghese Z, Moorhead JF. 1989. Logistic regression model for the assessment of microalbuminuria and microproteinuria in insulin-dependent diabetic patients at risk. *Comput Methods Programs Biomed* 29(3):197-203.

- Ravid M, Brosh D, Ravid-Safran D, Levy Z, Rachmani R. 1998. Main risk factors for nephropathy in type 2 diabetes mellitus are plasma cholesterol levels, mean blood pressure, and hyperglycemia. *Arch Intern Med* 158(9):998-1004.
- Redon J, Martinez F. 2012. Microalbuminuria as surrogate endpoint in therapeutic trials. *Current Hypertension Reports* 14(4):345-349.
- Redon J, Rovira E, Miralles A, Julve R, Pascual JM. 2002. Factors related to the occurrence of microalbuminuria during antihypertensive treatment in essential hypertension. *Hypertension* 39(3):794-8.
- Redon J, Williams B. 2002. Microalbuminuria in essential hypertension: redefining the threshold. *J Hypertens* 20(3):353-5.
- Ridker PM. 2003. Cardiology Patient Page. C-reactive protein: a simple test to help predict risk of heart attack and stroke. *Circulation* 108(12):e81-5.
- Ritz E, Viberti GC, Ruilope LM, Rabelink AJ, Izzo JL, Jr., Katayama S, Ito S, Mimran A, Menne J, Rump LC et al. . 2010. Determinants of urinary albumin excretion within the normal range in patients with type 2 diabetes: the Randomised Olmesartan and Diabetes Microalbuminuria Prevention (ROADMAP) study. *Diabetologia* 53(1):49-57.
- Robles NR, Felix FJ, Fernandez-Berges D, Perez-Castan J, Zaro MJ, Lozano L, Alvarez-Palacios P, Garcia-Trigo A, Tejero V, Morcillo Y et al. . 2013. Prevalence of abnormal urinary albumin excretion in elderly people: a Spanish survey. *Int Urol Nephrol* 45(2):553-60.
- Robles NR, Felix FJ, Fernandez-Berges D, Perez-Castan JF, Zaro MJ, Lozano L, Alvarez-Palacios P, Garcia-Trigo A, Tejero V, Morcillo Y et al. . 2012. Prevalence of abnormal urinary albumin excretion in a population-based study in Spain: results from the HERMEX Study. *Eur J Clin Invest* 42(12):1272-7.
- Ross R. 1999. Atherosclerosis--an inflammatory disease. *N Engl J Med* 340(2):115-26.
- Rothman KJ. 1976. Causes. *Am J Epidemiol* 104(6):587-92.
- Rothman KJ, Greenland S, Lash TL. 2008. Modern epidemiology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Royston P, Moons KG, Altman DG, Vergouwe Y. 2009. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 338:b604.
- Ruggenti P, Remuzzi G. 2006. Time to abandon microalbuminuria? *Kidney Int* 70(7):1214-22.
- Sabanayagam C, Lee J, Shankar A, Lim SC, Wong TY, Tai ES. 2010. C-reactive protein and microalbuminuria in a multi-ethnic Asian population. *Nephrol Dial Transplant* 25(4):1167-72.
- Saha TK, Bhattarai A, Batra HS, Banerjee M, Misra P, Ambade V. 2014. Correlation of Microalbuminuria with Estimated GFR (eGFR) by Cockcroft–Gault and MDRD Formula in Type 2 Diabetics and Hypertensives. *Indian Journal of Clinical Biochemistry*:1-4.
- Sánchez Azcona P, Quaglia N. A higher prevalence of microalbuminuria in women in a diabetes management program at a public hospital in the south of Santa Fe province.
- Satchell SC, Tooke JE. 2008. What is the mechanism of microalbuminuria in diabetes: a role for the glomerular endothelium? *Diabetologia* 51(5):714-25.
- Schaeffner E, Ritz E. 2012. Alcohol and kidney damage: a Janus-faced relationship. *Kidney Int* 81(9):816-8.
- Scheuer H, Gwinner W, Hohbach J, Grone EF, Brandes RP, Malle E, Olbricht CJ, Walli AK, Grone HJ. 2000. Oxidant stress in hyperlipidemia-induced renal damage. *Am J Physiol Renal Physiol* 278(1):F63-74.
- Scheven L, Halbesma N, de Jong PE, de Zeeuw D, Bakker SJ, Gansevoort RT. 2013a. Predictors of progression in albuminuria in the general population: results from the PREVEND cohort. *PLoS One* 8(5):e61119.
- Scheven L, Van Der Velde M, Heerspink HJL, De Jong PE, Gansevoort RT. 2013b. Isolated microalbuminuria indicates a poor medical prognosis. *Nephrology Dialysis Transplantation* 28(7):1794-1801.
- Scheven L, Van der Velde M, Lambers Heerspink HJ, De Jong PE, Gansevoort RT. 2013c. Isolated microalbuminuria indicates a poor medical prognosis. *Nephrol Dial Transplant* 28(7):1794-801.

- Schmieder RE, Mann JF, Schumacher H, Gao P, Mancia G, Weber MA, McQueen M, Koon T, Yusuf S, Investigators O. 2011. Changes in albuminuria predict mortality and morbidity in patients with vascular disease. *J Am Soc Nephrol* 22(7):1353-64.
- Schrader J, Luders S, Kulschewski A, Hammersen F, Zuchner C, Venneklaas U, Schrandt G, Schnieders M, Rangoonwala B, Berger J et al. . 2006. Microalbuminuria and tubular proteinuria as risk predictors of cardiovascular morbidity and mortality in essential hypertension: final results of a prospective long-term study (MARPLE Study)*. *J Hypertens* 24(3):541-8.
- Sernyak MJ, Rosenheck R. 2003. Risk adjustment in studies using administrative data. *Schizophrenia Bulletin* 29(2):267-271.
- Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM. 2003. C-reactive protein and the risk of developing hypertension. *JAMA* 290(22):2945-51.
- Simon J, Amde M, Poggio ED. 2011. Interpreting the estimated glomerular filtration rate in primary care: benefits and pitfalls. *Cleve Clin J Med* 78(3):189-95.
- Singh A, Satchell SC. 2011. Microalbuminuria: causes and implications. *Pediatr Nephrol* 26(11):1957-65.
- Sleight P. 2000. The HOPE Study (Heart Outcomes Prevention Evaluation). *J Renin Angiotensin Aldosterone Syst* 1(1):18-20.
- Smink PA, Lambers Heerspink HJ, Gansevoort RT, de Jong PE, Hillege HL, Bakker SJ, de Zeeuw D. 2012. Albuminuria, estimated GFR, traditional risk factors, and incident cardiovascular disease: the PREVEND (Prevention of Renal and Vascular Endstage Disease) study. *Am J Kidney Dis* 60(5):804-11.
- Sobell L, Sobell M. 1992. Timeline Follow-Back. In: Litten R, Allen J, editors. *Measuring Alcohol Consumption*. Humana Press. p. 41-72.
- Squires DA. 2012. Explaining high health care spending in the United States: an international comparison of supply, utilization, prices, and quality. *Issue Brief (Commonw Fund)* 10:1-14.
- Stankeviciute N, Sabah S, Singh A, Shaykh M, Bakir AA, Arruda JA, Dunea G. 1998. Can total urinary protein measurements predict microalbuminuria? *Am J Nephrol* 18(4):285-90.
- Stanton M. The high concentration of U.S. health care expenditures [Internet]. Rockville (MD): Agency for Healthcare Research and Quality. Available from: <http://www.ahrq.gov/research/findings/factsheets/costs/expriach/expendria.pdf>
- Stephenson AJ, Kattan MW. 2006. Nomograms for prostate cancer. *BJU Int* 98(1):39-46.
- Steyerberg EW. 2009. *Clinical prediction models : a practical approach to development, validation, and updating*. New York ; London: Springer.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21(1):128-38.
- Stuveling EM, Bakker SJ, Hillege HL, Burgerhof JG, de Jong PE, Gans RO, de Zeeuw D, Group PS. 2004. C-reactive protein modifies the relationship between blood pressure and microalbuminuria. *Hypertension* 43(4):791-6.
- Sullivan LM, Massaro JM, D'Agostino RB, Sr. 2004. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 23(10):1631-60.
- Szumilas M. 2010. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry* 19(3):227-9.
- Tagle R, Gonzalez F, Acevedo M. 2012. [Microalbuminuria and urinary albumin excretion in clinical practice]. *Rev Med Chil* 140(6):797-805.
- Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, Levin A, Levey AS. 2011. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 305(15):1553-9.
- Tapp RJ, Shaw JE, Zimmet PZ, Balkau B, Chadban SJ, Tonkin AM, Welborn TA, Atkins RC. 2004. Albuminuria is evident in the early stages of diabetes onset: results from the Australian Diabetes, Obesity, and Lifestyle Study (AusDiab). *Am J Kidney Dis* 44(5):792-8.
- Thakkar* B, Arora** K, Vekariya*** R, Lulania**** M, ***** ASA. 2011. Prevalence of Microalbuminuria in Newly Diagnosed Type 2 Diabetes Mellitus. *NJIRM* 2(4):22-25.
- Toto RD. 2004. Microalbuminuria: definition, detection, and clinical significance. *J Clin Hypertens (Greenwich)* 6(11 Suppl 3):2-7.

- Trevisan R, Dodesini AR, Lepore G. 2006. Lipids and renal disease. *J Am Soc Nephrol* 17(4 Suppl 2):S145-7.
- Tseng CH. 2005. Correlation of uric acid and urinary albumin excretion rate in patients with type 2 diabetes mellitus in Taiwan. *Kidney Int* 68(2):796-801.
- van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B. 2010. The global burden of diabetes and its complications: an emerging pandemic. *Eur J Cardiovasc Prev Rehabil* 17 Suppl 1:S3-8.
- Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, Initiative S. 2007. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 4(10):e297.
- Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. 2005. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 58(5):475-83.
- Verhave JC, Hillege HL, Burgerhof JGM, Navis G, de Zeeuw D, de Jong PE, Group ftPS. 2003. Cardiovascular Risk Factors Are Differently Associated with Urinary Albumin Excretion in Men and Women. *Journal of the American Society of Nephrology* 14(5):1330-1335.
- Viazzi F, Parodi D, Leoncini G, Parodi A, Falqui V, Ratto E, Vettoretti S, Bezante GP, Del Sette M, Deferrari G et al. . 2005. Serum uric acid and target organ damage in primary hypertension. *Hypertension* 45(5):991-6.
- Volpe M. 2008. Microalbuminuria screening in patients with hypertension: recommendations for clinical practice. *Int J Clin Pract* 62(1):97-108.
- Waden J, Forsblom C, Thorn LM, Gordin D, Saraheimo M, Groop PH, Finnish Diabetic Nephropathy Study G. 2009. A1C variability predicts incident cardiovascular events, microalbuminuria, and overt diabetic nephropathy in patients with type 1 diabetes. *Diabetes* 58(11):2649-55.
- Weir MR. 2007. Microalbuminuria and cardiovascular disease. *Clin J Am Soc Nephrol* 2(3):581-90.
- Wharam JF, Weiner JP. 2012. The promise and peril of healthcare forecasting. *Am J Manag Care* 18(3):e82-5.
- White SL, Polkinghorne KR, Cass A, Shaw JE, Atkins RC, Chadban SJ. 2009. Alcohol consumption and 5-year onset of chronic kidney disease: the AusDiab study. *Nephrol Dial Transplant* 24(8):2464-72.
- Winocour PH, Marshall SM. 1998. Microalbuminuria : from epidemiology to clinical management. Cambridge: Cambridge University Press.
- Woloshin S, Schwartz LM. 2005. Distribution of C-reactive protein values in the United States. *N Engl J Med* 352(15):1611-3.
- Wood AM, White IR, Royston P. 2008. How should variable selection be performed with multiply imputed data? *Stat Med* 27(17):3227-46.
- Xu J, Lee ET, Devereux RB, Umans JG, Bella JN, Shara NM, Yeh J, Fabsitz RR, Howard BV. 2008. A longitudinal study of risk factors for incident albuminuria in diabetic American Indians: the Strong Heart Study. *Am J Kidney Dis* 51(3):415-24.
- Yudkin JS, Forrest RD, Jackson CA. 1988. Microalbuminuria as predictor of vascular disease in non-diabetic subjects. Islington Diabetes Survey. *Lancet* 2(8610):530-3.
- Zakkerkish M, Shahbazian HB, Shahbazian H, Latifi SM, Moravej Aleali A. 2013. Albuminuria and its correlates in type 2 diabetic patients. *Iran J Kidney Dis* 7(4):268-76.
- Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115(5):654-7.