UCLA
UCLA Electronic Theses and Dissertations

Title

A Psychometric Analysis of Patient-Reported Outcomes in Chronic Kidney Disease

Permalink

https://escholarship.org/uc/item/1r43h6bn

Author

Peipert, John Devin

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Psychometric Analysis of Patient-Reported Outcomes in Chronic Kidney Disease

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Community Health Sciences

by

John Devin Peipert

2017

ABSTRACT OF THE DISSERTATION

A Psychometric Analysis of Patient-Reported Outcomes in Chronic Kidney Disease

by

John Devin Peipert

Doctor of Philosophy in Community Health Sciences

University of California, Los Angeles, 2017

Professor Donald Morisky, Chair

Background: Survival is a critical outcome in chronic kidney disease (CKD), but it provides a limited view of how well patients are doing. Many aspects of patients' health can only be obtained by patient reported measures (PRMs), such as health-related quality of life (HRQOL). This dissertation examines the psychometric properties of currently used PRMs in chronic kidney disease (CKD) and make recommendations for how collection and reporting of PRMs can be systematized in the CKD field.

Methods: This dissertation used data from three separate sources, each containing kidney patients' responses to PRMs: treatment decision-making, medication adherence, and HRQOL. The treatment decision-making PRMs examined include patients' Decisional Balance (perceived Pros and Cons) and Self-Efficacy to pursue both living and deceased donor kidney transplant (LDKT, DDKT; 6 measures in total). The 8-item Morisky Medication Adherence Scale (MMAS-

8) was examined as a measure of medication adherence. Finally, the Kidney Disease Quality of Life (KDQOL)-36 was examined as a measure of HRQOL. For each measure, internal consistency reliability was estimated. Dimensional structure was examined with exploratory and confirmatory factor analysis (EFA, CFA) and item-scale correlations, corrected for item overlap. To determine the dimensionality suggested by the exploratory factor analysis, several criteria were used, including the scree "elbow" test, parallel analysis, and the Tucker-Lewis reliability coefficient. For CFA models, model fit was determined with the Satorra-Bentler chi-square, the comparative fit index (CFI), Tucker-Lewis Index (TLI) and root mean square error of approximation (RMSEA). Good model fit is evidenced by a non-significant Satorra-Bentler chi-square, CFI and TLI values of above 0.95, and RMSEA of 0.06 or less. Next, the measurement invariance for each measure was examined between Black and White patients. Finally, recommendations for improvements to each measure were made, including calculations to determine the number of items needed to achieve good ($\geq$0.80) and excellent ($\geq$0.90) reliability (reliability of $\geq$0.70 is considered adequate). Generally, excellent reliability is required for use with individuals.

Summary of Results: For both the LDKT and DDKT Pros and Cons measures, 2 correlated factors CFA models fit the data well, verifying the original dimensional structures for these measures. Additionally, the LDKT and DDKT Self-Efficacy measures were supported by single factor CFA models, also supporting their original dimensional structure. However, several of these scales only evidenced adequate internal consistency reliability (LDKT Pros, LDKT Cons, DDKT Cons). These scales would require the addition of up to 15-17 parallel items to achieve excellent reliability ($\geq$0.90). Regarding the MMAS-8, the original factor structure was not supported by EFA and CFA models. One item, "Did you take your medicine yesterday?," was

weakly correlated (r<0.243) with the others, and did not load highly (λ<0.40) in EFA models. A CFA model with this item removed fit the data well (RMSEA = 0.06; CFI = 0.99; TLI = 0.98). The internal consistency reliability of the modified scale was 0.78, and 18 items would need to be added to achieve excellent reliability. Finally, regarding the KDQOL-36, the original factor structure was supported, and the internal consistency reliability for each KDQOL-36 scale exceeded the criterion for "good" reliability ($\geq$0.80), though the addition of up to 11 items would be required to increase reliability to "excellent" (Symptoms/Problems scale). For all scales in the dissertation, recommendations were made for increasing reliability using classical test theory and item response theory.

The dissertation of John Devin Peipert is approved.

Linda B. Bourque

Courtney Thomas

Peter Bentler

Ron D. Hays

Donald E. Morisky, Committee Chair

University of California, Los Angeles

2017

To my Committee and mentors, whose guidance made this project possible. To my wife, who always encouraged me to innovate, and to my parents, who believed it was possible to achieve my goals.

TABLE OF CONTENTS

LIST OF ACRONYMS AND ABBREVIATIONS

Acronyms and abbreviations listed under the chapter where they are introduced.

<u>Chapter 1</u>

ASCQ-ME: Adult Sickle Cell Quality of Life Measurement Information System

CAT: Computer adaptive testing

CKD: Chronic kidney disease

CMS: Centers for Medicare and Medicaid Services

ESRD: End-stage renal disease

HIE: Health Insurance Experiment

HRQOL: Health-related quality of life

IRT: Item response theory

KDQOL: Kidney Disease Quality of Life

MOS: Medical Outcomes Study

PRM: Patient-reported measures

PRO: Patient-reported outcomes

PROMIS: Patient Reported Outcomes Measurement Information System

USRDS: United States Renal Data System

<u>Chapter 2</u>

BMQ: Brief Medication Questionnaire

CHOICE: Choices for Health Outcomes in Caring for ESRD

CTT: Classical Test Theory

HMO: Health maintenance organization

LDKT: Living donor kidney transplant

MAQ: Medication Adherence Questionnaire

MARS: Medication Adherence Report Scale

MMAS: Morisky Medication Adherence Scale

MGL: Morisky, Green and Levine medication adherence scale

RRT: Renal replacement therapy

SMAQ: Simplified Medication Adherence Questionnaire

TTM: Transtheoretical Model of Behavior Change

Chapter 4

AV: Arteriovenous

CFI: Comparative fit index

CRIC: Chronic Renal Insufficiency Cohort

CFA: Confirmatory factor analysis

DDKT: Deceased donor kidney transplant

DOPPS: Dialysis Outcomes and Practice Patterns Study

EFA: Exploratory factor analysis

EQ-VAS: European Quality of Life Visual-Analog Scale

IRB: Institutional Review Board

KDCS: Kidney Disease Component Score

MCS: Mental Health Component Score

MEI: Medical Education Institute

PCS: Physical Health Component Score

RMSEA: Root mean squared error approximation

TLI: Tucker-Lewis Index

WLSMV: Weighted least squares with mean and variance adjustment

LIST OF FIGURES

LIST OF TABLES

LIST OF DATASETS

Throughout the dissertation, the following datasets will be referred to by the names given below.

TREATMENT DECISION-MAKING DATASET: Described on p. 46. Covers constructs relevant to living and deceased donor transplant decision-making, including Decisional Balance and Self-Efficacy

MEDICATION ADHERENCE DATASET: Described on pp. 46-47. Covers the CKD medication adherence construct.

HEALTH-RELATED QUALITY OF LIFE DATASET: Covers the measure of health-related quality of life for CKD.

Date: 5/21/2017

EDUCATION

- PhD candidate in Community Health Sciences, Fielding School of Public Health, University of California, Los Angeles
- Masters of Social Work (MSW) focused on Health and Human Development, 2009; Specialization in Applied Statistics*;* Brown School of Social Work and Public Health, Washington University in St. Louis (***additional coursework in epidemiological/statistical methods, especially survival analysis, clinical study design, large database management***)
- MA in Philosophy, 2007; Focus on philosophy of science, social theory, and philosophy of language, University of Missouri, St. Louis
- BA in Philosophy, 2003; Southern Illinois University at Edwardsville, Concentration in formal logic

EXPERIENCE

November 2013-present: Principal Statistician, Division of Nephrology, David Geffen School of Medicine, University of California Los Angeles

November 2011-November 2013: Research Statistician, Division of General Medical Sciences, Washington University School of Medicine, Washington University in St. Louis

August 2010-November 2011: Research Data Analyst, Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University

HONORS AND AWARDS

Dolores Baja-Lasan Scholarship, Brown School of Social Work, Washington University in St. Louis, 2007-2010
University Fellowship, Fielding School of Public Health, UCLA, 2014-present

PROFESSIONAL SOCIETY AND ORGANIZATION MEMBERSHIP

American Society of Transplantation (AST)
American Society of Transplantation Psychosocial Community of Practice (PSCoP)
American Society of Transplantation Living Donor Community of Practice (LDCoP)
Southern California Chapter of the American Statistical Association (SCASA)

SELECTED PEER-REVIEWED PUBLICATIONS

1. Butt Z, Peipert JD, Webster K, Chen C, Cella D. General population norms for the Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI). *Cancer.* Jan 15 2013;119(2):429-437.
2. Waterman AD, Peipert JD, Hyland SS, McCabe MS, Schenk EA, Liu J. Modifiable Patient Characteristics and Racial Disparities in Evaluation Completion and Living Donor Transplant. *Clin J Am Soc Nephrol.* 2013; 8(6):995-1002.
3. Peipert JD, Beaumont JL, Bode R, Cella D, Garcia SF, Hahn EA. Development and validation of the functional assessment of chronic illness therapy treatment satisfaction (FACIT TS) measures. *Qual Life Res*. 2014; 23(3):815-824.
4. Perman, T., Yanez, B., Peipert JD, Wortman, K., Beaumont JL, Bode R, Cella D. Ambulatory Cancer and U.S. General Population Reference Values and Cutoff Scores for the Functional Assessment of Cancer Therapy. *Cancer.* 2014; 120(18):2902-2909.
5. Waterman AD, Robbins ML, Paiva AL, Peipert JD., et al. Your path to transplant: a randomized controlled trial of a tailored computer education intervention to increase living donor kidney transplant. *BMC Nephrol.* 2014;15(1):166.
6. Waterman AD, Robbins ML, Paiva AL, Peipert JD., et al. Measuring kidney patients' motivation to pursue living donor kidney transplant: Development of Stage of Change, Decisional Balance and Self-Efficacy measures. *J Health Psychol.* 2015;20(2):210-221.
7. Brick L, Sorensen D, Robbins M, [...], Peipert, JD, et al. Invariance of Measures to Understand Decision Making for Pursuing Living Donor Kidney Transplant. *J Health Psychol*. 2015; Epub 2015/06/25.
8. Waterman AD, McSorley A-MM, Peipert JD, Goalby CJ, Peace LJ, Lutz PA, Thein JL. Explore Transplant at Home: A randomized control trial of an educational intervention to increase transplant knowledge for Black and White socioeconomically disadvantaged dialysis patients. *BMC Nephrol*. 2015; 16(150):1-11.
9. Waterman AD, Peipert JD, Goalby CJ, Dinkel KM, Xiao H, Lentine KL. Assessing Transplant Education Practices in Dialysis Centers: Comparing Educator Reported and Medicare Data. *Clin J Am Soc Nephrol*. 2015; 10(9); 1617-1625.
10. Waterman AD, Robbins ML, Peipert JD. Educating Prospective Kidney Transplant Recipients and Living Donors about Living Donation: Practical and Theoretical Recommendations for Increasing Living Donation Rates. *Curr Transpl Rep.* 2016;3:1-9.
11. Talamantes E, Norris KC, Mangione CM, Moreno G, Waterman AD, Peipert JD, Bunnapradist S, Huang E. Incidence of Conversion to Active Waitlist Status Among Temporarily Inactive Transplant Candidates from Linguistically Isolated Communities. *Clin J Am Soc Nephrol*. 2017; 12(3); 483-492.
12. Axelrod DA, Kynard-Amerson CS, Wojciechowski D, Jacobs M, Lentine KL, Schnitzler M, Peipert JD, Waterman AD. Cultural Competency of a Mobile, Customized Patient Education Tool for Improving Potential Kidney Transplant Recipients' Knowledge and Decision-Making. *Clin Transplant*. 2017; 31(5); e12944.
13. Peipert JD, Hays RD. Using Patient Reported Measures in Dialysis Clinics. *Clin J Am Soc Nephrol*. In press.
14. Weng FL, Brown D, Holland B, Peipert JD, Waterman AD. A Cluster Randomized Trial of an Educational Intervention to Increase Knowledge of Living Donor Kidney Transplant among Potential Transplant Candidates. *Prog Transplant*. In press.

CHAPTER 1. PROBLEM AND SIGNIFICANCE

The field of chronic kidney disease (CKD) is an area of medicine wherein the use of patient-reported measures (PRMs) is highly relevant, and to date, many excellent measures have been developed.[1-3] For example, the Kidney Disease Quality of Life (KDQOL™) and In-Center Hemodialysis Consumer Assessment of Healthcare and Providers and Systems (CAHPS®) measures are widely used in research and clinical applications.[1,3] However, to date no large, systematic effort that could be classified as a "measurement information system" has been made to account for and coordinate extant measures across the range PRMs. CKD is a clinical area ripe for such a systematic effort due to the extensive oversight of CKD health services like dialysis and transplant by Centers for Medicare and Medicaid Services (CMS). Since CMS covers the cost of patient care for most renal-replacement therapy, an extensive effort is made to track outcomes, with large data collection projects funded by government agencies. For example, the United States Renal Data System (USRDS), funded by the NIH, collects data and publishes annual reports on CKD outcomes.[4] The USRDS provides data to examine outcomes like hospitalization and patient survival, but has not incorporated comprehensive measures of PRMs or patient-reported outcomes (PROs).

A very common outcome measure in medical studies is disease progression free survival.[5-8] In the field of kidney transplantation, CMS monitors each transplant center's rates of patient survival and survival of the transplanted kidney as a performance measure on which reimbursement for services is contingent.[9] However, focus on survival only provides a limited view of how well patients are doing. Many aspects of patients' health can only be determined through patient reports. Patient reports provide an opportunity to understand health and health care from the patient's perspective. For instance, health-related quality of life (HRQOL) and

1

reports and ratings of health care can shed light on the effectiveness of medical interventions or the severity of impact of progressing disease.

The U.S. Food and Drug Administration (FDA) defines PRMs as: ''any report coming from patients about a health condition and its treatment, without interpretation of the patient's response by a clinician or anyone else'' (p. 2).[10] PRMs include PROs. The acronym "PROs" is often used to refer to the entire range of PRMs, but it has been noted that PROs should be limited to HRQOL and patient satisfaction with care.[11] Other important types of PRMs include health preferences, behaviors, and even key patient characteristics.

Use of PRMs in public health research dates back many years. For example, in the 1950's, Suchman and colleagues examined the measurement properties of patient reports of their health with items like "How would you rate your health at the present time?" (rated with 5 response options from "Excellent" to "Very Poor",[12] which is still used today.[13] Additionally, reports of satisfaction with health care also date back to the 1950's.[14] RAND's Health Insurance Experiment (HIE), which began in 1974 and lasted through 1982, was among the first large, systematic studies of PRMs.[15] Additionally, the HIE focused on other types of PRMs, like patient characteristics and health behaviors (e.g., smoking).

The Medical Outcomes Study (MOS) followed the HIE and generated several of the most commonly used PROs.[16,17] The general aims of the MOS were to determine if characteristics of care delivery systems impacted outcomes for patients with one or more of four chronic conditions: hypertension, diabetes, heart disease, and depression. The MOS produced a battery of PRMs. The MOS was a longitudinal study conducted at multiple study sites, including Los Angeles, Chicago, and Boston. Physicians and patients were enrolled from multiple health care

systems. Ultimately, 362 clinicians contributed responses to questionnaires and patients to the

study, and 22,785 patients participated in the cross-sectional component while 2,056 patients

completed the longitudinal component. The project produced widely-used PROs including the

most widely used PRO to date: the SF-36[18,19] and the SF-12.[20]

During the last decade, there have been large, national efforts to improve and standardize

Figure 1. PROMIS Conceptual Model
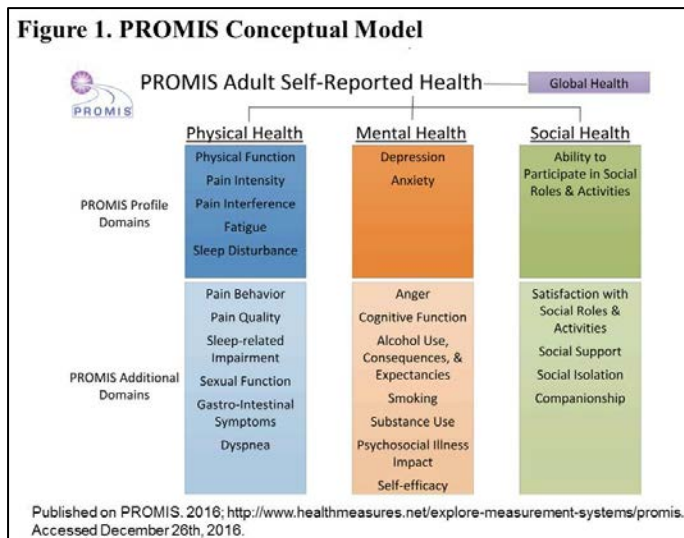
PROs. The largest recent effort is the Patient Reported Outcomes Measurement Information System (PROMIS)[21,22] project. PROMIS was funded as part of the NIH Roadmap for Medical Research,[23] a pan-Institute, strategic funding plan aiming to solve major medical problems in the 21st

Century. PROMIS takes an innovative approach to the development and evaluation of PROs

through the use of item response theory (IRT) and computer adaptive testing (CAT) that can

draw from large banks of items to generate efficient, reliable, and parsimonious individually-

tailored measures of patients' HRQOL. The PROMIS conceptual model of key dimensions is

shown in Figure 1. Note that PROMIS HRQOL measures are not disease-specific and may not

capture all the specific concerns germane to patients with particular conditions. There are many

PROMIS measures, but the PROMIS-29[24] is to the SF-36, with physical and mental health

components arising out of 8 domains: 1) physical function; 2) anxiety; 3) depression; 4) fatigue;

5) sleep disturbance; 6) ability to participate in social roles and activities; 7) pain interference; 8)

pain intensity.

Other large, systematic, HRQOL measure development and evaluation efforts are condition or disease state-targeted. The NIH Neuro-QOL (Quality of Life in Neurological Disorders) study developed a similar system of HRQOL measures to PROMIS but these were targeted toward patients with neurological disorders.[25] Neuro-QOL adopted a similar conceptual framework to PROMIS with Physical Health, Mental Health, and Social Health overarching dimensions and domains within these broader dimensions being of particular relevance to patients with neurological disorders.[26] (Figure 2) Neuro-QOL used the same basic methodology used in PROMIS, including IRT and computer adaptive testing with

Figure 2. Neuro-QOL Conceptual Model

Neuro·QoL — Adult Self-Reported Health

Physical Health	Mental Health	Social Health
Fatigue	Anxiety	Ability to Participate in Social Roles & Activities
Sleep Disturbance	Depression	
Lower Extremity Function (Mobility)	Positive Affect & Well-Being	Satisfaction with Social Roles & Activities
Upper Extremity Function (Fine Motor, ADL)	Emotional & Behavioral Dyscontrol	
	Cognitive Function	
Bowel Function*	Communication	
Urinary/Bladder Function*	Stigma	
Sexual Function*	End of Life Concerns	

*Untested Item Pools

Published on Neuro-QOL. 2016; http://www.healthmeasures.net/explore-measurement-systems/neuro-qol/. Accessed December 26th, 2016.

large item banks to generate individualized, parsimonious HRQOL assessments. However, the item banks consist of condition-targeted items. Other examples of state-of-the-science PRO measurement projects currently underway include ASCQ-ME (Adult Sickle Cell Quality of Life Measurement Information System)[27] and the NIH Toolbox project,[28] which features significant use of PROs in its emotion-oriented measures as well as in its vision-targeted HRQOL set of measures.[29]

In summary, there has been significant effort to develop and validate PRMs for kidney patients specifically, and for all patients generically, over the last two decades. Despite this significant effort, there are still important psychometric questions to be answered about the measures that have been developed. Some of these questions regard the basic psychometric

properties of very commonly used measures with the general CKD population, while some regard developing a better understanding of certain measures' psychometric properties with specific populations of patients (e.g., with dialysis patients only). Further, as the science of health measure has advanced over the last decade, it is critical to determine if some standard measures meet our current standards. This dissertation will examine the psychometric properties of currently used PRMs in CKD and make recommendations for how these PRMs can be improved and made appropriate to all CKD patients. The overall goal of the dissertation is to determine to what extent there are psychometrically-sound PRMs for major constructs related to CKD patients' health and health care and to determine next steps in improving the science and clinical application of PRMs in CKD.

CHAPTER 2. THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1. OVERVIEW OF CKD

2.1.1 CKD EPIDEMIOLOGY

Presently, there are over 30 million patients with CKD and 650,000 with ESRD, or kidney failure, in the United States.[30] ESRD patients must have medical therapy to replace their kidney function to sustain life. The two options available are dialysis, where a machine filters toxins from the blood, or a kidney transplant. Compared to dialysis, kidney transplant is associated with much longer survival and better HRQOL. Specifically, depending on the patient's age, transplant offers 6-16 additional years of life over dialysis.[30] Despite these benefits, most ESRD patients, 70%, remain on dialysis and do not pursue kidney transplant.[30]

Many CKD patients may have a difficult time choosing the best treatments for their disease because they are uninformed,[31] due a lack of factual information about all treatments' risks and benefits and risks (e.g., length of additional life offered, chance of problems with a medical procedure).[32] Additionally, several transplant education programs have been developed over the last decade to help inform dialysis and transplant patients about the risks and benefits of getting a transplant.[33-39]

There are significant racial disparities in the prevalence and incidence of diabetes and hypertension.[40] The incidence of ESRD due to diabetes increased by 33% between 2000 to 2012 for Blacks, but only 2% for Whites during the same period.[41] Blacks are 3.9 times more likely to develop ESRD than Whites. Despite its clinical superiority as a therapy, Blacks are 4.2 times less likely to receive transplants than Whites.[41] This is especially the case regarding transplants from living donors, which are associated with the best outcomes. Recent evidence shows that Blacks

patients are less likely to receive transplants at every transplant center in the United States.[42] This evidence creates a picture of a detrimental double disparity: African-Americans are much more likely to develop ESRD than Whites, but much less likely to receive its optimal treatment, transplantation.

Over the past few decades, an interesting line of research has been developed that suggests African-American dialysis patients have better HRQOL than Whites.[43,44] These findings have sometimes been used in clinical decision-making to reduce transplant referrals for transplant among African-American dialysis patients, despite better survival and HRQOL outcomes for transplant among the general population of CKD patients. It is currently unclear whether or not there is an actual different in HRQOL between White and Black patients, or whether there is simply a difference in the way patients from these racial groups respond to common HRQOL measures. Before it is known whether this clinical judgement is appropriate, the measurement equivalence of common HRQOL measures between African-Americans and patients of other races should be established.

2.2 OVERALL CONCEPTUAL MODEL

To identify the most important PRMs for chronic kidney disease, a conceptual framework is needed that incorporate all key health constructs and specifies the relationships between them. Fung and Hays developed a conceptual model that identifies multiple types of PRMs.[11] (Figure 3) In their original

Figure 3. Fung and Hays Conceptual Model for Patient Reported Measures

Originally published in Fung CH, Hays RD. Prospects and challenges in using patient-reported outcomes in clinical practice. *Qual Life Res.* 2008; 17(10):1297-1302.
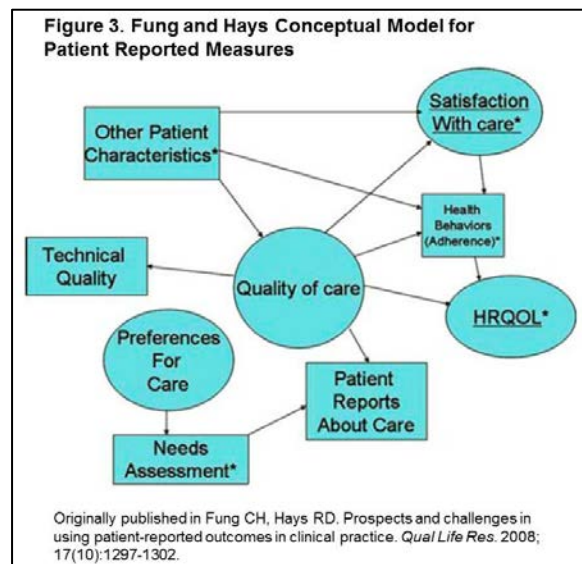
framework, Fung and Hays distinguish between underlying constructs and direct indicators that

7

can be operationalized as PRMs. For example, a health behavior such as adherence to medications can be measured by patient reports on whether they have or have not adhered to a prescribed medication regimen in response to questions asking directly about medication taking (e.g., "Did you take your medications yesterday at the prescribed times?"); such questions assess whether the behavior occurred or not, and they do not intend to represent a larger, underlying construct. On the other hand, HRQOL is a construct that cannot be indicated directly and usually requires a series of questions that each represents related aspects of the larger concept.

Major constructs in this framework include preferences for care, HRQOL, quality of care, and satisfaction with care. The direct indicators include patient characteristics, technical quality of care, needs assessment, patient reports about care, and health behaviors. Of the many relationships between these constructs and indicators, it is important to note that HRQOL is exclusively an outcome and does not influence the other depicted constructs or indicators. On the other hand, HRQOL is influenced by other concepts and indicators, including health behaviors and quality of care, which do not count as outcomes, and therefore are not PROs. Though satisfaction with care is generally considered a PRO, it also impacts health behaviors. It is important to distinguish between patients' *satisfaction with care*, and patients' *experiences with care*. *Satisfaction with care* regards discrepancies between patients' expectations for care and the care they actually receive.[45] *Experience with care* refers to aspects of the care patients receive and interactions with different elements of the health care system.[45] Indicators like patient reports about care may be used to measure experiences with care, along with preferences for care and ratings of care. Of these, the needs assessment and patient reports about care would be distinguished as PRMs, while technical quality of care would not because it is assessed using

8

expert consensus. Additionally, other patient characteristics, like demographics, are also considered PRMs.

Though this framework lays excellent groundwork for understanding the types of PRMs important for CKD patients, the following adjustments will be made to adjust the framework for this project. First, many PRMs could be implemented as underlying constructs or indicators. An example includes the adherence to medications, a health behavior. As described above, assessments of medication taking may only ask about specific medication-taking behaviors and not indicate a larger construct. However, some patient reports of medication adherence actually use multi-item scales to measure a latent behavioral construct. It is not necessarily required that the conceptual model specifies whether the PRM be measured as an underlying construct or an indicator unless a specific recommendation is made about which type of measure should be used. Therefore, the adaptation of Fung and Hays's framework will not distinguish between underlying constructs and indicators. Fung and Hays also recognize that many constructs can be measured as latent variables or indicators.

However, taking the National Quality Forum's lead,[46] there are advantages to distinguishing between the constructs to be measured with PRMs, and the different types of PRMs that can be used to measure the range of salient constructs and concepts. Doing so may help non-technical audiences discriminate between the basic ideas they want to measure and the range of specific instruments they can choose from to do so. Therefore, the adapted Fung and Hays conceptual model features two separate parts, one featuring the constructs and concepts relevant to CKD patients and the health services they use, and another indicating the types of PRMs used to measure these constructs and concepts.

A third set of adjustments for the Fung and Hays framework regard a more explicit role for use of health services, including decision-making about which treatments to use. CKD patients may have many treatment options available to them. Further, if kidney failure is reached, referred to often as CKD stage 5, patients must obtain a renal replacement therapy (RRT) to remain living. Their two general options are dialysis and kidney transplantation. Within dialysis options, patient can choose: 1) hemodialysis, where a dialysis machine cleans the blood by removing and returning it from the body while cycling it through an "artificial kidney" that contains dialysate, a blood cleansing fluid; 2) peritoneal dialysis, where the dialysate is kept in the tissue of the patient's own belly, then removed after the blood is cleansed through a catheter. Hemodialysis often requires going to a dialysis center 3 times per week for many hours. Individuals may pursue a transplant from a deceased or a living donor. Finally, kidney patients may choose no RRT, especially if their kidneys have not yet failed, opting instead to make lifestyle changes, like improvements in diet, exercise, increased adherence to hypertension of diabetes medications, or ceasing tobacco and alcohol usage in order to improve kidney function.

Given this set of choices, there is significant clinical interest in understanding how CKD patients choose a RRT, and making sure that they understand the risks and benefits of each treatment. For this reason, the CMS requires that dialysis patients be informed of their option for transplant within the first 45 days of starting treatment.[47] In their present framework, Fung and Hays point to several aspects of use of, and decision-making about, health services, including patients' preferences for care. However, preferences for care do not include all aspects of treatment decision-making, and should instead be treated as a sub-construct of Treatment Decision-Making along with other sub-constructs from a range of health behavior theories. First, the Transtheoretical Model (TTM) of Behavior Change employs the construct Decisional

Balance, which is a patients' weighting of the pros and cons of a specific treatment.[48,49] Second, Self-Efficacy is defined in the TTM as a patient's ability to overcome perceived barriers to getting a specific treatment, which is somewhat different than the general definition of self-efficacy used in other health behavior theories.[50-52] Finally, knowledge of the treatment is a key construct related to treatment decision-making in the Health Belief Model, the Theory of Reasoned Action/Theory of Planned Behavior (or their integration).[53] All these constructs should be represented, along with preferences for care, in the conceptual model as treatment decision-making constructs.

Finally, additional causal pathways are specified in the newer version of the model. The additional causal pathways are influenced by Andersen's Behavioral Model of Access to Health Care,[54] a very frequently used conceptual framework of health services use. Like the Fung and Hays framework, Andersen's model includes HRQOL as an outcome. However, in Andersen's model, HRQOL has a recursive impact on health services usage and health behaviors and health care usage. This is especially appropriate for CKD patients, who may use multiple primary types of RRT through the course of their disease trajectory such that changes in RRTs might be influenced by the HRQOL and satisfaction with the current treatment.

Figure 4. Conceptual Framework of Patient Reported Measures in Chronic Kidney Disease
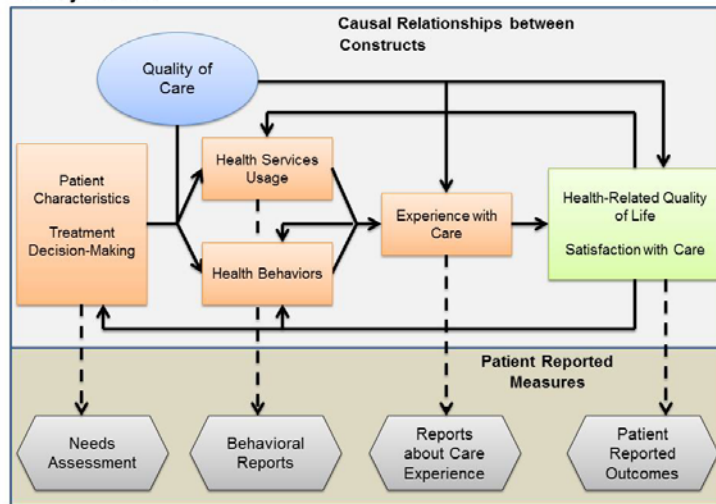
These modifications of Fung and Hays's conceptual model lead to a new conceptual model, shown in Figure 4, which can be employed to identify PRMs in CKD. This conceptual model serves two purposes: 1) to identify the most important constructs to be measured with PRMs and the causal relationships between them (upper panel); and 2) to identify the types of PRMs used to operationalize each of these constructs (lower panel). In Figure 4, constructs are represented with boxes and categories of PRMs are represented with hexagons. Only one non-patient reported measure is included, Quality of Care, which is represented in an oval. In the upper panel, solid lines represent causal relationships between constructs. Dashed lines between the constructs and PRMs indicate which types of PRMs are used to measure groups of constructs.

Regarding the causal part of the model, Patient Characteristics (e.g., demographic characteristics) and Treatment Decision-Making influence both Health Services Usage and Heath Behaviors. Health Services Usage represents which treatment patients use; in the case of CKD patients, this would indicate dialysis vs. transplant, use of primary care services in addition to nephrology care, use of preventative services [for patients who have not reached end-stage renal

disease (ESRD) yet], among other services. Health behaviors include many behaviors that patients may take that impact their health, including their diet, exercise routines, use of tobacco, alcohol, and other substances. In addition, there is a recursive path from the behavior construct to itself to reflect substantial research indicating that previous health behaviors are among the strongest predictors of future health behaviors.[55] Among the most important health behaviors patients take is their medication adherence. Depending on their stage of kidney disease and the treatment they receive, kidney patients may have an extensive medication burden, on average taking 8-10 medications daily,[56] with some patients taking more than 20 pills per day.[57] Two of the most frequently studied medications among dialysis patients include phosphate binders and anti-hypertension medications.[58] In addition, patients on dialysis manage adherence to their dialysis treatment schedule, diet, and fluid intake requirements.[58] Among kidney patients on dialysis, as well as those with transplant, medication adherence is among the strongest predictors of health outcomes, especially hospitalizations[58] and mortality.[59]

Quality of care impacts both Health Services Usage and Health Behaviors, along with patients' Experience with Care. Experience with Care is defined as "the range of interactions that patients have with the health care system, including their care from health plans, and from doctors, nurses, and staff in hospitals, physician practices, and other health care facilities".[60] CMS has adopted the In-Center Hemodialysis Consumer Assessment of Healthcare Providers and Systems (ICH-CAHPS) as the standard measure of Experience with Care in dialysis centers. Such measures can be used for internal quality improvement, as well as for kidney clinics' performance monitoring. The ICH-CAHPS includes 3 composite scales, including Nephrologists Communication and Caring, Providing Information to Patients, and Quality of Dialysis Center

Care and Operations. Experience with Care is directly impacted by Health Services Usage and Health Behaviors, as well as Quality of Care.

The last set of constructs is Outcomes, which includes patients' HRQOL and their Satisfaction with Care. HRQOOL assessment is made with standardized instruments that include multiple aspects of health, including physical, mental, and social health. Although RRT's are life saving for patients with ESRD, it has been argued that their primary clinical objective is to improve patients' HRQOL.[61,62] A significant effort is made to improve to HRQOL in dialysis patients especially. In their Conditions for Coverage (42 CFR §494.90), CMS mandated that each dialysis patient's physical and mental health be monitored, and this often occurs with the use of a standardized HRQOL measure.[63] The patient reports of HRQOL are then used to create individually-tailored interventions that focus on the areas where the patient's HRQOL needs most improvement. In this conceptual model, HRQOL is directly impacted by Experience with Care. For example, if dialysis patients do not receive the information they need about caring for their dialysis access site, they may develop an infection that impacts HRQOL. In turn, HRQOL impacts Health Services Usage and Health Behaviors. For example, if dialysis patients have depressive symptoms, they may need to use mental health services. Similarly, depression may impact dialysis patients' Health Behaviors, like adherence to their medication or exercise regimen.

Dialysis patient's Satisfaction with Care is an important outcome of RRT and can be used to help understand the quality and effectiveness of kidney care. In assessments of kidney care, many studies have successfully assessed the extent to which care meets their expectations,[64-67] instead of objective dimensions of care that represent patients' experiences with care. In the proposed conceptual model, Satisfaction with Care is directly impacted by Experience with Care,

14

since the actual experiences with care either will or will not be consistent with the patient's expectations. For example, the satisfaction questionnaire developed for the Choices for Health Outcomes in Caring for ESRD (CHOICE) study asked patients to rate the quality of different aspects of care as "Poor" to "Excellent" represented in statements like, "How often the nephrologist sees you," "The amount of information you are being given to help you choose between hemodialysis and peritoneal dialysis," and "How much fluid is removed during your dialysis session."[65] Each of these statements on its own represents an experience with care, or an objective feature of the care patients have received. However, when rated on a normative scale from "Poor" to "Excellent", the scale elicits the extent to which these objective aspects of care meet the patients' expectations and therefore capture information about the patient's satisfaction about care. Hays and Arnold put forth a theory of patient satisfaction with medical care that defines this construct as an attitude as opposed to belief or behavioral intention; therefore, satisfaction with care is thought of as the level of satisfaction with their care.[68]

In the measurement part of this conceptual model, four types of PRMs are identified: 1) Needs Assessments; 2) Behavioral Reports; 3) Reports about Experiences with Care; and 4) Patient-Reported Outcomes (PROs). As in the original Fung and Hays conceptual model, implementation of PRMs in clinical settings is guided by IDEAL patient-clinician encounter framework, which has the following steps: 1) **I**dentify the health problem; 2) **D**iscuss the health problem with the patient and offer options for treatment; 3) **E**nact action to co-create the treatment plan with the patient; 4) take **A**ction on the treatment plan; 5) **L**earn about the effects of the treatment plan.[11] Though HRQOL is a PRO, along with biological assessments, a baseline HRQOL assessment can be made to help identify the health problem. Once the health problem is identified, a Needs Assessment can be used in a clinical context to elicit patients' decision-

making and preferences around the treatment approach. Then, along with reports about patients'

use of health services and their health behaviors assessed with Behavioral Reports, patient

reports about their decision-making and preferences for care can be used to facilitate discussions

about, and co-creation of, the treatment approach. Once the treatment approach is enacted, its

objective characteristics can be monitored by Reports about Care Experiences. Finally, the

effects of treatment approach can be assessed with PROs, including a follow-up HRQOL

assessment to determine if change has occurred since the baseline assessment. Additionally,

patients' Satisfaction with Care can also be used to determine if the treatment is successful or

needs to be changed.

2.3. THEORETICAL UNDERPINNING OF KEY CONSTRUCTS
2.3.1 TREATMENT DECISION-MAKING THEORY

Many theories explain patients' decision-making about their treatment and the behaviors

associated with seeking these treatments. Among the most popular theories include the Health

Belief Model, the Theory of Reasoned Action/Planned Behavior, and the Precaution Adoption

Process Model.[69] Another of the primary theories used to understand how patients make

decisions about their treatment options, and to take health behaviors, is the Transtheoretical

Model of Behavior Change (TTM). The TTM includes several key constructs, including stages

of change, processes of change, decisional balance, and self-efficacy.[70] Stages of change

represent the level of willingness a person has to change. There are typically 5 stages of change

specified: *Precontemplation* (not considering change), *Contemplation* (considering change in the

next 6 months), *Preparation* (preparing for change in the next 30 days), *Action* (taken action

toward change), and *Maintenance* (sustained action for 6 months).[70] Another key construct,

Decisional Balance, is the differential weight of the Pros and Cons of changing a health

behavior. Finally, *Self-Efficacy* is the confidence a person has to change even when confronted with challenges.

These constructs work in tandem to describe a person's change in health behavior, primarily around the relationship between changes in Decisional Balance associated with changes in stage of change. First, the balance of Pros and Cons varies by stage such that in the *Precontemplation* stage the Cons outweigh the Pros, the Pros begin to overtake the Cons in either the *Contemplation* or *Preparation* stage, and the Pros outweigh the Cons in the Action stage.[71] A similar pattern often emerges for Self-Efficacy. (Figure 5) In a meta-analysis, it was found that a patient's movement from *Precontemplation* to *Action* is associated with a one standard deviation increase in the perceived Pros to change (part of Decisional Balance) while a change from *Contemplation* to *Action* is associated with a one-half standard deviation decrease in the perceived Cons to change.[71]

Figure 5. Changes in Decisional Balance & Self-Efficacy by Stage of Readiness
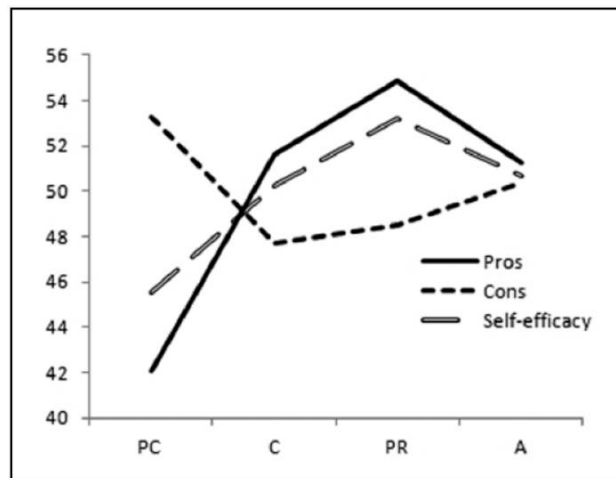
Published in: Waterman AD, Robbins ML, Paiva AL, et al. Measuring kidney patients' motivation to pursue living donor kidney transplant: development of stage of change, decisional balance and self-efficacy measures. *Journal of health psychology.* 2015;20(2):210-221.

As a theory of change (vs. an explanatory theory),[72] the TTM holds that efforts to help people change health behaviors must recognize which stage of change the person is in, and

provide strategies to match that stage. The previous work on the relationship between the TTM's

constructs helps to guide ways to intervene with patients in particular stages. For instance, in the

circumstance of a smoking cessation intervention, asking a person in the *Precontemplation* stage

to quit smoking would not likely respond successfully to a request to quit smoking today.

Instead, asking that person to learn more about and consider the Pros of quitting smoking might

be a more appropriate interventional step.

Interestingly, for some CKD treatment-seeking behaviors, namely transplantation,

achieving the treatment, a typical marker for the Action and Maintenance stages, is not totally in

the patient's control. For instance, patients may play a large role in whether or not they join the

waiting list for transplant, or how intently they seek a living donor, but factors out of their

control may determine whether or not they receive a transplant once on the waitlist, or if anyone

agrees to donate their kidneys. For this reason, current research in stages of change for CKD

patients seeking transplant have determined that the Action and Maintenance stages for pursuit

of deceased donor transplantation are centered on joining the waitlist, not receiving a

transplant.[73] Further, the Action stage for pursuit of living donor transplantation regards taking

any action toward living donor transplantation, and there is no Maintenance stage for this

behavior.[74] These examples show how the definitions of stages of change are dependent on the

specific behavior under consideration.

2.3.2 MEDICATION ADHERENCE THEORY

In any chronic or infectious condition, medication adherence is influenced by many

factors, and CKD is no exception. Medication adherence can be defined as medication taking

that accurately and precisely corresponds with a prescribed medication regimen.[57] The broader

construct of adherence is comprised of two smaller behaviors. The first, persistence refers to

initiation of medication-taking to its discontinuation; i.e., a persistent patient is one who has begun taking the medication and has not stopped. The second, execution, refers to the accuracy and precision with which the regimen is followed. Patients may be taking medication, but not in the prescribed doses, with the prescribed frequency, or at the prescribed times. More specific conceptualizations of types of adherence can include "percent adherence", or the proportion of prescribed doses that are taken, and "dose-timing adherence ", which refers to the amount of error in adherence to the schedule of doses prescribed.[75,76]

Non-adherence to key medications, like anti-hypertensives, is associated with worse clinical outcomes for CKD patients. For example, in study of 295 CKD patients in Thailand, lower adherence reported on the 8-item Morisky Medication Adherence Scale (MMAS-8) was associated with a 2-fold increase in CKD progression.[77] Among dialysis patients, common medications include phosphate binders and hypertension control medications. Patients who have received a transplant must take multiple immunosuppressant medications to prevent rejection of the transplanted graft. A recent study of medication burden among 1,238 hemodialysis patients found the average daily medication burden was 9.7 pills, and only 48% of patients were adherent to their medication regimen.[78] This study also measured patient-perceived burden of medication therapy (e.g., bothered by the number of pills in the regimen, bothered by medication side effects) and found that high medication burden was the best predictor of medication non-adherence (odds ratio: 0.48; 95% CI: 0.35-0.66). In a recent review of 44 studies of medication adherence among hemodialysis patients, non-adherence ranged widely, from 13%-99%.[79] Another recent review of 25 studies of medication adherence among peritoneal patients, non-adherence ranged widely across studies, from 3%-53%.[80]

Medication non-adherence among kidney transplantation patients is also a major issue, with many medications required to prevent the transplanted graft from rejecting, and eventually losing the transplanted kidney. Indeed, a leading cause of graft rejection and loss is non-adherence to the immunosuppression medication regimen.[81,82] The complexity of the medication regimen may be a critical factor leading to medication non-adherence among kidney transplant patients. One study estimated the daily pill burden among kidney transplant recipients to be 16-25 pills taken daily.[83] Non-adherence rates are also very high among kidney transplant patients. A meta-analysis found that the median proportion of non-adherent patients among kidney transplant recipients was 22%, but was estimated to be 67% at the top of the range.[82]

Non-adherence may be intentional or unintentional. One interpretation of intentional non-adherence is supported by social cognition theories,[84,85] such as the health belief model,[53,86] theory of reasoned action/theory of planned behavior,[87,88] and social cognitive/learning theory.[50-52] These theories and models explain non-adherence as the execution of intentional behaviors that result from rational decision-making by patients about whether or not they should take their medication. Typically, this decision-making is influenced by careful consideration of information available to the patient. Certainly, some medication non-adherence is intentional. For example, patients may develop a belief that the medication is not effective for them, and based on that belief, discontinue the medication or fail to execute the regimen as prescribed. One way to understand intentional non-adherence is through social-learning theories that suggest patients may weigh the costs or perceived barriers to taking medication before deciding whether or not to do so.[89] On the other hand, many times non-adherence to medication is unintentional. Unintentional medication non-adherence is not the result of specific beliefs about medication-

taking, but instead is tied to factors like misunderstanding the medication regimen or facing other barriers that prevent intended medication taking.[84,85]

Much of the research fails to distinguish between intentional and unintentional non-adherence,[84,85] leading to incorrect or inappropriate conclusions about which types of adherence-increasing interventions to pursue. A study of 201 home-based peritoneal dialysis patients found that intentional non-adherence was more common for dialysis therapy (e.g., skipped or shortened appointments) than for medication, whereas the opposite was true for unintentional non-adherence.[90] A recent study of adherence to phosphate binders among 79 hemodialysis patients found that 61% of patients reported unintentional non-adherence and 48% reported intentional non-adherence.[91] The same study also found that patient-reported lower non-adherence was associated with lower odds of phosphate control (odds ratio: 0.71; 95% CI: 0.49-1.00). Another study of adherence to phosphate binders in 135 hemodialysis patients found that, over 2 months, 78% of prescribed doses were taken.[92] In this study, the best predictors of perfect adherence were social support and HRQOL, which together accounted for 26% of the variance in adherence. A recent analysis of Medicare data by Wang and colleagues of 30,933 in-center hemodialysis patients found that 50% reported a discontinuation of phosphate binders.[93] The most common reasons for discontinuation included perceiving that phosphorus was under control (27%) and side effects (11%).

In addition to the distinction between intentional and unintentional non-adherence, standard health behavior theories, like the Health Belief Model,[94] have been applied to understanding whether a patient will be adherent to their medication. Notably, the TTM applies its general theory about health behavior to understanding how ready patients are to adopt and adhere to their medication regimen. In a theoretical exploration, developers of the TTM posited

that stages of change for adherence could be measured in the following way.[95] First, adherence is defined for the respondent as "taking pills as directed," then the question is asked: "Do you consistently take all your pills as directed by your doctor?" If patients respond "No," then they are asked if they intend to do so in the next 6 months or in the next 30 days. If they do not intend to in the next six months, they are coded in the *Precontemplation* stage; if they intend to do so in the next six months but not 30 days, they are coded in the *Contemplation* stage; and if they intend to do so within the next 30 days, they are coded in the *Preparation* stage. Patients who respond "Yes" to the original question and who say they have been taking pills as directly for six months or less are coded in the *Action* stage and those who say they have been doing so for more than six months are coded in the *Maintenance* stage. This stage of change algorithm can be used in studies applying the TTM to medication adherence.

Decisional Balance and Self-Efficacy are also applied to adherence[95]. In the context of adherence, self-efficacy describes the patient's confidence to be adherent in the face of barriers. For example, whether the patient can stay adherent even when they are very busy or very sick would constitute a potential marker of self-efficacy. Regarding medication adherence, decisional Balance describes the weighing of Pros and Cons to adherence. Examples of Pros would be staying healthy, feeling better, being more available to one's family, and being able to work more. Examples of Cons would be the cost of medications and their side-effects. In the context of a study applying the TTM to adherence, each of these constructs would be assessed separately, often in the context of testing an intervention to increase adherence. Then a patient's individual stage of change, perceived Pros and Cons, and Self-Efficacy would be taken into account to tailor recommendations about how to improve adherence.

One example of a study in this style includes a randomized controlled trial investigating a computerized health education intervention aiming to increase patients' adherence to lipid-lowering medications among 1,227 hypertensive patients in New England.[96] The primary outcome of the trial was the stage of change to adhere to their medication regimen as defined above, applied to the lipid-lowering medications, i.e., "Do you consistently take all your lipid-lowering medications as directed by your doctor?" They also employed another measure of adherence wherein five items representing common ways patient may be non-adherent to medications were represented (e.g., forgotten to take a dose, stopped taking medications because the patient felt better) with categorical responses summed to create a scale ranging from less to more non-adherent. Patients receiving the intervention were significantly more likely to increase their stage of change to be adherent and had significantly lower non-adherence than patients in the control group.

A second study by the same investigators also examined the impact of a very similar intervention to increase adherence to lipid-lowering medications with 404 patients in New England, and came to similar conclusions about the efficacy of the TTM-based education to increase medication adherence.[97] Each of these trials used a population-based approach wherein participants were proactively recruited from a pool of patients served by a large health maintenance organization (HMO) who were not necessarily looking for assistance with improving their medication adherence, as opposed to a clinical trial where patients seeking treatment are randomized to usual care or an experimental treatment. This approach demonstrates the ability of the TTM to be successful in improving medication adherence as a public health objective, where it is commonly thought of as a medical concern only. In the second of these studies,[97] ancillary impacts on improved exercise participation and reduced

intake of dietary fat was observed as well, indicating the capacity for TTM-based interventions to target multiple behaviors.

In line with the TTM approach to understanding adherence, many patient-related factors have been examined as influences on medication non-adherence among CKD patients, and dialysis patients specifically. A study of 502 hemodialysis patients found that modifiable factors representing both intentional and unintentional medication adherence were associated with higher odds of phosphate binder non-adherence.[98] The odds of non-adherence were very high for patients who reported discontinuing phosphate binders because they felt worse (OR: 11.04; 95% CI: 1.69-68.03), felt better (OR: 7.09; 95% CI: 1.45-14.25), or because of a lack of understanding of the medication regimen (OR: 7.09; 95% CI: 2.10-23.95).

Depression is a key clinical comorbidity to CKD that may be significant risk factor for medication non-adherence, especially among dialysis patients.[57] Cukor and colleagues compared the association of depression and medication non-adherence among dialysis (n=65) and transplant (n=94) patients and found that the correlation between adherence levels and depression was stronger for dialysis patients ($r = -47$; $p<0.001$) than for transplant patients ($r= -38$; $p<0.001$).[99] In another study among a sample of 2,180 hypertensive patients, those with depression had nearly a 2 fold increase in odds of medication non-adherence (odds ratio: 1.96; 95% CI: 1.30-2.70).[100]

A review conducted by Ghimere and colleagues found that the most common patient-level correlates of medication non-adherence were younger age, non-White race, smoking, reporting that one's kidney disease impacted their family, and living alone.[79] The review of medication non-adherence among peritoneal dialysis patients found that the most common

patient-level factors associated with non-adherence were lower age, lower socioeconomic status, non-White race, and smoking.[80]

2.3.3 HEALTH-RELATED QUALITY OF LIFE THEORY

Cella and Tulsky identified three central reasons to assess HRQOL: to assess the need for rehabilitation; to assess the efficacy or effectiveness of a medical treatment; and as a predictor of response to future treatments. Additionally, as the use of HRQOL measures has increased and norms for general populations and key patient populations have been generated, HRQOL is an approach to tracking population health.[21,101-103] Among the major projects that have generated HRQOL, it is often defined, in a basic form, as functioning and well-being, or what the patient can do and how the patient feels.[18] Multidimensional HROQL in this approach entails physical, emotional, and social health, the general domains used in the PROMIS project.[21] Within these primary domains are several important subdomains, including role functioning, sexual health, ability to participate in leisure, spiritual health, and family functioning.

HRQOL measures come in many shapes and sizes. From the start of theory-building around HRQOL, its scholars began categorizing the types of HRQOL measures by the methods used to generate them, the population they are intended to reflect, and how the ratings of HRQOL resulting from them can be used. There are two major types of HRQOL measures: profile and preference-based.[104,105] These two types of measures can often, but not always, be distinguished by the methods used to generate them and the types of applications for which they are most appropriate. Profile measures tend to be generated through psychometric approaches (e.g., factor analyses). Preference-based measures tend to be generated for, and are most applicable to, economic analyses, especially decision analysis. A profile measure captures all relevant domains of HRQOL and often features scales for individual domains of HRQOL (e.g.,

25

physical health, mental health). A preference-based measure, otherwise known as a utility

measure,[104] captures patients' preferences for levels of health and potential treatments. The

single utility HRQOL score will often be the difference between the expected benefit to HRQOL

from a new treatment subtracting any burdens associated with the treatment, like side-effects.

Another key distinction between types of HRQOL measures regards the distinction

between generic and targeted measures.[104-106] Generic measures of HRQOL are intended to be

appropriate for all disease conditions. Targeted measures, on the other hand, are aimed toward a

specific subgroup of the population, either in terms of a specific disease condition or a specific

age group or racial group. Targeted measures may be modifications of generic measures that

make minor adjustments to the items to suit the targeted population, or may be completed

tailored to the population of interest, with subscales and items that apply only to their concerns

and interests.

Generic PROs tend to have the benefit of being contained in one instrument that

incorporate multiple subdomains of the general PROs. Generic PROs are also more widely used,

providing the opportunity to compare scores across applications in different interventions and

populations. On the other hand, a significant weakness of generic PROs regards its omission of

specific items that may be most relevant to specific populations. Targeted PROs have the benefit

of including content that is highly relevant to specific patient populations, but may be limited in

their application. Further, since the application of targeted instruments is often limited to specific

populations, the psychometric properties of targeted instruments may not be understood as well

as those of generic instruments. Often times, using a combination of generic and targeted PROs

is used to capture the benefits of both.

As noted above, the PROMIS project represents the gold standard in HRQOL research to date. As a set of generic measures for HRQOL, PROMIS measures cover the major dimensions of HRQOL.[21] In its first wave, PROMIS brought together multiple content area experts, methodological experts, and clinicians from academia, as well as the NIH, to conduct a series of nationally-representative surveys with 1,532 patients from 6 primary study sites representing both a clinic sample (n=803) and the general population (n=729). Additionally, internet survey panels were used to recruit larger samples of patients from the general population (n=12,521) and with chronic conditions, including heart disease, cancer, rheumatoid arthritis, osteoarthritis, psychiatric illnesses, chronic obstructive pulmonary disease, spinal cord injury, among others (n=7,080).[24] The internet-panel sample was drawn through a sample-matching procedure that draws a random sample from the target population then selects appropriate matches from the internet panel; the target population was defined in terms of desired distributions of demographic characteristics: gender (50% female, 50% male), age (20% each of age groups, including 18-29, 30-44, 45-59, 60-74, $\geq$75), race/ethnicity (12.3% Black and 12.5% Hispanic), and education (10% with < high school education). From this sample, a subsample representing the U.S. general population was derived.

PROMIS used a rigorous approach to PRO development, starting with broad efforts around domain mapping, archival data analysis, and qualitative item review. The domain-mapping protocol began with the WHO conceptual framework of health with physical, mental, and social health as its broadest components.[21,24] To come to consensus on domains, the investigative team then conducted a Delphi process. From this process, 5 subdomains were identified, including physical functioning, fatigue, pain, emotional distress, and social role participation. Then, the PROMIS team identified large datasets with previously-developed PROs

and developed a protocol for IRT analyses with the objective of improving the understanding of the 5 subdomains' dimensionality and identifying the best candidate items to include in item banks. Approximately 7,000 items were chosen for review. Finally, a qualitative review of the selected items was conducted in order to classify and revise of each item – binning and winnowing. Additionally, focus groups with patients and providers were conducted to further validate the PROMIS conceptual framework. Additionally, since this initial phase, some modifications were made to the PROMIS conceptual framework.[22]

After these initial activities, the item banks were created. Item banks, defined by PROMIS as a set of "carefully calibrated questions that define and quantify a common concept and thus provide an operational definition of a trait" (p. 2)[21] were created for all key domains in the PROMIS conceptual framework. The number of items ranged from 12 (satisfaction with participation in discretionary social activities) to 124 (physical function). A first set of item calibration analyses using IRT models revealed a typical trait range of -4 to +4 on the logit scale. However, for more efficient interpretation, all PROMIS scales were put on the T-Score metric. The T-score metric sets the mean at 50 and standard deviation at 10 in a referent group; in PROMIS, the referent group is the United States general population. Higher scores are interpreted as higher levels of the construct being measured (e.g., higher physical functioning). Due to these properties, T-score are very easy to interpret.

All correlations between domain short forms and full item banks were high at $r>0.95$. For example, the 124 item physical function item bank and 10 item short form were correlated at $r=0.96$. Reliabilities for each item bank were very high, especially for T-scores above 40, where reliabilities were >0.90, with some exceptions. PROMIS represents the gold standard for PRMs and PROs and offers a set of generic health measures that can be used in any field.

2.3.4 PSYCHOMETRIC APPROACHES TO PRM/PRO DEVELOPMENT AND VALIDATION

Classical Test Theory (CTT) is based on true score theory, which holds that a person's

observed score (X) on a PRM or PRO is the sum of their level of trait (or true score) that is

measured by the instrument (T) and error (E),[107] or:

$$X = T + E$$

The errors in this equation (E) are assumed to be normally distributed and uncorrelated with the

true score. Reliability in CTT is commonly thought of as the ability of an item or a scale to

reveal the same score for the same trait level; perhaps within the same individual at different

points in time, perhaps for two or more individuals with the same trait levels. Formally,

reliability (ρ) in CTT is defined in terms of the ratio of variances (σ^2) of true scores to total

observed scores, whether these are for a single item or a sum score across a set of items. If Y is a

score created by summing multiple items, this ratio is:

$$\rho_Y = \frac{\sigma_T^2}{\sigma_Y^2}$$

This equation means that scales with a high reliability will have a high proportion of true score

variance to total variance, implying low error variance. There are many types of reliability, for

example, internal reliability consistency, which is measured with a statistic ranging from 0-1

with the following standards for interpretation at the group level: 0.70, >0.80, and >0.90

indicating acceptable, good, and excellent reliability.[108] Additionally, for individual use, a

minimum of 0.90 is required.

One common formulation of internal consistency reliability is Cronbach's α. Cronbach's

α is defined as:

$$\frac{K}{K-1}\left(1 - \frac{\Sigma_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_X^2}\right)$$

where K is the number of items in the scale, σ_X^2 is the observed variance of the total test scores, and $\sigma_{Y_i}^2$ is the variance of the ith item in the scale.[109] Cronbach's α can be understood as the correlation between two scales measuring the same construct.

In CTT, a first examination of the dimensionality of a scale can be performed by calculating the item-to-total correlation for scales (often corrected for overlap of items across scales)[107]. After examining item-to-total correlations, a common next step is factor analyses, including exploratory and confirmatory factor analyses. In exploratory factor analyses, the number of factors in the scale are not known *a priori*, and this type of analysis is often used to generate first impressions about the number of factors the scale may have. Confirmatory factor analyses then may be used to try to confirm the factor, or dimensional, structure suggested by the exploratory analyses. To ensure that the underlying constructs examined in factor models is equally meaningful for all patient subgroups (e.g., racial groups, gender groups), and that items equivalently measure underlying constructs for these subgroups, it is necessary to establish measurement invariance, i.e., that the construct is measured the same way for different groups of patients.

2.4. USE OF PATIENT-REPORTED OUTCOMES IN CKD
2.4.1 TREATMENT DECISION-MAKING

To date, few investigators have developed and evaluated measures of treatment decision-making for CKD patients, especially regarding the decision about whether or not to get a transplant. A handful of studies employ self-created measures for which no psychometric analyses (e.g., investigation of factor structure, construct validity tests) are reported. Examples

include measures of transplant knowledge,[110-112] readiness for transplant in general,[110-113] willingness to take specific behaviors toward living donor kidney transplant (LDKT),[114] and transplant concerns.[110,111] CKD studies may also use generic, non-condition targeted decision-making measures, such as O'Connor and colleagues' Decisional Self-Efficacy measure.[115,116] Therefore, there is a significant gap in the CKD research about which measures of treatment decision-making have the best psychometric properties.

Among measures about which development and evaluation information has been reported, Ismail and colleagues developed the Rotterdam Renal Replacement Knowledge Test (R3K-T), a CKD treatment knowledge scale among dialysis and transplant patients in the United States using IRT methods and found evidence for a two factor, 21 item scale with subscales for "dialysis and transplantation" and "living donation".[117] Wright and colleagues developed a 28-item measure of CKD knowledge in a study that examined reliability [Kuder-Richardson (KR)-20 = 0.72] and construct validity via associations with related constructs.[118] Prakash and colleagues created a set of dialysis decision-making measures using qualitative methods and exploratory factor analysis. Though no results from the factor analyses were reported, reliabilities for the following scales were reported: dialysis modality knowledge, $\alpha = 0.91$; lifestyle barriers, $\alpha = 0.71$; and CKD self-efficacy, $\alpha = 0.39$.

Waterman and colleagues have developed and conducted initial evaluation of a set of TTM-based Decisional Balance and Self-Efficacy measures for both deceased and living donor kidney transplant (four measures in total). Each of these measures is currently being employed in two randomized controlled trials that seek to increase kidney patients' knowledge of transplant, readiness for transplant, and pursuit of transplant.[119,120] These measures were developed with the

31

input of kidney patients and clinical experts in nephrology. The initial analysis plan followed that typical for these constructs in other health conditions.[121,122]

For the living donor transplant Decisional Balance measure, the factor structure was verified with exploratory and confirmatory factor analyses. Again, a two factor solution exhibited the best fit to the data, yielding the Pros and Cons subscales (average item loadings of 0.6-0.8). Both subscales evidenced "good" reliability (Pros = 0.86, Cons = 0.80).[74] For the living donor transplant Self-Efficacy measure, the factor structure of this scale was also verified with exploratory and confirmatory factor analyses, demonstrating a good fit for a one factor solution (loadings ranging between 0.61-0.91).[74] The Cronbach's alpha was 0.88, indicating "good" internal consistency reliability.

For the deceased donor transplant Decisional Balance measure, the factor structure of this measure was determined with exploratory and verified with confirmatory factor analyses, in which this scale was reduced from 16 to 12 items. A two-factor solution exhibited the best fit, yielding the Pros and Cons subscales (average item loadings of 0.6-0.8). The Pros subscale had a Cronbach's alpha of 0.75 and the Cons subscale had a Cronbach's alpha 0.76, evidencing adequate internal consistency reliability for both scales.[73] For the deceased donor transplant Self-Efficacy measure, the factor structure of this scale was verified with exploratory and confirmatory factor analyses, demonstrating a good fit for a one factor solution (loadings ranging between 0.50-0.73).[73] The alpha was 0.85, indicating "good" internal consistency reliability. Waterman, et al's measures represent the state of the science in CKD treatment decision-making PRMs, but have not yet been examined among a sample of dialysis patients not seeking transplant. Additionally, while measurement invariance between racial groups, genders, and

levels of education have been established for the LDKT decision-making measures,[123] no such study has been performed for the DDKT decision-making measures.

2.4.2 MEDICATION ADHERENCE

In studies of medication non-adherence in dialysis, several patient-reported instruments have been used. Ghimere and colleagues' review of medication non-adherence in dialysis revealed the use of the following instruments in at least one study: Brief Medication Questionnaire (BMQ); Medication Adherence Report Scale (MARS); Morisky, Green and Levine medication adherence scale (MGL); MMAS-8; and Simplified Medication Adherence Questionnaire (SMAQ).[79] Each of these instruments has been validated either as generic instruments meant to be applicable to any chronic or infectious condition requiring medication, or as a targeted instrument for CKD patients.

The BMQ is a widely used, generic patient-reported medication adherence screening tool.[124] The BMQ is comprised of 3 independent adherence screening tools with 9 total items: a 5-item Regimen screener, a 2-item medication Beliefs screener, and a 2-item Recall screener. A validation study among 43 patients taking ACE inhibitors examined criterion validity with an electronic medication event monitoring device used as the criterion, and found that each of the screeners often was sensitive and specific, and overall accuracy for either sporadic and repeated non-adherence, though tended not to have these properties for both types of non-adherence; for example, the regimen screener had 80% sensitivity to repeat non-adherence but 0% sensitivity to sporadic non-adherence. However, no psychometrically-oriented analyses were conducted to determine the factor structure of the instrument.

The Medication Adherence Report Scale (MARS) is a generic medication adherence measure that has been validated among multiple chronic and infectious conditions.[125-129] For example, validity of an asthma-focused version of the MARS was established through examining correlations between it and electronic medication event monitoring-reported adherence.[125] However, though reliabilities have been generated for the MARS among several conditions, including hemodialysis patients,[129] and descriptions of MARS scores have been reported in hemodialysis patients,[130-132] its dimensionality is not reported in general or hemodialysis populations.

The Simplified Medication Adherence Questionnaire (SMAQ) was originally developed and evaluated among HIV patients using 3 items from the MMAS-4 and adding 3 items that ask specifically about the number of missed doses in the past 24 hours, frequency in general of missed doses, and number of days with missed doses.[133] Other than reporting an internal consistency reliability of 0.75, no other psychometric analyses were conducted. Instead, validation was evidence through ability to distinguish between adherence levels as measured by an electronic medication event monitoring device, and between clinical measures of disease control. When used with CKD patients, the SMAQ has been found to distinguish between phosphate levels, though not between blood pressure levels.[134]

The 4-item Morisky Green Levine Test (MGL), otherwise known as the Medication Adherence Questionnaire (MAQ), is among the most widely-used self-reported adherence scales, cited by >3,500 peer-reviewed studies.[135] Developed in 1986, the Morisky, Green and Levine medication adherence scale's dimensionality was established with principal components analysis and was bolstered by high corrected item-to-total correlations, ranging from 0.479-0.561.[135] The internal consistency reliability for this scale was only 0.61. More recent analyses have found

evidence for two factors within this scale, one representing intentional non-adherence and one representing unintentional non-adherence.[136] Despite its overall widespread use, this scale has only been used a handful of times among CKD patients.[78,137,138] However, these studies do not report psychometric characteristics of the MGL medication adherence scale for CKD patients.

More recently, the MMAS-8, an 8 item version of the Morisky adherence scale, was developed. The MMAS-8 is a generic assessment of medication-taking behavior and has been used in a number of health conditions, both chronic and infectious diseases. It has proven reliability (internal consistency = .83) for subjects under care for high blood pressure and has shown support for concurrent and predictive validity[139] (criterion-related validity) with blood pressure control, HgA1c levels for subjects treated for diabetes, HDL levels for high cholesterol, and high correlations with social support, mental health and satisfaction with health care. The MMAS-8 assesses both unintentional (such as forgetting) and intentional (stopping when you feel better) types of adherence. No studies were found that report psychometric properties of the MMAS-8 in CKD patients.

2.4.3 HEALTH RELATED QUALITY OF LIFE

HRQOL is a key outcome in CKD, and is often used as a justification for making decisions about which treatments are appropriate for patients. In general, transplantation, especially kidney transplantation, is thought to offer better HRQOL in comparison to all forms of dialysis,[140,141] a finding that has been consistent in the medical literature for more than 20 years.[142,143] In the field of CKD, both generic and disease-targeted scales have been used, both offering important elements to the understanding of HRQOL in the CKD population.

Several generic HRQOL scales have been used with CKD patients having received dialysis and transplant, including but not limited to: the Spitzer QL-index, the Nottingham Health Profile, the Campbell Index of Well-Being, Cantril's Self-Anchoring Scale, and the Life Satisfaction Scale.[2,141,144] Additionally, the Karnofsky Index, Sickness Impact Profile, Medical Outcome Study (MOS) Short Form 36-Item Health Survey (SF-36), EuroQOL 5D scale (EQ-5D), World Health Organization Quality of Life Questionnaire, and Time Trade-Off approaches have been widely used.[2,141,144] Additionally, instruments aimed at mental health, like the Beck Depression Index, CES-D Depression scale, Affect Balance Scale, and the Self-Esteem Inventory, as well as social health scales have been used.[2,144] Systematic reviews of the use of these instruments among various CKD populations have been conducted, so their results will not be rehashed here[2,141].

However, of particular note, in the review of the literature, only one study of CKD stage 1-5 patients has included assessment of the PROMIS measure. The Midwest Pediatric Nephrology Consortium recruited 223 patients aged 8-17 from their respective clinics and assessed the PROMIS depression, anxiety, social-peer relationships, pain interference, fatigue, mobility, and upper extremity function short forms.[145] Though no additional psychometric analyses were performed among this sample of patients, effect sizes for differences across important clinical subgroups were reported; e.g., CKD stage, hospitalized or not, presence of medical comorbidities.

The most widely-used targeted scales are the Kidney Disease Quality of Life (KDQOL) scales developed by Hays and colleagues in 1994.[1] The original KDQOL instrument, also referred to as the KDQOL-SF, combines the SF-36 as a measure of generic HRQOL with 11 kidney disease targeted subscales, including Symptoms/Problems of Kidney disease (34 items),

36

Effects of Kidney Disease (20 items), Burden of Kidney Disease (4 items), Work Status (4 items), Cognitive Function (6 items), Quality of Social Interactions (4 items), Sexual Function (4 items), Sleep (9 items), Social Support (4 items), Dialysis Staff Encouragement (6 items), and Patient Satisfaction (2 items). A psychometric and validation study of the KDQOL-SF, these targeted subscales all had acceptable to excellent internal consistency reliabilities,[1] ranging from 0.76 (Sleep) to 0.94 (Effects of Kidney Disease). A factor analysis revealed that these 11 subscales, in addition to subscales from the SF-36, all loaded moderately to highly on 4 factors representing physical health, mental health, disease-targeted, and patient satisfaction. Construct validity for the subscales was evidenced through significant associations with patient-reported good days, patient –reported bad days, self-rating of quality of life compared to those without kidney disease, ability to do everything the patient wants to do, number of disability days in the last 30 days, and overall rating of health.

Though the KDQOL-SF has demonstrated attractive psychometric and measurement properties, it was too long to administer in many settings, especially in clinical contexts. Clinical application of the KDQOL instruments is an important concern. According to their Conditions for Coverage (§494.90), the Centers for Medicare and Medicaid Services (CMS) require that all dialysis patients be assessed with a physical and mental health tool, and the KDQOL-36 is the preferred tool for this purpose.[63] Then, patients' responses to the KDQOL-36 must be used by dialysis providers to develop a personalized health plan for each patient. The demands of clinical dialysis practice put significant limitations on the time available for health assessment, creating the need for a brief KDQOL assessment. With this need in mind, a shorter version of the KDQOL, the KDQOL-36, has been deemed more appropriate for clinical use.

The KDQOL-36 employs the SF-12 as a measure of global health and adds disease-specific subscales tailored to issues faced by kidney patients. The SF-12 is a brief version of its parent instrument, the SF-36. As its name implies, the SF-36 has 36 items covering several domains, each with its own subscale that has demonstrated internal consistency reliability: including physical functioning (10 items: $\alpha = 0.93$), role- physical (4 items: $\alpha = 0.89$), bodily pain (2 items: $\alpha = 0.90$), general health (5 items: $\alpha = 0.81$), vitality (4 items: $\alpha = 0.86$), social functioning (2 items: $\alpha = 0.68$), role-emotional (3 items: $\alpha = 0.86$), and mental health (5 items: $\alpha = 0.84$), with another item for health transitions (1 item).[146] The most commonly reported score from the SF-36 are two component scores, the Physical Health Component ($\alpha = 0.92$) and the Mental Health Component ($\alpha = 0.88$). The SF-12 has the same subscales as the SF-36, drawing 1-2 items from each.[147] Figure 6, shows the measurement model used for the SF-36 and SF-12, with boxed items indicating those chosen for the SF-12.

Figure 6. SF-36/SF-12 Measurement Model
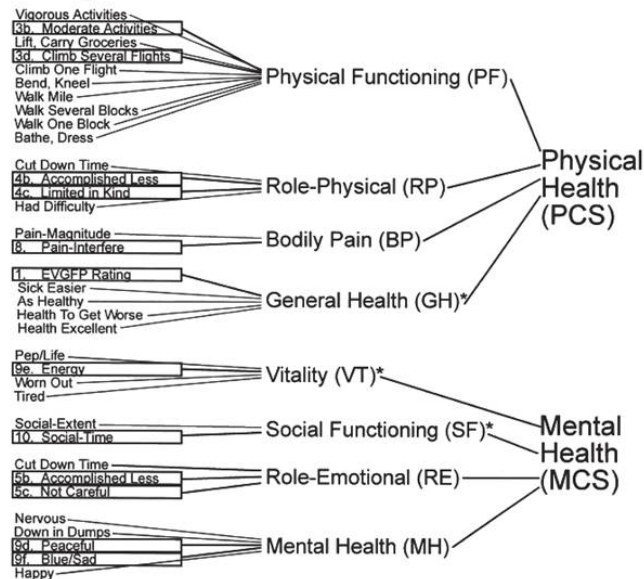
Published in Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *Journal of clinical epidemiology.* 1998;51(11):903-912.

In addition to the SF-12, the KDQOL-36 includes 3 kidney-disease targeted scales, including Burdens of Kidney Disease, Symptoms and Problems with Kidney Disease, and Effects of Kidney Disease. These items are subsets of the KDQOL-SF. Though it was created in 1994, to date, there has been little examination of the KDQOL-36's psychometric characteristics in a large, U.S. sample.

CHAPTER 3. SPECIFIC AIMS

The overall objective of this dissertation is to evaluate the psychometric properties of patient-reported measures commonly-used with kidney patients in research and in clinical settings. This examination includes general psychometric evaluations and evaluations of whether measures are equally reliable and valid for subgroups of patients. These analyses will contribute to the evidence base on the psychometric performance of patient-reported measures used in kidney disease and yield information on the use of these measures for regulatory oversight and quality improvement efforts. The specific aims of the dissertation are:

Aim 1. **To describe the psychometric properties of existing health decision-making, health-related quality of life, and medication adherence measures in the CKD population.**

As reviewed above, PRMs have played a significant role in understanding how CKD patients make decisions about treatments available to them, whether or not patients are compliant with the treatments, and in determining the effectiveness of these treatments in improving patients' lives. With PRMs playing an increasingly important role in CKD care quality improvement and oversight by CMS, it is essential to evaluate the psychometric properties of the measures. In this aim, I will select measures that operationalize broader constructs of PRMs, including treatment decision-making, health behaviors, and health-related quality of life, then examine the psychometric properties of these measures using a classical test theory approach.

Aim 2. **To examine measurement invariance for each measure between Black and White CKD patients.**

One of the benefits of PRMs is the inclusion the patient's voice. Because of disparities in access to and benefit from CKD treatments between different patient subgroups there is a need to

ensure that PRMs work equally well for each group. When PRMs are developed and evaluated only with single groups of patients (e.g., non-Hispanic, White males), it is unknown whether their psychometric properties are the same for different groups of patients (e.g., racial/ethnic groups, gender groups, different stages of disease). In this aim, I will identify whether selected treatment decision-making, medication adherence, and health-related quality of life measures have measurement equivalence for Black and White CKD patients subgroups for whom the largest disparities in access and outcomes of CKD care are present.[148]

Aim 3. **Based on the results of Aims 1 and 2, to make recommendations for improvements to each measure.**

The results of Aims 1 and 2 may reveal that the standard measures analyzed in this dissertation have some undesirable measurement properties, either in general or for specific groups of patients. Anticipated issues with these measures include a different dimensionality (factor structure) than what has previously been reported, lower reliability than is required for intended uses (especially for individual use, which requires a reliability of ≥ 0.90), items that do not load on hypothesized factors, or items that fail to demonstrate measurement invariance for Black and White patients. In the face of these issues, recommendations for ways to improve the measures will be made. The intended audience for these recommendations will primarily include clinicians, health care providers, or regulatory agencies that use the measures considered in this dissertation for patient assessment and CKD care oversight.

Figure 7 shows the general approach to be taken to incorporate each of the Aims with each of the constructs. For each construct or set of constructs, the analyses specified for Aims 1-3 will be applied to the PRMs used to measure these constructs. Under this approach, a similar

41

set of analyses will be applied to each measure used to operationalize each construct. Any

deviation from this standard approach will be noted and justified. This approach will create

consistency in the evaluation criteria for each measure, and lead to a clear rationale for

recommendations to improve each measure.

Figure 7. General Dissertation Approach
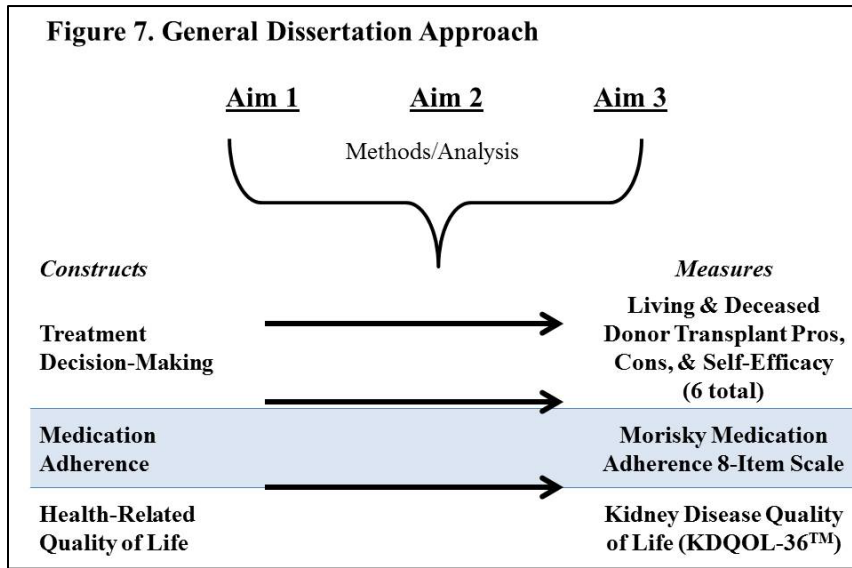
CHAPTER 4. METHODS

4.1. METHODS OVERVIEW

To meet these aims, I have collected data from three separate sources, each containing kidney

patient responses to patient-reported measures around the constructs of interest for this

dissertation: treatment decision-making, medication adherence, and health-related quality of life.

Unfortunately, it was not possible to locate a single dataset that contained the measures of

interest for each of these constructs. Therefore, one dataset each, three in total, will be used for

each of the three sets of constructs separately. Below, the measures to be used in this dissertation

will be described, along with each dataset. Finally, the statistical analyses used for each aim will

be described. In general, the same set of analyses will be applied to each dataset, except where

noted. Since each of these constructs can be operationalized and measured with a PRM that

captures a unique portion of patients' characteristics along their experience with and outcomes of

CKD and its treatments, the structure of this dissertation will be a continuous monograph instead

of a set of discrete papers.

4.2. MEASURES

4.2.1 TREATMENT DECISION-MAKING: TRANSPLANT SELF-EFFICACY & DECISIONAL BALANCE

For this dissertation, the LDKT and deceased donor kidney transplant (DDKT)

Decisional-Balance and Self-Efficacy measures described above in Chapter 2 will be evaluated.

The analyses conducted for this dissertation are only the second analyses of these measures to

date, so little additional data on the psychometric properties, reliability and validity of these

measures is available for reference.

The Decisional Balance measures, which rate the relative importance CKD patients place

on the benefits and disadvantages of LDKT and DDKT, and the LDKT and DDKT Self-Efficacy

43

scales, which measure the confidence an individual has in their ability to pursue a transplant in a wide variety of challenging situations, were initially evaluated in a sample of patients that had already presented to a transplant center to begin transplant evaluation.[73,74] This leaves open the possibility that constructs such as Decisional-Balance and Self-Efficacy, which indicate how patients are making decisions about their treatment, would be different for patients who have not yet presented to a transplant center for evaluation. Specifically, patients who have not yet presented to a transplant center for evaluation may be earlier in their decision-making process about which treatment they would like to pursue. Therefore, examining the psychometric properties of these measures among a dialysis-only sample will determine if they are appropriate for use with that subgroup.

In the randomized controlled trial (RCT) from which the data on the measures was taken, each measure was administered using a structured phone interview by research staff at UCLA.[119] These measures were given as part of a baseline assessment before patients were randomized to receive one of three possible transplant education interventions.

The LDKT Pros and Cons measures each have 6 items, prompted by "How important is this statement to your decision about living donor transplant?", then each are rated on a 5-point scale ranging from, "not important" (1) to "extremely important" (5). An example of a LDKT Pros item is "With a living donor transplant, I will be able to contribute to my family and friends sooner," and an example of a LDKT Cons item is "The surgery will inconvenience the living donor's work or life too much." To assess the LDKT self-efficacy items, patients were given the context: "How confident are you that you could get a living donor transplant?" on a 5-point response scale from "not at all confident" (1) to "completely confident" (5). An example of a LDKT self-efficacy item is: "A potential living donor who was evaluated did not match you". It

44

is important to note that this definition of "Self-Efficacy" varies somewhat from the more common, general definition used in behavioral health sciences, which focuses on individuals' beliefs about whether or not they can control their own behaviors and environmental factors impacting these behaviors.[52] None of the items from any of the LDKT measures were recoded or reverse coded, and summed scale scores are reported in the raw score metric.

The DDKT Pros and Cons measures also each have 6 items given with the context, "How important is this statement to your decision about transplant?", then rated on a 5-point scale ranging from, "Not important" (1) to "Extremely important" (5). An example of a DDKT Pros item is "I would not have to be on dialysis," and an example of a DDKT Cons item is "I would have to take a lot of medicine after transplant." Similar to the LDKT self-efficacy items, the 8 DDKT self-efficacy items are first given the context, "How confident are you that you could get a transplant even if:", and each item is then rated from "not at all confident" (1) to "completely confident" (5). An example of a DDKT Self-Efficacy is "You didn't have your own transportation to the transplant center." Like the LDKT measures, none of the items from the DDKT measures were recoded or reverse coded, and all summed scale scores are reported in raw metric. No analyses of these measures have been conducted other than the original validation analyses.[73,74]

4.2.2 MEDICATION ADHERENCE: MMAS-8

The 8-item version of the Morisky Medication Adherence Scale (MMAS-8) was selected as the measure to examine for the Medication Adherence construct. Though used widely in hypertension research,[100,149-152] as well as with other chronic conditions,[153,154] reports of the MMAS-8's psychometric properties among CKD populations are infrequent. Among a sample of 295 CKD stage 3-5 patients in Thailand, lower scores on the MMAS-8 were associated with

progression in stage of CKD, evidencing criterion validity.[77] Alkatheri and colleagues examined

adherence to an unspecified medication in 89 hemodialysis patients in Saudi Arabia using an

Arabic translation of the instrument.[155] This study found that 28% of patients reported "high"

adherence (MMAS-8 score of 8), 40% reported "medium" adherence (MMAS-8 score of <8 to

6), and 31% reported "low" adherence (MMAS-8 score of <6). However, no psychometric

properties of the measure from this study were reported.

The MMAS-8 is an expansion of the MMAS-4-item self-reported adherence scale.[135] The

MMAS-4 was found to predict completion of tuberculosis treatment among patients diagnosed

with active tuberculosis, but its internal consistency reliability was below the commonly used

0.70 threshold for acceptable reliability for group comparisons ($\alpha=0.61$).[135] Chemical markers

were correlated with the MMAS-4 and found to have a strong linear significant association

between the presence of the marker and self-reported medication-taking behavior ($p<0.01$).[156]

Three items of the MMAS-4 are included in the MMAS-8.[157] However, the MMAS-8 adds

additional reasons for non-adherence and increased reliability. In a study with hypertensive

patients, the MMAS-8 demonstrated acceptable internal consistency reliability ($\alpha=0.83$), with

corrected item-to-total correlations ranging between 0.40-0.59.[135] Additionally, a confirmatory

factor model provided support for a single dimension for the 8 items in the scale. The MMAS-8

was associated with related constructs in the hypothesized direction, including more medication

knowledge (positive association), having more social support (positive association), greater

patient satisfaction with clinic visits (positive association), better coping strategies (positive

association), higher reported stress (negative association), and medication complexity (negative

association).[139] Additionally, the MMAS-8 significantly predicted whether patients had their

blood pressure under control.

The MMAS-8 was administered to patients by Kantar health through telephonic interviews or internet-based self-administration along with other survey measures as determined by the patient's health condition and treatment or health services used. Of the MMAS-8's items, 7 have "yes/no" responses; e.g., "When you travel or leave home, do you sometimes forget to bring along your medications?" The 8th item, "How often do you have difficulty remembering to take all your blood pressure medication?", has a 5-point set of response options from "Never" (1) to "All the time" (5). For scoring, this item is dichotomized to "Never/Rarely" vs. "Once in a while/Sometimes/Usually/All the time". Use of this measure requires obtaining a license, and its scoring cannot be described here because it is proprietary. The final score ranges from 0-8, with higher scores indicating better adherence. A previous analysis by Morisky and colleagues indicated three categories of adherence derivable from the MMAS-8: High adherence = 8; Medium adherence = 6 - <8; Low adherence = <6.[139]

As demonstrated in the literature review, several key factors hypothesized to be associated with the MMAS-8 in this study's conceptual model (Figure 4.) have been examined among CKD patients. However, not all hypothesized relationships have been evidenced. For example, in the study of adherence to phosphate binders among dialysis patients, Joson and colleagues did not find a significant association between phosphate knowledge, an example of a Predisposing Characteristic in Andersen's model, and the MMAS-8.[158] The association between the MMAS-8 and patient satisfaction was evidenced in a study of 151 kidney transplant patients, wherein Alkatheri and colleagues found a significant, positive association between MMAS-8 scores and scores on the Treatment Satisfaction Scale.[159] These previous analyses can be used to guide analyses of the MMAS-8 among the current study's sample.

4.2.3 HEALTH-RELATED QUALITY OF LIFE: KDQOL-36

The HRQOL measure examined for this dissertation is the KDQOL-36. Investigators using data from the Chronic Renal Insufficiency Cohort (CRIC) Study examined the reliability of the KDQOL-36 among a sample of 836 English-speaking Hispanic (n=150), Spanish-Speaking Hispanics (n=270), and non-Hispanic Whites (n=409), and found comparable internal consistency reliability for the scales across these groups: Burden of Kidney Disease α ranging from 0.84-0.87; Symptoms/Problems of Kidney Disease α ranging from 0.82-0.83; and Effects of Kidney Disease α ranging from 0.81-0.83.[160] This study also examined construct validity of the KDQOL-36 subscales, but did not conduct factor analyses or use other latent-variable approaches, like item response theory analyses, to determine if the factor structure of the scales is supported.

There have been several recent reports of its psychometric properties in non-English language or non-American samples. Chen and colleagues examined a Chinese-language (Cantonese) version of the KDQOL-36 among 356 maintenance dialysis patients, both hemodialysis (n=253) and peritoneal (n=103).[161] Using confirmatory factor analysis, this study found support for the 3 targeted subscales, though model fit statistics all indicated less than excellent fit: root mean squared error approximation (RMSEA) = 0.08 (excellent fit is ≤0.06);[162,163] and comparative fit index (CFI) = 0.80 (excellent fit is ≥0.95).[163] Internal consistency and test-rest reliability were ≥0.80 for each scale.

Chao, et al. examined another Chinese-language (Mandarin) version of the KDQOL-36 in a sample of 428 CKD patients from stages 1-5.[164] Using a confirmatory factor model with one item removed yielded less than excellent, fit: RMSEA = 0.063; goodness-of-fit index (GFI) = 0.83. Yang and colleagues conducted a study of the KDQOL-36 among 394 hemodialysis patients in Singapore.[165] Item-to-scale correlations in this study ranged between 0.76-0.90, while

48

a confirmatory factor model showed somewhat less than excellent fit: RMSEA = 0.09; CFI = 0.93. Internal consistency reliability for the scales ranged between 0.80-0.91.

A South Indian (Kannada) version of the KDQOL-36 was examined in a sample of 82 hemodialysis patients.[166] Internal consistency reliability for the KDQOL-36 scales ranged from 0.72-0.77, and test-retest reliability examined in a subsample of these patients (n=45) ranged from 0.83-0.99. The scales were also significantly correlated with the European Quality of Life Visual-Analog Scale (EQ-VAS). In a Cantonese-speaking Chinese sample of 110 dialysis and 122 kidney transplant patients, and test-retest reliability ranged between 0.98-1.00, though internal consistency reliabilities ranged from 0.61-0.83.[167] Another Mandarin Chinese version of the KDQOL-36 was examined by Tao, et al. in a sample of 103 dialysis patients.[168] Internal consistency reliabilities ranged between 0.76-0.78 for the scales, and test-retest reliabilities ranged between 0.84-0.86. Support for construct validity was provided by significant correlations between the scales and a rating of overall health and the Beck Depression Index version 2 in the hypothesized directions.

The associations between the KDQOL-36 and other, similar constructs support construct validity based on hypotheses generated by the conceptual model given in Figure 4. For example, the CRIC study described above found significant associations with depression and reported CKD symptoms and the KDQOL-36.[160] However, other hypothesized relationships, like the one between medication adherence and the KDQOL-36, have been scarcely or never examined. In the literature review for this dissertation, no studies examining the association between medication adherence and the KDQOL-36 or KDQOL-SF (SF-36 and 11 kidney disease targeted scales) were found. A few studies examine the association between KDQOL scores and other health behaviors. Among these, Bayoumi and colleagues found that longer dialysis duration was

associated with a lower total KDQOL-SF score (non-standard scoring).[169] Finally, a lower average of the 11 KDQOL-SF subscale scores was associated with a higher risk of hospitalization and mortality.[170,171]

Of note, a non-standard scoring of the KDQOL-SF was employed in the international Dialysis Outcomes and Practice Patterns Study (DOPPS), which administered the KDQOL-SF among 7,378 dialysis patients from Europe (n=2,406), Japan (n=2,087), and the United States (n=2,885).[172] This non-standard score, labelled the Kidney Disease Component Score (KDCS), was generated by taking the average of the KDQOL-SF's 11 targeted scales (excludes SF-36 scales). Though predictive of mortality, as reported above, no psychometric analyses have been conducted to support the use of this score as a representative of a unidimensional scale. Nonetheless, the DOPPS study did report norms for each of the scales individually, including country-specific norms to which the scores of this study can be compared.

The KDQOL-36 was administered to dialysis patients during their dialysis treatment or over the phone by staff of the Medical Education Institute (MEI). The data were collected as part of MEI's effort to assist dialysis providers meet their CMS requirement to collect the KDQOL-36 and incorporate the responses into a patient's personalized care plan. The KDQOL-36 items are given in Appendix 1. The SF-12 is embedded with the KDQOL-36 and its Physical Health Component Score (PCS) and the Mental Health Component Score (MCS) are scored on a T-score (mean = 50, SD = 10, in U.S. general population). The three kidney-targeted scales (Burdens of Kidney Disease, Symptoms/Problems of Kidney Disease, and Effects of Kidney Disease) are each score by linearly transforming all items to a 0-100 possible range and averaging the items. KDQOL-36 items are all scored so that higher scores indicate better health. Finally, a preference-based scale, the SF-6D, can be derived from SF-12 items within the

KDQOL-36.[173] The SF-6D is estimated using 7 of the SF-12 items, covering 6 domains, including: Physical Functioning (1 item), Role Limitations (2 items), Social Functioning (1 item), Pain, Mental Health (1 item), and Vitality (1 item). The SF-6D is scored by weighing responses to each of these items then summing each respondent's weighted responses. The scores range between 0-1 with the theoretical minimum anchored at dead (0) and the maximum anchored at full health (1). Unless patients die during a study's follow-up, values of 0 are unlikely, and the bottom range of observed scores has been reported at 0.30.[174]

Table 1. Summary of Measures	
Construct	Measure(s)
Treatment Decision-Making	Waterman Living & Deceased Donor Kidney Transplant • Pros (1 LDKT, 1 DDKT) • Cons (1 LDKT, 1 DDKT) • Self-Efficacy (1 LDKT, 1 DDKT)
Medication Adherence	Morisky Medication Adherence Scale 8-Item Version (MMAS-8)
Health-Related Quality of Life	Kidney Disease Quality of Life (KDQOL™) 36 Item Version

4.3. DESCRIPTION OF DATASETS

4.3.1 IRB AND DATA PROTECTIONS

All datasets employed in this dissertation are de-identified, with all patient identifiers stripped from the datasets. Nonetheless, an application to the UCLA Institutional Review Board (IRB) was made describing the study's protocol. The study protocol poses minimal to no risk for kidney patients' whose data is included in the study. Without individual identifiers included in the data, there is no chance that the patients' identities can be disclosed. Further, data used in this study will only be shared in aggregated form. An IRB exemption was granted by the UCLA Human Subjects Protection Committee [UCLA IRB # 17-000399 (Treatment Decision-Making Dataset), #17-000403 (Medication Adherence Dataset), #17-000313 (Health-Related Quality of Life Dataset)].

4.3.2 TREATMENT DECISION-MAKING

The data for this dimension of the project are from a RCT that assesses the efficacy of kidney transplant education programs to increase ESRD patients' knowledge of transplant. The data was collected in a study lead by Dr. Amy Waterman in the Division of Nephrology, David Geffen School of Medicine at UCLA, who has granted permission to use a completely de-identified version of this dataset for the present study.

The RCT was conducted with ESRD patients who are currently receiving dialysis treatment in clinics throughout the state of Missouri. In this study, all participants are on dialysis and may not yet have presented to a transplant center to begin evaluation. Eligible patients include those who are: 18 years or older; identify as Black or White race; currently receive dialysis; have a household income of no more than 250% of the federal poverty level; and can read and speak English. At the start of the study, patients are randomized to receive either theoretically-grounded, tailored transplant education or standard-of-care transplant education received in dialysis centers. Participants' self-efficacy to pursue transplant, decisional balance (weighing of pros and cons to transplant), and knowledge of kidney transplant is assessed at the beginning (pretest) and end (posttest) of an 8 month period. The protocol for this trial has also been published. For the reasons described in Chapter 2, only the baseline assessments of Decisional Balance and Self-Efficacy will be employed for the present project. Five hundred sixty one patients completed the baseline assessment, and their data will be used for this project.

4.3.3 MEDICATION ADHERENCE

Data for medication adherence dimension of the study come from a private research organization, Kantar Health. Their *National Health and Wellness Survey* includes many patient reported health measures, including the 8-item version of the Morisky Medication Adherence

Scale (MMAS-8). For this study, those reporting kidney disease and kidney failure were selected. Additionally, the survey includes information on other health behaviors (e.g., use of tobacco and alcohol), multiple medical comorbidities diagnosed by a physician (e.g., diabetes, depression), information about medical treatments received (e.g., use of dialysis), and patient-reported outcomes (e.g., health related quality of life, satisfaction with care).

Kantar Health agreed to provide a completely de-identified dataset for adult CKD patients participating in their 2012-2013 *National Health and Wellness Survey*. The primary data of interest are responses to the MMAS-8, though a range of other variables were provided, including patient demographic information, presence of medical comorbidities, use of dialysis, health behaviors, and health related quality of life, measured with the MOS SF-36.

4.3.4 HEALTH-RELATED QUALITY OF LIFE

Data for HRQOL was obtained from the Medical Education Institute (MEI)[175]. MEI administered the KDQOL-36 as part of their KDQOL-Complete program. KDQOL-Complete is aimed at helping dialysis providers meet the CMS requirement to report the number of patients completing the KDQOL-36 instrument each year, and to use patients' responses to the KDQOL-36 in the development of a personalized Plan of Care.[175] The KDQOL-Complete program offers a computerized interface for entering KDQOL-36 data then helps providers interpret each patient's case-mix adjusted scores adjusting for patients' demographic and medical characteristics with graphic data displays and comparisons to population norms.

The KDQOL-Complete data reports can then be used for patient education or quality improvement purposes. Along with the KDQOL-36 survey, MEI collects dialysis patients' demographic and clinical characteristics. MEI provided a de-identified dataset containing all

patients' responses to the KDQOL-36 from 6/1/2015 to 5/31/2016. Additionally, patients'

race/ethnicity, age, whether etiology of ESRD was diabetes, dialysis type (in-center

hemodialysis, peritoneal dialysis, conventional home hemodialysis), dialysis access site

[arteriovenous (AV) fistula, AV graft, venous catheter, PD catheter), employment status, and

language (including English, Spanish, and multiple others) the survey was administered in were

provided.

4.4. STATISTICAL ANALYSIS
4.4.1 AIM 1. ANALYSES

First, all items and variables were described with category frequencies and proportions,

central tendency (e.g., mean and median) and dispersion (e.g., standard deviation). For multi-

item scales, standard scoring procedures were used, reverse-scoring or recoding as needed. After

multi-item scales were created, their distributions, along with the number of complete scale

responses, were reported. Missing data were handed using complete case analyses. Additionally,

scores will be reported for important patient subgroups. For example, for the KDQOL-36, scores

within groups of patients with different dialysis access sites will be reported.

To conduct some psychometric analyses, the datasets were split into two random halves.

One of these datasets was used for exploratory factor analysis (EFA) and the other was used for

confirmatory factor analysis (CFA). Among the first, exploratory dataset, polychoric or

tetrachoric correlations were computed between the items and examined. Cohen's conventions[176]

for magnitude of correlations will be used: >0.10 to <0.243 = small; >0.243 to <0.371 = medium;

≥0.371= large. Then, this correlation matrix was used to perform exploratory factor analysis

using maximum likelihood estimation with oblique Promax solutions with squared multiple

correlation used for prior communalities. To determine the dimensionality suggested by the

exploratory factor analysis, several criteria were used, including the scree "elbow" test, parallel analysis, and the Tucker-Lewis reliability coefficient (values >0.95 indicate good fit). Parallel analysis uses simulations to generate "expected" eigenvalues and plots these against the observed eigenvalues; the highest number of observed eigenvalues that are larger than the expected indicates the number of factors to retain. If more than one factor was retained using these criteria, then standardized loadings on rotated factor pattern were reported. Otherwise, the loadings from the unrotated factor were reported.

Next, the other half of the randomly divided dataset was used for confirmatory factor models,[177] unless the factor structure evidenced in the EFA was the same as the original factor structure for the scale. In that case, the full sample was used. The CFA models were run with the polychoric/tetrachoric correlation matrix. Where appropriate, multilevel confirmatory factor models were employed.[178] These models used robust maximum likelihood estimation.[179] When more than one factor was considered, covariances between the factors were computed, and item loadings were set to zero on factors on which the items were not hypothesized to load. Model fit was determined with the Satorra-Bentler chi-square,[162] the comparative fit index (CFI), Tucker-Lewis Index (TLI) and root mean square error of approximation (RMSEA). Good model fit is evidenced by a non-significant Satorra-Bentler chi-square, CFI and TLI values above 0.95, and RMSEA of 0.06 or less.[162] The corrected item-to-total correlations were examined using multitrait scaling analysis, correcting for item overlap in cases with multiple scales[180] using the user-created SAS macro %MULTI.[181] It is important to note that the item-to-total scale correlations generated through this analysis are product moment correlations, not polychoric correlations. Coefficient alpha was used to estimate internal consistency reliability, with the following standards: ≥ 0.70, ≥ 0.80, and ≥ 0.90 indicating acceptable, good, and excellent

reliability, respectively, for use in examining group differences; ≥ 0.90 indicates suitability for use with individuals.[182] Additionally, for measures that can be reported at the medical facility level, I also estimated facility level reliability using 1-way ANOVA models to partition between versus within facility variance.[183]

4.4.2 AIM 2. ANALYSES

Measurement invariance was evaluated for the treatment decision-making, medication adherence, and HRQOL measures. Since a major interest in CKD is whether, in comparison to White patients, Black patients have equal access to the best treatments, take the same health behaviors to maintain their health, and, in consequence, have the same health outcomes, measurement invariance will be tested evaluated for Black and White patients. First, measurement invariance for each measure was based on confirmatory factor analysis (CFA) models using SEM. After examining the final CFA model from Aim 1, tests of 3 sequential, nested, multi-group CFA models were conducted using the Mplus software with the robust weighted least squares with mean and variance adjustment (WLSMV) estimator for categorical items. First, configural invariance models wherein all model parameters are freely estimated across groups were fit. This step shows whether the factor structure is the same for each measure across each group. Second, if configural invariance was found, then metric invariance was examined by constraining the factor loadings to be equal across groups. Third, if metric invariance was assured, then scalar invariance was examined by constraining intercepts (equal) as well as factor loadings across the groups. This step in the procedure determines if patients from the different groups provide similar ratings on each item or not. In each of these series of nested models, the Satorra-Bentler χ^2 difference test for categorical models and change in CFI (<0.01)[184] was examined to determine if the models significantly differ across groups. For

example, for a test of measurement invariance between Blacks and Whites at the configural

level, a significant χ^2 difference test or a difference in CFI of ≥ 0.01 indicated lack of configural

invariance. In examining differences in factor loadings across groups, the standards set by Yoon

and Millsap were adopted: ≥ 0.1 = small; ≥ 0.2 = medium; ≥ 0.3 = large.[185]

This sequential approach has been conducted already for the LDKT Treatment Decision-

Making measures (LDKT Pros, Cons, and Self-Efficacy)[123] and full or partial strict invariance

for each measure was found across race, gender, and education groups. Therefore, the

measurement invariance analyses were new for all measures except the LDKT Treatment

Decision-Making measures, though measurement invariance will be tested among these

measures to determine if these results are consistent with the former results.

4.4.3 AIM 3. ANALYSES

After the analyses in Aims 1 and 2 were completed for PRMs corresponding to each construct or

set of constructs, possible alterations of items and scales featured in this dissertation were

considered. The types of results considered in this Aim included, but were not limited to: 1)

dimensional/factor structure that deviated from originally-published factor structures; 2) items

with loadings < 0.40 in EFA or CFA analyses; 3) measures with internal consistency reliability $<$

0.70; and 4) measures that were not invariant across group. Some of these criteria have no

numeric cut-off as indicated by an index or other score, but instead were judged relative to item

and measure performance in published literature.

For Aim 3, the Spearman-Brown Prophecy Formula was used to determine how many

more items need to be added to a scale to increase its reliability. The Spearman-Brown Prophecy

Formula is:

$$\rho^{new} = \frac{k\rho^{old}}{1 + (k-1)\rho^{old}}$$

where ρ^{new} is the new reliability and ρ^{old} is the reliability already estimated for the scale; in this dissertation, ρ^{old} will be internal consistency reliability estimated with Cronbach's α. k is defined as the new test length over the current test length.[109] For these analyses, the desired ρ^{new} was set, then the formula was used to solve for k.

CHAPTER 5. TREATMENT DECISION-MAKING RESULTS

5.1 TREATMENT DECISION-MAKING AIM 1 RESULTS

The 561 participants in this study were mostly Black (71%), with similar proportions of females (49%) and males (51%). The average age was 54 years, and the majority of participants had achieved a high school diploma or less education. As expected from the study's inclusion criteria, the largest proportion of participants reported that they would be able to live <1 month in their current situation if they lost their income (42%), a measure of financial security, and only 2% reported having private health insurance. A very large majority were on hemodialysis (92%). Many of the participants reported higher levels of health literacy, with 48% reporting that they were confident filling out hospital forms by themselves "all of the time", and 55% reporting that they had someone help them read hospital materials "none of the time. (Table 2).

Table 2. Dialysis Patient Characteristics from Treatment Decision-Making Dataset (n=561)	
Race (%, n)	
Black	71% (399)
White	29% (162)
Sex (%, n)	
Female	49% (274)
Male	51% (287)
Age (mean, range)	54, 23-75
Etiology of ESRD: PKD (%, n)	7% (38)
Education level (%, n)	
High school diploma or less	53% (298)
Some college	32% (181)
College graduate or higher	14% (81)
Health insurance (%, n)	
Medicare (National medical card)	87% (489)
Medicaid (State medical card)	8% (46)
Private insurance (HMO or PPO)	3% (14)
Other insurance	2% (12)
If family lost current income, how long could you live in your current situation? (%, n)	
<1 month	42% (236)
1-2 months	23% (131)
3-6 months	9% (49)
7-12 months	4% (24)
>12 months	20% (110)

Table 2. Continued	
Dialysis Type (%, n)	
Hemodialysis	93% (519)
Peritoneal dialysis	7% (41)
How often has someone help read hospital materials? (%, n)	
None of the time	55% (309)
A little of the time/Some of the time	33% (184)
Most of the time/All of the time	12% (68)
How confident are you in filling out hospital forms by yourself? (%, n)	
All of the time	48% (271)
Most of the time/ Some of the time	38% (212)
A little of the time/None of the time	13% (75)
Has needed social support (%, n)	76% (426)
Medical mistrust[b] (mean, sd)	2.8 (0.6)
CDC HRQOL-4	
General health[c] (mean, sd)	3.4 (1.0)
Note: HLOC = Health locus of control	
[a]Score ranges from 6-36, with higher scores reflecting higher HLOC of this type.	
[b]Score ranges from 1-4, with higher scores reflecting higher medical mistrust.	
[c]Score ranges from 1-5, with higher scores reflecting better health.	

Table 3 shows the distribution of the LDKT and DDKT Pros, Cons, and Self-Efficacy

scale scores. Very few cases (4 of 561) included in the study sample were missing items from

these measures. All of these scales exhibited a ceiling effect, with extensive proportions at the

ceiling for the LDKT Pros (24%), LDKT Self-Efficacy (25%), DDKT Pros (38%), and DDKT

Self-Efficacy (19%) scales. Polychoric correlations among the items in each of these 6 scales are

shown in Tables 4a-4d and 5a-5b. For all the scales, the majority of items were highly or

moderately (medium) correlated with one another, with very few small correlations observed.

Table 3. LDKT and DDKT Pros, Cons, and Self-Efficacy (SE) Scales from Treatment Decision-Making Dataset						
	LDKT Pros	LDKT Cons	LDKT SE	DDKT Pros	DDKT Cons	DDKT SE
n of observations	558	557	558	559	558	561
Mean	25.0	20.1	22.5	26.4	21.3	32.6
Standard Deviation	4.9	5.9	6.9	4.7	5.9	6.9
% at Floor	0.2%	1%	3%	0.2%	0.4%	1%
% at Ceiling	24%	7%	25%	38%	8%	19%
Minimum Score Observed	6.0	6.0	6.0	6.0	6.0	8.0
25th Percentile	22.0	16.0	18.0	25.0	18.0	29.0
50th Percentile (Median)	26.0	20.0	24.0	28.0	22.0	34.0
75th Percentile	29.0	25.0	29.0	30.0	26.0	38.0
Maximum Score Observed	30.0	30.0	30.0	30.0	30.0	40.0
LDKT = Living donor kidney transplant; DDKT = Deceased donor kidney transplant						

Table 4a. Polychoric Correlations between Living Donor Kidney Transplant (LDKT) Pros Items from Treatment Decision-Making Dataset

	1	2	3	4	5	6
With a living donor transplant, I will be able to contribute to my family and friends sooner (1)	1.0					
I will be healthier because I spent less time on dialysis (2)	**.43**	1.0				
With a living donor transplant, I can return to my normal activities sooner (3)	**.74**	**.48**	1.0			
A living donor kidney generally lasts longer than a deceased donor kidney (4)	**.48**	.29	**.41**	1.0		
A living donor transplant could happen more quickly because I don't have to wait for a kidney on the waiting list (5)	.34	**.45**	**.45**	**.55**	1.0	
My living donor will feel good seeing my health improve (6)	**.65**	.35	**.50**	.31	.30	1.0

Table 4b. Polychoric Correlations between LDKT Cons Items from Treatment Decision-Making Dataset

	1	2	3	4	5	6
The surgery will inconvenience the living donor's work or life too much (1)	1.0					
I will feel guilty having someone donate to me (2)	**.40**	1.0				
I don't want to involve anyone else in my health problems (3)	**.38**	**.55**	1.0			
Donation could harm my relationship with a living donor (4)	**.48**	**.39**	**.41**	1.0		
The living donor could not donate again if someone closer to them ever need a kidney (5)	**.44**	.24	.15	.31	1.0	
A living donor could have health problems due to donating (6)	**.50**	.36	**.38**	**.55**	.28	1.0

Table 4c. Polychoric Correlations between LDKT Self-Efficacy Items from Treatment Decision-Making Dataset

	1	2	3	4	5	6
You asked someone to donate and they turned you down (1)	1.0					
A potential living donor changed their mind and decided not to be evaluated (2)	**.84**	1.0				
A potential living donor who was evaluated did not match you (3)	**.78**	**.82**	1.0			
You don't know anyone who might be a living donor for you (4)	**.74**	**.70**	**.60**	1.0		
You didn't know how to discuss living donation with potential donors (5)	**.77**	**.71**	**.59**	**.79**	1.0	
Other people were not supportive of you having a living donor transplant (6)	**.72**	**.63**	**.64**	**.59**	**.66**	1.0

Bold correlations are large according to Cohen's rule, ≥ .371; n's ranged from 521 – 557.

Table 5a. Polychoric Correlations between Deceased Donor Kidney Transplant (DDKT) Pros Items from Treatment Decision-Making Dataset

	1	2	3	4	5	6
I would not have to be on dialysis. (1)	1.0					
I would live a longer life with a transplant. (2)	**.43**	1.0				
I would feel better and have more energy with a transplant. (3)	**.43**	**.68**	1.0			
If I got a transplant, my family's life could return to normal. (4)	.36	**.55**	**.55**	1.0		
If I got a transplant, my friends and family would have me in their lives longer.(5)	**.40**	**.69**	**.68**	**.66**	1.0	
I could do more of the things I like to do with a transplant. (6)	**.48**	**.57**	**.67**	**.56**	**.70**	1.0

Table 5b. Polychoric Correlations between DDKT Cons Items from Treatment Decision-Making Dataset

	1	2	3	4	5	6
Transplant surgery would be very painful for me. (1)	1.0					
If the transplant fails, it would have been a lot of work and pain for nothing. (2)	.36	1.0				
I would have to take a lot of medicine after transplant. (3)	.30	.34	1.0			
I could die during the transplant surgery. (4)	**.48**	.34	**.44**	1.0		
I could have health problems due to the transplant. (5)	**.48**	.37	**.40**	**.51**	1.0	
I might not be able to pay for the drugs to prevent transplant rejection. (6)	.34	.28	.35	.33	**.42**	1.0

Table 5c. Polychoric Correlations between DDKT Self-Efficacy Items from Treatment Decision-Making Dataset								
	1	2	3	4	5	6	7	8
Your friends and family were unsupportive of you getting a transplant. (1)	1.0							
You didn't have transportation to the transplant center. (2)	**.50**	1.0						
The evaluation process took several months to finish. (3)	**.71**	**.63**	1.0					
The transplant tests and surgery were very painful. (4)	**.54**	**.54**	**.62**	1.0				
You had to lose weight or change your lifestyle in some way to be eligible for a transplant. (5)	**.56**	**.54**	**.66**	**.56**	1.0			
You didn't have someone to help you with your family responsibilities when you were recovering from surgery. (6)	**.51**	**.53**	**.50**	**.50**	**.59**	1.0		
The transplant evaluation and surgery scared you. (7)	**.47**	**.54**	**.63**	**.60**	**.52**	**.63**	1.0	
You didn't know how to pay for the drugs to prevent rejection after the surgery. (8)	**.42**	**.53**	**.44**	**.48**	**.52**	**.50**	**.52**	1.0
Bold correlations are large according to Cohen's rule, $\geq$ .371; n's ranged from 548-560.								

The 12 items for the LDKT Pros and Cons scales were examined in an exploratory factor analysis. The ML model for these scales resulted in a communality >1, called an ultra-Heywood case. Heywood cases in ML estimation may be more likely with small sample sizes, and principal axis factoring may be a better estimator than ML estimation in this cases.[186] Therefore, in the presence of an ultra-Heywood case with ML estimation, principal axis factoring was selected. In this model, the 12 eigenvalues were 3.47, 1.81, 0.47, 0.28, 0.16, 0.08, 0.04, -0.07, -0.12, -0.20, -0.22, -0.26 (recall that the priors are set to squared multiple correlations, causing negative eigenvalues). Figure 8 shows the plot of actual to simulated eigenvalues from the parallel analysis on this scale, which also indicated support for 2 factors. Additionally, the scree plot from this model (shows as the "Actual" eigenvalue plot in Figure 8), also suggested a 2 factor solution.

Figure 8. Parallel Analysis on LDKT Pros and Cons Scales

Parallel Analysis -Median Simulated Eigenvalues
12 Variables, &niterIterations, 281 Observations
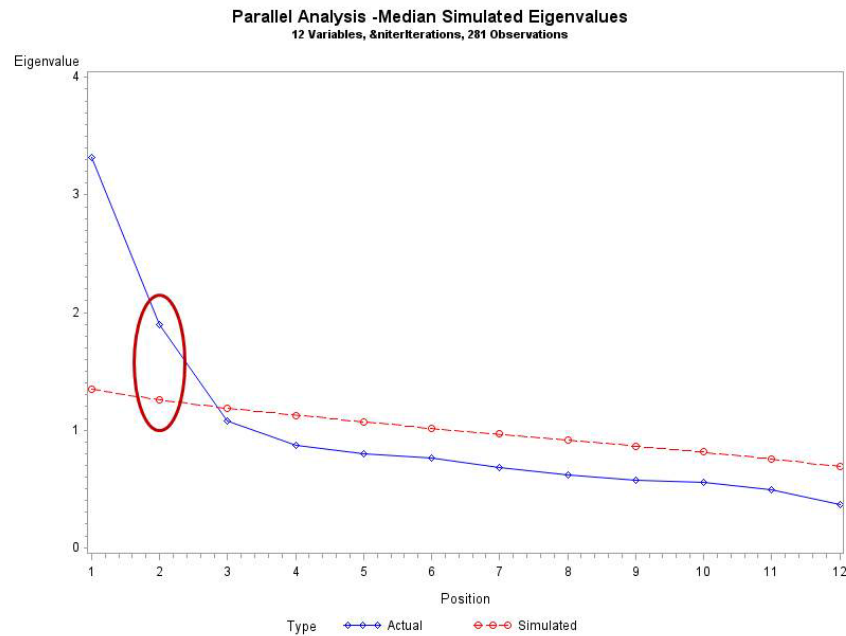
Table 5. shows the loadings for these scales. All loaded on their respective factors at 0.40 or greater. Given that the dimensionality of the LDKT Pros and Cons scale matched that of the original dimensionality,[74] the confirmatory factor analysis was conducted on the entire sample (n=561). A CFA model (Table 5) for these scales was fit wherein the LDKT Pros and Cons items (12 items total) were considered in a 2 correlated factors model for categorical variables (polychoric correlation matrix) with robust ML estimation. This approach was taken in the original construction of these scales. The two factor model showed good fit: Satorra-Bentler $\chi 2$ = 119.53, df = 53, p<0.001; RMSEA = 0.05; CFI = 0.96; and TLI = 0.95. The inter-correlation between the factors was 0.18. All factor loadings were greater than 0.40.

63

Table 6. Factor Analyses and Item-to-Total Correlations for Living Donor Kidney Transplant (LDKT) Pros and Cons Scales from Treatment Decision-Making Dataset	Factor 1 (LDKT Pros)			Factor 2 (LDKT Cons)		
	EFA Loading	CFA Loading	Item-Total-Corr.	EFA Loading	CFA Loading	Item-Total-Corr
LDKT Pros (α = 0.72)						
With a living donor transplant, I will be able to contribute to my family and friends sooner (LDPro1)	0.85	0.83	0.53	-0.06	-	0.07
I will be healthier because I spent less time on dialysis (LDPro2)	0.55	0.55	0.35	0.07	-	0.17
With a living donor transplant, I can return to my normal activities sooner (LDPro3)	0.80	0.84	0.57	-0.04	-	0.11
A living donor kidney generally lasts longer than a deceased donor kidney (LDPro4)	0.55	0.47	0.35	0.21	-	0.21
A living donor transplant could happen more quickly because I don't have to wait for a kidney on the waiting list (LDPro5)	0.59	0.57	0.43	0.08	-	0.12
My living donor will feel good seeing my health improve (LDPro6)	0.67	0.69	0.44	-0.11	-	0.05
LDKT Cons (α = 0.71)						
The surgery will inconvenience the living donor's work or life too much (LDCon1)	0.01	-	0.14	0.69	0.76	0.56
I will feel guilty having someone donate to me (LDCon2)	-0.02	-	0.08	0.62	0.63	0.48
I don't want to involve anyone else in my health problems (LDCon3)	-0.05	-	0.10	0.63	0.58	0.44
Donation could harm my relationship with a living donor (LDCon4)	0.01	-	0.16	0.68	0.62	0.43
The living donor could not donate again if someone closer to them ever need a kidney (LDCon5)	0.20	-	0.17	0.41	0.43	0.29
A living donor could have health problems due to donating (LDCon6)	0.003	-	0.12	0.67	0.68	0.48

Table 6 also shows the item-to-total scale correlations (corrected for overlap) for each

item and the internal consistency reliability for each scale (both calculated on the entire sample).

Most items had large correlations with their hypothesized scale (≥ 0.371), or very near large

correlations (e.g., *LDPro2* and *LDPro4* were both correlated with the LDKT Pros Scale at

r=0.35). The lowest item-to-scale correlation was between *LDCon5* and the LDKT Cons scale, at

r=0.29, still medium in magnitude. Internal consistency reliabilities were only acceptable for

each of these scales (≥ 0.70 and <0.80).

Regarding the LDKT Self-Efficacy scale, the EFA using ML estimation yielded the following eigenvalues: 19.13, 0.70, 0.12, -0.02, -0.32, -0.48. Parallel analysis and the scree plot for the LDKT Self-Efficacy scale indicated a 1 factor solution. (Figure 9)

Figure 9. Parallel Analysis on LDKT Self-Efficacy Scale
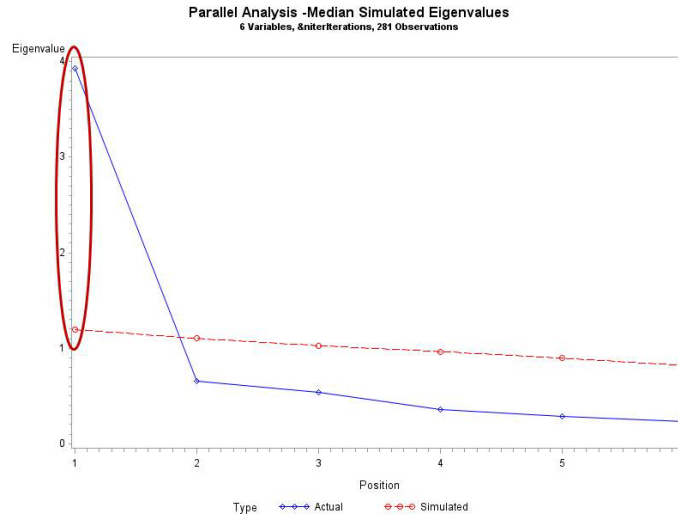
Again, since the EFA dimensionality matched that of the original scale, the CFA was run on the full dataset. The CFA model for categorical variables (polychoric correlations) with robust ML estimation for the LDKT Self-Efficacy scale for a single factor showed mixed results. The Satorra-Bentler was significant: χ^2 = 62.09, df = 9, p<0.001. Additionally, the RMSEA did not indicate good model fit at 0.10. However, other fit indexes indicated excellent fit: CFI = 0.99; and TLI = 0.99. Table 7 shows the factor loadings from the EFA and CFA models. All loadings for both the EFA and CFA models were >0.40.

Table 7. Factor Analyses and Item-to-Total Correlations for Living Donor Kidney Transplant (LDKT) Self-Efficacy from Treatment Decision-Making Dataset			
LDKT Self-Efficacy (α = 0.89)	EFA Loading	CFA Loading	Item-Total-Corr
You asked someone to donate and they turned you down (LDSE1)	0.94	0.76	0.79
A potential living donor decided not to be evaluated (LDSE2)	0.90	0.80	0.76
A potential living donor who was evaluated did not match you (LDSE3)	0.83	0.89	0.70
You don't know anyone who might be a living donor for you (LDSE4)	0.79	0.86	0.64
You didn't know how to discuss living donation with potential donors (LDSE5)	0.82	0.82	0.69
Other people were not supportive of you having a LDKT (LDSE6)	0.75	0.79	0.66
EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis			

In addition, Table 7 shows the item-to-scale (corrected for overlap) and internal consistency reliability of the LDKT Self-Efficacy scale. All the item-to-scale correlation coefficients were large in magnitude (>0.371), and the internal consistency reliability was good

Figure 10. Parallel Analysis on DDKT Pros and Cons Scales
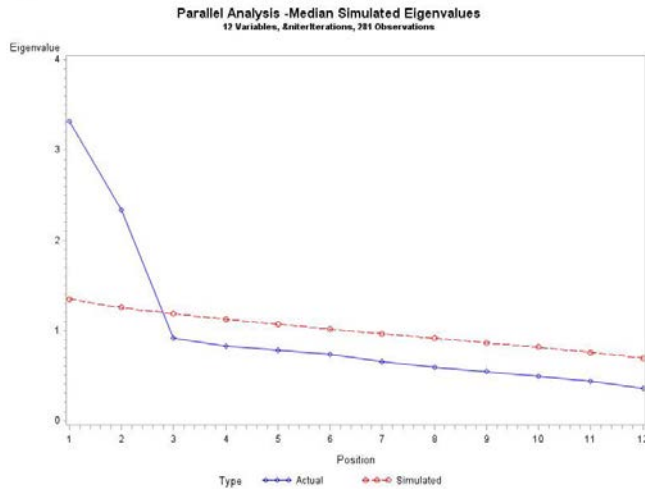
(≥0.80), bordering on excellent (≥0.90) at r=0.89.

Regarding the EFA for the DDKT Pros and Cons items, the ML solution also yielded an ultra-Heywood case (communality >1), so principal axis factoring was used for estimation. This model, containing all 12 DDKT Pros and Cons items, yielded the following eigenvalues: 3.80, 2.10, 0.29, 0.22, 0.15, 0.09, 0.02, -0.04, -0.09, -0.14, -0.20, and -0.26. Figure 10 shows that parallel analysis and the scree plot supported 2 factors.

Table 8 shows the factor loadings from this EFA model. All items loaded on their hypothesized scale at 0.40 or higher. Since this dimensionality matched the original scale, the CFA was run on the full dataset. The 2 correlated factors CFA model for categorical variables (polychoric correlation matrix) with robust ML estimation was an excellent fit to the data: Satorra-Bentler χ^2 = 69.92, df = 53, p=0. 06; RMSEA = 0.03; CFI = 0.99; and TLI = 0.99. The inter-correlation between the factors was 0.23.

Table 8. Factor Analyses and Item-to-Total Correlations for Deceased Donor Kidney Transplant (DDKT) Pros and Cons Scales from Treatment Decision-Making Dataset						
	Factor 1 (DDKT Pros)			Factor 2 (DDKT Cons)		
	EFA Loading	CFA Loading	Item-Total-Corr.	EFA Loading	CFA Loading	Item-Total-Corr
DDKT Pros (α = 0.80)						
I would not have to be on dialysis. (DDPro1)	0.54	0.61	0.40	-0.01	-	0.10
I would live a longer life with a transplant. (DDPro2)	0.82	0.80	0.61	-0.13	-	0.06
I would feel better and have more energy with a transplant. (DDPro3)	0.83	0.81	0.59	-0.05	-	0.06
If I got a transplant, my family's life could return to normal. (DDPro4)	0.71	0.72	0.55	0.05	-	0.17
If I got a transplant, my friends and family would have me in their lives longer. (DDPro5)	0.84	0.81	0.58	0.10	-	0.14
I could do more of the things I like to do with a transplant. (DDPro6)	0.78	0.83	0.64	0.07	-	0.12
DDKT Cons (α = 0.72)						
Transplant surgery would be very painful for me. (DDCon1)	-0.07	-	0.06	0.65	0.64	0.50
If the transplant fails, it would have been a lot of work and pain for nothing. (DDCon2)	-0.06	-	0.06	0.56	0.57	0.43
I would have to take a lot of medicine after transplant. (DDCon3)	0.17	-	0.15	0.54	0.58	0.40
I could die during the transplant surgery. (DDCon4)	-0.01	-	0.12	0.71	0.66	0.47
I could have health problems due to the transplant. (DDCon5)	-0.05	-	0.10	0.72	0.74	0.56
I might not be able to pay for the drugs to prevent transplant rejection. (DDCon6)	0.12	-	0.12	0.52	0.58	0.39
EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis						

Table 8 also shows the corrected item-to-total correlations for the DDKT Pros and Cons scales. All the item-to-scale correlations were of large magnitude (≥0.371). Additionally, while the DDKT Pros scale exhibited good reliability (≥0.80), the DDKT Cons scale only exhibited acceptable reliability (≥0.70). The reliability of the DDKT Cons scale was comparable to that of its validation study (0.80 vs. 0.81, respectively), though the DDKT Cons scale had worse reliability in the present study (0.72 vs. 0.76).[73]

Finally, regarding the DDKT Self-Efficacy scale, like many of the other scales, an ultra-Heywood case emerged for the ML model (communality > 1). Therefore, the principal axis factor model was used. This model yielded the following eigenvalues: 4.39, 0.25, 0.07, 0.04, -0.03, -0.04, -0.09, and -0.21. Parallel analysis and the scree plot of this scale indicated a 1 factor solution. (Figure 11).

Figure 11. Parallel Analysis on DDKT Self-Efficacy Scale

The loadings from the model are shown in Table 9. All loadings were greater than 0.40, ranging from 0.65 (DDSE8) to 0.83 (DDSE3). Like the models above, since the dimensionality suggested by the EFA matched the original dimensionality, the CFA was run on the full sample (n=561). The results of a CFA model for categorical variables (polychoric correlations) with a single factor using robust ML estimation demonstrated good fit to the data: Satorra-Bentler $\chi^2 =$ 53.00, df = 20, p<0.001; RMSEA = 0.06; CFI = 0.99; and TLI = 0.99. All of the loadings were >0.40.

Table 9. Factor Analyses and Item-to-Total Correlations for Deceased Donor Kidney Transplant (DDKT) Self-Efficacy from Treatment Decision-Making Dataset			
DDKT Self-Efficacy (α = 0.83)	EFA Loading	CFA Loading	Item-Total-Corr
Your friends & family were unsupportive of you getting a transplant. (DDSE1)	0.73	0.71	0.55
You didn't have transportation to the transplant center. (DDSE2)	0.73	0.66	0.51
The evaluation process took several months to finish. (DDSE3)	0.83	0.83	0.64
The transplant tests and surgery were very painful. (DDSE4)	0.74	0.75	0.58
You had to lose weight or change your lifestyle in some way to be eligible for a transplant. (DDSE5)	0.76	0.76	0.60
You didn't have someone to help you with your family responsibilities when you were recovering from surgery. (DDSE6)	0.73	0.65	0.58
The transplant evaluation and surgery scared you. (DDSE7)	0.76	0.67	0.59
You didn't know how to pay for drugs to prevent rejection after surgery. (DDSE8)	0.65	0.59	0.47
EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis			

Table 9 also shows the item-to-total scale correlations (corrected for overlap) and internal consistency reliability for the DDKT SE scale. All item-to-total scale correlations were large in magnitude. Additionally, the internal consistency reliability was good at 0.83.

5.2 TREATMENT DECISION-MAKING AIM 2 RESULTS

Measurement invariance between Black and White patients was examined for the 2 correlated factors DDKT Decisional Balance (Pros and Cons) and a 1 factor model for DDKT Self-Efficacy using multiple-group CFA using the robust weighted least squares with mean and variance adjustment (WLSMV) estimator. Table 10 shows the loadings for White vs. Black patients for DDKT Pros and Cons scales from the Configural model. Loadings on the Pros and Cons scales were largely similar across the racial groups and all fell within the range of small differences ($0.1 \geq$ loading <0.20).

Table 10. Configural Robust Weighted Least Squares Model Factor Loadings for White and Black Patients on Deceased Donor Kidney Transplant (DDKT) Pros and Cons Scales				
	DDKT Pros		DDKT Cons	
	Black	White	Black	White
DDPro1	0.54	0.70	-	-
DDPro2	0.74	0.86	-	-
DDPro3	0.78	0.88	-	-
DDPro4	0.74	0.72	-	-
DDPro5	0.76	0.90	-	-
DDPro6	0.79	0.93	-	-
DDCon1	-	-	0.58	0.66
DDCon2	-	-	0.53	0.58
DDCon3	-	-	0.56	0.70
DDCon4	-	-	0.66	0.63
DDCon5	-	-	0.75	0.74
DDCon6	-	-	0.61	0.56

Even though some small differences in loadings were observed, the formal tests of measurement invariance did not offer evidence to reject invariance hypotheses. Table 11 shows these results. The χ^2 difference tests between the Configural, Metric, and Scale models between the increasingly restrictive models were all non-significant, and the ΔCFI values fell below the critical value for rejecting the null hypothesis invariance at every step (0.01).

Table 11. Measurement Invariance White vs. Black for DDKT Decisional Balance									
Model[b]	Hypothesis	χ^2	Df	P	CFI	$\Delta\chi^{2,a}$	Δdf	p	ΔCFI
Configural	Invariant factor structure	153.21	106	0.002	0.984	-	-	-	-
Metric	Above + invariant factor loadings	161.90	116	0.003	0.985	10.02	10	0.44	0.001
Scalar	Above + invariant intercepts	200.13	150	0.004	0.983	40.84	34	0.20	-0.002
[a]$\Delta\chi^2$ difference test for model in row above: Metric vs. Configural, Scalar vs. Metric									
[b]All models estimated with Robust Weighted Least Squares.									

Regarding the measurement invariance results for the DDKT SE scale, all differences in loadings in the Configural Model between Black and White patients were below the threshold for small differences ($\geq$0.1 = small; $\geq$0.2 = medium; $\geq$0.3 = large.). (Table 12.)

Table 12. Configural Robust Weighted Least Squares Model Factor Loadings for White and Black Patients on the DDKT Self-Efficacy Scale		
	DDKT SE	
	Black	White
DDSE1	0.70	0.72
DDSE2	0.68	0.62
DDSE3	0.86	0.77
DDSE4	0.76	0.75
DDSE5	0.76	0.76
DDSE6	0.74	0.64
DDSE7	0.74	0.72
DDSE8	0.61	0.56

Additionally, although the χ^2 differences test between the Metric and Scalar models was statistically significant at 0.04, the ΔCFI between these models was very low at 0.001, well below the 0.01 threshold, with model fit actually improving somewhat in the Scalar model. (Table 13.)

Table 13. Measurement Invariance White vs. Black for DDKT Self-Efficacy									
Model[b]	Hypothesis	χ^2	df	p	CFI	$\Delta\chi^{2,a}$	Δdf	p	ΔCFI
Configural	Invariant factor structure	102.45	40	<0.001	0.979	-	-	-	-
Metric	Above + invariant factor loadings	96.83	47	<0.001	0.983	3.95	7	0.97	0.004
Scalar	Above + invariant intercepts	118.19	70	<0.001	0.984	36.78	23	0.04	0.001
[a]$\Delta\chi^2$ difference test for model in row above: Metric vs. Configural, Scalar vs. Metric									
[b]All models estimated with Robust Weighted Least Squares.									

5.3 TREATMENT DECISION-MAKING AIM 3 RESULTS

The biggest deficiency of these scales was their reliability. In particular, the LDKT Pros and Cons, and the DDKT Cons scales had Cronbach's alphas just over the cut-off for acceptable reliability, 0.70. Using the Spearman-Brown Prophecy Formula, the number of items required to bring all the scales to good ($\alpha = 0.80$, if not already achieved) and excellent reliability was

calculated ($\alpha = 0.90$). Table 14. shows the number of total items for each scale needed to achieve good and excellent reliability.

Table 14. Spearman Brown Prophecy Formula Calculations for Increased Reliability of Living Donor Kidney Transplant (LDKT) and Deceased Donor Kidney Transplant (DDKT) Pros, Cons, and Self-Efficacy Scales			
Scale	Current α, n of Items	Total Items Needed for Good Reliability ($\alpha = 0.80$)	Total Items Needed for Excellent Reliability ($\alpha = 0.90$)
LDKT Pros	0.72, 6	10	21
LDKT Cons	0.71, 6	10	23
LDKT Self-Efficacy	0.89, 6	-[a]	7
DDKT Pros	0.80, 6	-[a]	14
DDKT Cons	0.72, 6	10	21
DDKT Self-Efficacy	0.83, 8	-[a]	15
[a]Already achieved good reliability.			

For the LDKT Pros, LDKT Cons, and DDKT Cons scales, a total of 10 items would be required for these scales to achieve good reliability, entailing the addition of 4 items to each scale. Each scale would need to add several items to achieve excellent reliability, with items total for each scale ranging between 7-23 items.

Based on these analyses, the primary recommendation for future research on these scales includes their expansion with more items to achieve excellent reliability. Each of these scales is used for individual assessment in the context of educational interventions about kidney transplantation.[120] The current, low reliabilities of these scales are insufficient for this usage. However, additional scale length may also create administrator and respondent burden. Therefore, the most useful study of these scales would be to create item banks with many items, then use Item Response Theory to create short forms that achieve the reliability of the full item bank.

CHAPTER 6. MEDICATION ADHERENCE RESULTS

6.1 MEDICATION ADHERENCE AIM 1 RESULTS

There were 130,088 total participants in the Kantar Health 2012-2013 *National Health and Wellness Survey*. Of these, the analysis sample consisted of 756 who reported CKD. The average age of these participants was 58 years (range: 20-93), and they were predominantly White (71%). The largest proportion of these participants had completed some college (46%) and most were married or cohabiting (58%). A large majority was not on dialysis (78%), and the most common medical comorbidities were hypertension (77%) and diabetes (45%), the two leading causes of kidney failure.[148] (Table 15).

Table 15. CKD Patient Characteristics from Medication Adherence Dataset (n= 756)	
Age (mean, SD, range)	57.6, 14.7, 20-93
Race/Ethnicity (%, n)	
White	71% (538)
Black	13% (97)
Asian	4% (27)
Hispanic	9.0% (68)
American Indian	2% (13)
Multi-racial	1% (9)
Other	0.5% (4)
Sex	
Female	38% (290)
Male	62% (466)
Education (%, n)	
High school grad or less	20% (152)
Some college	46% (345)
College grad or more	34% (259)
Marital Status (%, n)	
Married/Cohabitation	58% (440)
Single, never married	16% (122)
Divorced/separated	16% (123)
Widowed	10% (71)
On Dialysis (%, n)	
Yes	21% (158)
No	78% (588)
Missing	1% (10)
BMI (mean, SD, range)	31.2, 8.2, 11.7-82.6
Ever smoked cigarettes	
Yes	58% (435)
No	42% (321)

Table 15. Continued	
Comorbidities Diagnosed by Dr (%, n)	
Alcoholism	6% (43)
Anxiety	24% (180)
Hypertension	77% (581)
Congestive heart failure	17% (129)
Diabetes	45% (343)

Most respondents self-reported being adherent to their medication on the 8 items of the MMAS-8. (Table 16). For example, to the item "Do you sometimes forget to take your pills?", only 20% answered "yes". Additionally, to these items "Have you ever cut back or stopped taking your medication without telling your doctor because you felt worse when you took it? " and "When you feel like your condition is under control, do you sometimes stop taking your medicine?," only 7% and 8% responded "yes", respectively. The mean MMAS-8 score (0-8 possible range) in this sample was 6.7 (SD = 1.6).

Table 16. Distribution of MMAS-8 Items among CKD Patients from Medication Adherence Health Dataset (n= 756)	
Do you sometimes forget to take your pills?	
Yes	20% (149)
No	80% (607)
Over the past 2 weeks, were there any days when you did not take your medicine?	
Yes	13% (95)
No	87% (661)
Have you ever cut back or stopped taking your medication without telling your doctor because you felt worse when you took it?	
Yes	7% (52)
No	93% (704)
When you travel or leave home, do you sometimes forget to bring along your medications?	
Yes	11% (85)
No	89% (671)
Did you take your medicine yesterday?	
Yes	88% (663)
No	12% (93)
When you feel like your CKD is under control, do you sometimes stop taking your medicine?	
Yes	8% (57)
No	92% (699)
Taking medication every day is a real inconvenience for some people. Do you ever feel hassled about sticking to your treatment plan?	
Yes	25% (193)
No	75% (563)
How often do you have difficulty remembering to take all your medication?	
Never/Rarely	66% (498)
Once in a while/Sometimes/Usually/All the time	34% (258)

Examining polychoric correlations between the items, most items were highly inter-correlated (polychoric $r > 0.371$). A notable exception was the item "Did you take your medicine yesterday?", which had relatively low correlations with the other items. (Table 17.)

Table 17. Polychoric Correlations between MMAS-8 Items from Medication Adherence Dataset (full dataset)								
	1	2	3	4	5	6	7	8
Forgetting (Item 1)	1.0							
Any days (Item 2)	**.71**	1.0						
Cut-back (Item 3)	**.59**	**.81**	1.0					
Travel (Item 4)	**.58**	**.67**	**.65**	1.0				
Yesterday (Item 5)	-.01	.10	.19	.05	1.0			
Under control (Item 6)	**.69**	**.68**	**.82**	**.67**	.23	1.0		
Inconvenience (Item 7)	**.45**	**.59**	**.61**	**.56**	-.06	**.47**	1.0	
Remembering (Item 8)	**.59**	**.63**	**.52**	**.54**	.21	**.41**	**.51**	1.0
Bold correlations are large according to Cohen's rule, $\geq .371$								

Figure 12. Parallel Analysis of MMAS-8

Parallel Analysis - Median Simulated Eigenvalues
8 Variables, &niterIterations, 378 Observations
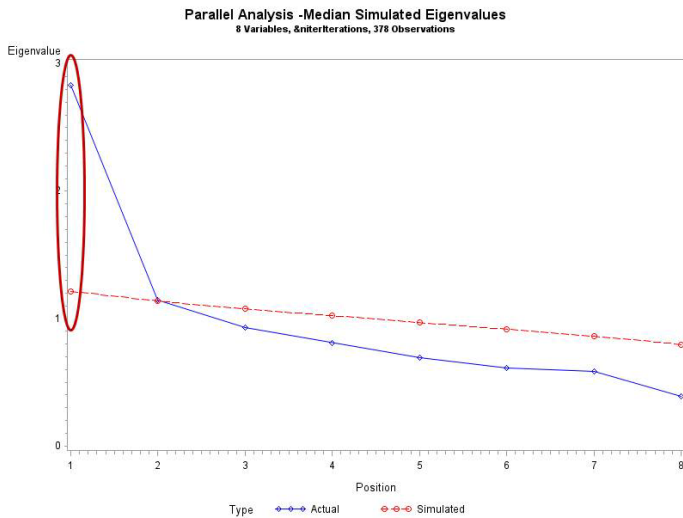
In the EFA, ML estimation yielded an ultra-Heywood case, so principal factoring was chosen. This model resulted in the following 8 eigenvalues: 4.17, 0.76, 0.30, 0.15, 0.05, -0.05, -0.11, and -0.18. The parallel analysis of this scale suggested a 1 factor solution. (Figure 12) The loadings for this EFA are presented in Table 11. All items loaded at >0.40 except the 5th item, "Yesterday". Since the factor structure of the EFA largely mirrored the original MMAS-8 factor structure, the CFA was conducted on the full sample (*n*=756). This CFA model, which employed polychoric correlations for the categorical variables and used robust ML estimation, fit the data very well: Satorra-Bentler $\chi^2 = 57.28$, $df = 20$, p<0.001; RMSEA = 0.05; CFI = 0.98; TLI = 0.97. However, Item 5 ("Yesterday") had a very low loading at 0.15. Therefore, a second

model omitting this item was examined: Satorra-Bentler χ^2 = 63.12, df = 13, p<0.001; RMSEA = 0.06; CFI = 0.99; TLI = 0.98. The likelihood ratio test to determine whether nested models differed significantly in fit (original model vs. original model with loading for Item 5 set to 0) yielded the following: $\chi^2_{diff} = \chi^2_{full} - \chi^2_{reduced}$ = abs(57.23-63.12) = 5.84; $df_{diff} = df_{full} - df_{reduced}$= abs(20 – 13) = 7: χ^2 = 5.84, df = 7, p=0.55. Since there is not a significant difference in fit, the more parsimonious model is preferable. In addition, the CFI and TLI values for the 7-item model were slightly higher, though its RMSEA was slightly higher as well (Table 18).

Table 18. Exploratory and Confirmatory Factor Analyses and Item-to-Total Correlations of MMAS-8 from Medication Adherence Dataset			
	EFA	CFA	Item-total Correlation[a]
Forgetting (Item 1)	0.74	0.75	0.52
Any days (Item 2)	0.86	0.83	0.60
Cut-back (Item 3)	0.88	0.82	0.52
Travel (Item 4)	0.73	0.76	0.50
Yesterday (Item 5)	*0.17*	-	-
Under control (Item 6)	0.88	0.78	0.48
Inconvenience (Item 7)	0.59	0.68	0.43
Remembering (Item 8)	0.66	0.68	0.44
[a]Corrected item-to-total correlations using multi-trait scaling method.			

The corrected item-to-total correlations for this 7 item scale ranged between 0.43-0.60 (Table 17). The internal consistency reliability for this scale was acceptable at 0.78.

6.2 MEDICATION ADHERENCE AIM 2 RESULTS

Regarding the measurement invariance models for the MMAS-8 scale, the Metric Model (constrained factor loadings to be equal across groups) is not available due to the binary distribution of the items with the WLSMV estimator, since this model is not identified because residual variances vary across groups in this model. Differences in loadings in the Configural Model (all parameters freely estimated) between Black and White patients were most often

within the threshold for small differences (≥ 0.10 and <0.20); however, Item 6 had a difference in

loadings between groups of medium magnitude (≥ 0.20 and <0.30). (Table 19.)

Table 19. Configural Robust Weighted Least Squares Model Factor Loadings for White and Black Patients on the MMAS-8			
	Black	White	Difference
Forgetting (Item 1)	0.50	0.50	0.00
Any days (Item 2)	0.78	0.88	0.10
Cut-back (Item 3)	0.92	0.79	0.13
Travel (Item 4)	0.62	0.50	0.12
Under control (Item 6)	0.90	0.65	0.25
Inconvenience (Item 7)	0.35	0.38	0.03
Remembering (Item 8)	0.37	0.51	0.14

While the $\Delta\chi^2$ was statistically significant between the Configural and the Scalar Models, the

ΔCFI did not quite reach the threshold to reject the hypothesis of invariance. It is important to

note, however, that the model fit declined slightly with the more restrictive, Scalar Model.

Table 20. Measurement Invariance White vs. Black for MMAS-8									
Model[b]	Hypothesis	χ^2	df	p	CFI	$\Delta\chi^{2,a}$	Δdf	p	ΔCFI
Configural	Invariant factor structure	65.67	42	0.01	0.982	-	-	-	-
Scalar	Above + invariant factor loadings + invariant intercepts	89.92	54	0.002	0.973	25.70	12	0.01	0.009
[a]$\Delta\chi^2$ difference test for model in row above									
[b]All models estimated with Robust Weighted Least Squares.									

6.3 MEDICATION ADHERENCE AIM 3 RESULTS

In this study, the modified (7 item) MMAS-8's internal consistency reliability was only

0.78, not meeting the cut-off for good (≥ 0.80) or excellent (≥ 0.90) reliability. Internal

consistency reliability was 0.83 in a previous study, exceeding the cut-off for good reliability.[139]

Therefore, the Spearman Brown prophecy formula was used to determine the number of items

that could be added to improve this scale's reliability.

Table 21. Spearman Brown Prophecy Formula Calculations for Increased Reliability of the MMAS Scale			
Scale	Current α, *n* of Items	Total Items for Good Reliability (α = 0.80)	Total Items for Excellent Reliability (α = 0.90)
MMAS 7 items	0.78, 7	8	18

Adding an additional item to the 7 item scale would achieve good reliability. However, it is important to note that, under the Spearman Brown prophecy formula, the item added must be *correlated equally or more strongly than the other items* to achieve higher reliability. With this in mind, adding back in the "Yesterday" (Item 5) would actually reduce the reliability of the MMAS scale considerably, resulting in a Cronbach's alpha of 0.73. Table 22. shows Cronbach's alphas for the MMAS-8 when the item in question is removed. As we can see, removing "Yesterday" (Item 5) achieves the alpha of 0.78, as noted above. Therefore, a more reliable item than "Yesterday" (Item 5) would need to be added to achieve good reliability. Likewise, 11 more items of each or higher quality to achieve excellent reliability. Therefore, a similar recommendation is made regarding the use of IRT to develop large item banks that have high reliability, then creating shorter forms of equivalent reliability from those banks.

Table 22. Cronbach's Alpha for MMAS-8 with Deleted Items	
Item	α
Forgetting (Item 1)	0.71
Any days (Item 2)	0.69
Cut-back (Item 3)	0.69
Travel (Item 4)	0.70
Yesterday (Item 5)	**0.78**
Under control (Item 6)	0.70
Inconvenience (Item 7)	0.72
Remembering (Item 8)	0.72

CHAPTER 7. HEALTH-RELATED QUALITY OF LIFE RESULTS

7.1 HEALTH RELATED QUALITY OF LIFE AIM 1 RESULTS

Between 06/01/2015 and 05/31/2016, MEI collected 77,072 assessments of the KDQOL-36 with dialysis patients in the United States. Of these, 69,068 were assessments with unique patients, and 8,004 records were of patients with multiple assessments. After selecting the first assessment from those with multiple assessments, 72,982 assessments remained. A further 1,273 assessments were excluded due to uncertainty about the assessment date, and 2 assessments were eliminated for incompleteness, leaving 71,707 assessments from unique patients. Finally, an additional 585 patient assessments were omitted due to age < 18 years, and 336 were omitted due to being pre-dialysis or having received a previous transplant, leaving a final sample of 70,786 patient assessments for analysis.

Table 23. provides patient characteristics from this dataset. Patients were, on average, 62 years old, and the most prevalent racial/ethnic group was White (46%), with the next most common racial group being Black (27.2%). Most of the questionnaires were administered in English (88.3%), with another 10.2% administered in Spanish. Additionally, the majority of patients were using in-center hemodialysis (83.0%).

Table 23. Dialysis Patient Characteristics from Health-Related Quality of Life (HRQOL) Dataset (n= 70,786)	
Age (mean, SD, range)	61.5, 14.5, 18-100
Race/Ethnicity (%, n)	
White	31% (21,972)
Black	27% (18,855)
Asian	5% (3,414)
Native Hawaiian/Pacific Islander	2% (1,401)
American Indian/Alaska Native	1% (1,016)
Hispanic/Latino	19% (13,594)
Missing	15% (10,534)

Table 23. Continued	
Language of Survey (%, n)	
English	88% (62,489)
Spanish	10% (7,228)
Other	2% (1,062)
Dialysis Type (%, n)	
In-Center Hemodialysis	83% (58,764)
Peritoneal Dialysis	12% (8,535)
Conventional Home Hemodialysis	3% (2,293)
Other	2% (1,194)
Dialysis Access Site (%, n)	
AV Fistula	37% (26,499)
AV Graft	6% (4,166)
Venous Catheter	20% (14,273)
PD Catheter	11% (7,585)
Missing	26% (18,263)
Diabetes as Etiology of ESRD (%, n)	53% (37,246)
Employment Status	
Retired Due to Disability	31% (21,647)
Retired Due to Age/Preference	25% (17,515)
Unemployed	10% (6,903)
Employed Full Time	6% (4,435)
Employed Part Time	4% (2,447)
Homemaker	2% (1,578)
Other	2% (1,542)
Missing	21% (14,719)

Table 24 provides the distribution of KDQOL-36 scales' scores. The mean SF-12 PCS score was

37.8, and the SF-12 MCS score was 50.7. The means of the Burden, Symptoms/Problems, and

Effects scales were 52.3, 77.9, and 73.9, respectively. These scales showed relatively high

proportions at the ceiling, but relatively few at the floor.

Table 24. SF-12 Physical and Mental Health Component Scores and KDQOL Scales from Health-Related Quality of Life (HRQOL) Dataset						
	SF-12 PCS	SF-12 MCS	KDQOL Burden	KDQOL Symptoms	KDQOL Effects	SF-6D
n of observations	69,763	69,763	70,099	70,088	70,023	70,051
Mean	37.8	50.7	52.3	77.9	73.9	0.71
Standard Deviation	10.4	10.3	29.9	16.2	21.6	0.15
% at Floor	0%	0%	5%	0.04%	0.3%	0%
% at Ceiling	0%	0%	9%	4%	10%	3%
Minimum Score Observed	10.6	11.4	0	0	0	0.35
25th Percentile	29.6	43.6	25.0	68.1	59.3	0.60
50th Percentile (Median)	37.3	53.0	50.0	81.8	78.1	0.70
75th Percentile	46.1	58.9	75.0	90.9	90.6	0.82
Maximum Score Observed	65.9	72.0	100.0	100.0	100.0	1.0

Polychoric correlations between the items are presented in Appendix 2. Correlations were high among the Burden scale items and the Effects scale items. However, correlations among Symptoms/Problems items were somewhat diffuse. In particular, item 28, regarding the access site, did not demonstrate a large correlation with any other items (range of correlations 0.24-0.31).

Figure 13. Parallel Analysis of KDQOL Items
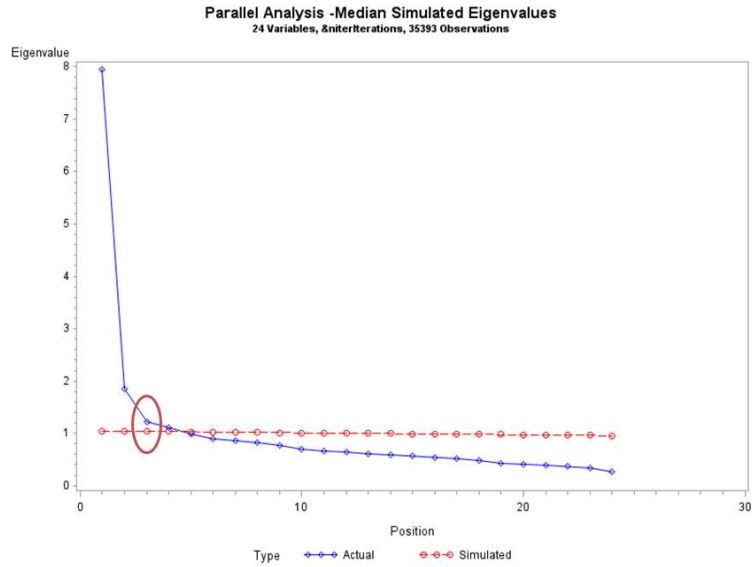
Parallel analysis suggested 3 factors for the KDQOL targeted items. (Figure 13) The top 5 eigenvalues in the EFA model were 21.18, 4.05, 1.79, 0.81, and 0.61. The scree plot (shown in Figure 13 as "Actual") indicated that 3 factors should be retained. The Tucker Lewis reliability coefficient for the 3 factor model was 0.86, indicating somewhat less than excellent fit. Therefore, the 3 factor model was chosen.

Table 25 shows the loadings and inter-factor correlations for this model. All items loaded on the factors put forth in the original KDQOL scales: i13-i16, Burdens of Kidney Disease; i17-i28, Symptoms/Problems with Kidney Disease; and i29-i36, Effects of kidney Disease.[1,187] Only

one item loaded lower than 0.40: item 28. There were no substantial (e.g., ≥ 0.40 on multiple factors) cross loadings. Finally, the inter-factor correlations were large, ranging from 0.49-0.65.

Table 25. Exploratory Factor Analysis of Kidney Disease Quality of Life (KDQOL)-36 Subscales from Health-Related Quality of Life (HRQOL) Dataset			
	Factor 1	*Factor 2*	*Factor 3*
My KD[a] interferes too much with my life (i13)[b]	**0.86**	-0.01	0.04
Too much time is spent dealing with KD (i14)[b]	**0.93**	-0.03	-0.02
I feel frustrated dealing with my KD (i15)[b]	**0.70**	0.05	0.13
I feel like a burden on my family (i16)[b]	**0.46**	0.10	0.21
Soreness in your muscles? (i17)[c]	0.01	**0.54**	0.12
Chest pain? (i18)[c]	0.04	**0.70**	-0.09
Cramps? (i19)[c]	0.02	**0.49**	0.02
Itchy skin? (i20)[c]	0.00	**0.54**	0.04
Dry skin? (i21)[c]	-0.03	**0.53**	0.10
Shortness of breath? (i22)[c]	0.03	**0.67**	-0.05
Faintness or dizziness? (i23)[c]	0.03	**0.64**	-0.01
Lack of appetite? (i24)[c]	-0.02	**0.53**	0.08
Washed out or drained? (i25)[c]	0.08	**0.55**	0.18
Numbness in hands or feet? (i26)[c]	-0.01	**0.54**	0.08
Nausea or upset stomach? (i27)[c]	-0.02	**0.62**	0.06
Problems with your access/catheter site? (i28)[c,e]	0.00	**0.30**	0.20
Fluid restriction? (i29)[d]	-0.01	0.06	**0.58**
Dietary restriction? (i30)[d]	-0.01	0.03	**0.65**
Your ability to work around the house? (i31)[d]	0.06	0.20	**0.54**
Your ability to travel? (i32)[d]	0.06	-0.03	**0.71**
Being dependent on doctors and other medical staff? (i33)[d]	0.08	-0.01	**0.72**
Stress or worries caused by kidney disease? (i34)[d]	0.18	0.10	**0.61**
Your sex life? (i35)[d]	0.01	-0.01	**0.60**
Your personal appearance? (i36)[d]	0.03	0.09	**0.66**
Factor 1	1.0	-	-
Factor 2	0.64	1.0	-
Factor 3	0.49	0.65	1.0
Note: Some items' wording have been reduced to fit table. [a]KD = Kidney Disease [b]Item stem is "How true or false is each of the following statements for you?" [c]Item stem is "During the past 4 weeks, to what extent were you bothered by each of the following?" [d]Item stem is "How much does kidney disease bother you in each of the following areas?" [e]For hemodialysis patients, access site is asked about; for peritoneal dialysis, catheter site is asked about.			

Given that the EFA reproduced the original KDQOL scale dimensionality, the CFA analyses were conducted on the full sample of patients. A 3 correlated factors CFA model for categorical variables using robust ML estimation was used. Table 26 shows this model, which fit the data very well. Satorra-Bentler χ^2 = 36,438.04, df = 249, p<0.001; RMSEA = 0.05; CFI = 0.99; TLI = 0.97.

Table 26. Confirmatory Factor Analysis of Kidney Disease Quality of Life (KDQOL)-36 Items			
	KDQOL Burden of Kidney Disease	KDQOL Symptoms/ Problems	KDQOL Effects of Kidney Disease
i13	0.87	-	-
i14	0.87	-	-
i15	0.84	-	-
i16	0.67	-	-
i17	-	0.65	-
i18	-	0.62	-
i19	-	0.53	-
i20	-	0.57	-
i21	-	0.59	-
i22	-	0.63	-
i23	-	0.64	-
i24	-	0.58	-
i25	-	0.73	-
i26	-	0.59	-
i27	-	0.66	-
i28	-	0.45	-
i29	-	-	0.61
i30	-	-	0.65
i31	-	-	0.72
i32	-	-	0.72
i33	-	-	0.76
i34	-	-	0.83
i35	-	-	0.59
i36	-	-	0.73

This model had superior fit in comparison to a multilevel model accounting for clustering within dialysis clinics: χ^2 = 96,949.86, df = 249, p<0.001; RMSEA = 0.07; CFI = 0.93; TLI = 0.92.

Finally, the corrected item-to-total correlations for each item to its subscale were high. All items were most correlated with the scales they were hypothesized to represent, as noted in bold font in Table 27. In some cases, large correlations were also observed between items and other scales. All 3 scales showed good internal consistency reliability, with Cronbach's alphas ≥ 0.80 but < 0.90. Facility-level reliabilities were slightly lower, but still acceptable to good, at 0.75, 0.76, and 0.83 for the Burdens, Symptoms/Problems, and Effects subscales, respectively. The average number of patients per dialysis center was 51. Both sets of reliabilities indicate that the KDQOL scales are reliable for use for comparisons of patient groups.

Table 27. Item-to-Total Correlations (Corrected for Overlap) of Kidney Disease Quality of Life (KDQOL)-36 Scales			
Item	KDQOL Burdens	KDQOL Symptoms/ Problems	KDQOL Effects
i13	**0.72***	0.39	0.52
i14	**0.73***	0.37	0.50
i15	**0.72***	0.42	0.53
i16	**0.56***	0.40	0.48
i17	0.33	**0.54***	0.43
i18	0.23	**0.44***	0.29
i19	0.25	**0.46***	0.33
i20	0.26	**0.51***	0.35
i21	0.28	**0.54***	0.39
i22	0.28	**0.51***	0.36
i23	0.28	**0.51***	0.36
i24	0.26	**0.45***	0.34
i25	0.43	**0.61***	0.53
i26	0.29	**0.50***	0.38
i27	0.29	**0.53***	0.39
i28	0.20	**0.31***	0.28
i29	0.36	0.39	**0.54***
i30	0.38	0.41	**0.58***
i31	0.47	0.53	**0.60***
i32	0.45	0.42	**0.62***
i33	0.47	0.45	**0.63***
i34	0.57	0.54	**0.68***
i35	0.34	0.33	**0.47***

Table 27. Continued			
i36	0.42	0.45	**0.59***
Cronbach's Coefficient Alpha	0.85	0.83	0.85
Center-Level Reliability[a]	0.75	0.76	0.83
Correlation with KDQOL Burdens of Kidney Disease Scale	1.0	-	-
Correlation with KDQOL Symptoms/Problems Scale	0.48	1.0	-
Correlation with KDQOL Effects of Kidney Disease Scale	0.62	0.62	1.0
Note: KDOQL scale for each item indicated by **bold font**.			
[a]Estimated from 1-way ANOVA partitioning between vs. within facility variance.			

7.2 HEALTH RELATED QUALITY OF LIFE AIM 2 RESULTS

As noted in the Methods Section 4.4, measurement invariance was examined for the 3 correlated factors KDQOL-36 using multiple-group CFA using the WLSMV estimator. Table 28 shows the loadings for White vs. Black patients for KDQOL-36 Burdens, Symptoms/Problems, and Effects of Kidney Disease scales. Many loadings were very similar across the racial groups and all fell below the range of small differences (<0.1).

Table 28. Configural Model Factor Loadings for White and Black Patients on Kidney Disease Quality of Life (KDQOL)-36 Burdens, Symptoms/Problems, and Effects of Kidney Disease Scales						
	KDQOL Burden of Kidney Disease		KDQOL Symptoms/ Problems		KDQOL Effects of Kidney Disease	
	White	Black	White	Black	White	Black
i13	0.87	0.87	-	-	-	-
i14	0.85	0.86	-	-	-	-
i15	0.84	0.85	-	-	-	-
i16	0.72	0.72	-	-	-	-
i17	-	-	0.65	0.66	-	-
i18	-	-	0.59	0.60	-	-
i19	-	-	0.49	0.54	-	-
i20	-	-	0.62	0.65	-	-
i21	-	-	0.65	0.68	-	-
i22	-	-	0.59	0.63	-	-
i23	-	-	0.59	0.65	-	-
i24	-	-	0.56	0.58	-	-
i25	-	-	0.77	0.78	-	-
i26	-	-	0.58	0.59	-	-
i27	-	-	0.64	0.64	-	-
i28	-	-	0.45	0.50	-	-
i29	-	-	-	-	0.63	0.68

Table 28. Continued						
i30	-	-	-	-	0.67	0.71
i31	-	-	-	-	0.75	0.75
i32	-	-	-	-	0.69	0.72
i33	-	-	-	-	0.74	0.76
i34	-	-	-	-	0.83	0.85
i35	-	-	-	-	0.58	0.59
i36	-	-	-	-	0.72	0.73

Model fit for both White and Black patients was also similar. White: $\chi^2 = 45{,}201.49$, $df = 249$, p<0.001; RMSEA = 0.08; CFI = 0.92; TLI = 0.92. Black: $\chi^2 = 25{,}476.63$, $df = 249$, p<0.001; RMSEA = 0.07; CFI = 0.93; TLI = 0.92. In Table 29., formal tests of measurement invariance are shown. As expected the χ^2 difference test between increasingly restricted models was significant, likely due to the sample size. However, and more importantly, CFI values increased with increasingly restrictive models, and the ΔCFI values fell below the critical value for rejecting the null hypothesis invariance at every step (0.01).

Table 29. Measurement Invariance White vs. Black for Kidney Disease Quality of Life (KDQOL)-36 Burdens, Symptoms/Problems, and Effects of Kidney Disease Scales

Model	Hypothesis	χ^2	df	p	CFI	$\Delta \chi^{2,a}$	Δdf	p	ΔCFI
Configural	Invariant factor structure	70,723.67	498	<0.001	0.925	-	-	-	-
Metric	Above + invariant factor loadings	65,648.63	519	<0.001	0.930	259.37	21	<0.001	0.005
Scalar	Above + invariant thresholds	61,263.48	588	<0.001	0.935	2442.80	69	<0.001	0.005
$\Delta \chi^2$ difference test for model in row above: Metric vs. Configural, Scalar vs. Metric									

7.3 HEALTH RELATED QUALITY OF LIFE AIM 3 RESULTS

Like the Treatment Decision-Making and Medication Adherence measures, the most important opportunity for improvement of the KDQOL subscales is to improve their reliability. Each of the scales achieved good reliability ($\geq$0.80), but not excellent reliability ($\geq$0.90). Given that these scales are used for individual assessment in dialysis centers, achieving scales with excellent reliability is an important way to improve them. Table 30 shows the Spearman Brown

Prophecy calculations for the number of items needed to reach excellent reliability for each of

the KDQOL scales.

Table 30. Spearman Brown Prophecy Formula Calculations for Increased Reliability of Kidney Disease Quality of Life (KDQOL)-36 Scales		
Scale	Current α, *n* of Items	Total Items for Excellent Reliability (α = 0.90)
KDQOL Burdens	0.85, 4	7
KDQOL Symptoms/ Problems	0.83, 12	23
KDQOL Effects	0.85, 8	13
Note: All scales already achieved good reliability.		

Additional study should seek to add more items to help increase the reliability of these scales.

The IRT methods described in the Aim 3 analyses above would also benefit the creation of

briefer and more reliable KDQOL scales.

CHAPTER 8. DISCUSSION

8.1 SUMMARY OF FINDINGS

This dissertation examined the psychometric properties of several patient-reported measures for use with chronic kidney disease patients. In most cases, the scales in their current form were acceptable for use. The largest departure from this trend was the MMAS-8, for which an item is recommended for removal. Otherwise, the primary recommendation is the use of item response theory (IRT) analyses to increase the reliability of the scales while keeping them brief.

Treatment Decision-Making. Regarding the LDKT measures, the LDKT Pros, Cons, and Self-Efficacy means were similar to the scores reported in their original validation study, and the dimensional structure of each of the original measures was evidenced in this current study among dialysis patients (in comparison to transplant patients examined in the original validation study).[74] That study recruited patients that had already presented for transplant evaluation, leading to the hypothesis that they would have higher LDKT Pros and Self-Efficacy, and lower LDKT Cons than patients in this study. Three separate samples were examined in that study: 1 for exploratory analyses (n = 135) and 2 for confirmatory models (n=145, n = 204). On the LDKT Pros scale, the means from these samples were 23.3 (SD=4.8), 23.8 (SD=5.2), and 26.4 (5.2), respectively. In the present study, the mean LDKT Pros score was 25.0 (SD=4.9). Similarly, the means on the LDKT Cons scale from the previous study were 18.1 (SD=5.7), 18.0 (SD=5.7), and 19.6 (SD=6.4), respectively, compared to 20.1 in the present study. Finally, the LDKT Self-Efficacy scale means from the previous study were 17.3 (SD=6.7), 17.1 (SD=6.5), and 21.5 (6.5), respectively, compared to 22.5 in the present study.

The similarity in mean scores across these two samples may have multiple substantive explanations. First, even though all of the patients in the previous sample had presented to a

transplant center (compared to only 40% in the current study) having committed the behavior of registering for transplant evaluation at a transplant center may indicate an increased readiness for transplant compared to dialysis patients who have not done so, but this readiness for transplant in general may not signal an increased readiness for living donor transplant, in which a family or friend donates the kidney to be transplanted. In other words, kidney patients who present for transplant evaluation may be comfortable with the signing-up for the kidney transplant waiting list, from which they could receive a transplant from a deceased donor when one is available, and they would likely be more ready for this type of transplant than kidney patients who do not come to the transplant center for transplant evaluation. However, patients presenting to be evaluated may not be any more ready for a transplant from a *living donor* transplant than those who have never come to the transplant center.

Previous research indicates that kidney patients may have many additional concerns about living donor transplant that they do not have about deceased donor transplant. For example, in a study of previous deceased donor kidney transplant recipients, Waterman and colleagues found that these patients may not have been willing to receive a transplant from a living donor because of their guilt over taking a living donor kidney, worry that they living donor would be harmed, and concern that the donor may need to donate to a closer friend or relative in the future.[188] Similarly, Salter and colleagues found that older kidney patients in particular were less comfortable asking for a living donor.[189] This evidence supports the hypothesis kidney patients at both dialysis centers (who have not presented to a transplant center) and those who have come to the transplant center for transplant evaluation may have similar levels of perceived pros, cons, and self-efficacy around living donor transplant.

A significant concern about the all the Treatment Decision-Making measures were the large proportions of patients at the theoretical maximum (ceiling). The scale with the lowest proportion at the ceiling was the LDKT Cons scale, with 7%, and the scale with the highest was the DDKT Pros scale, at 38%. Additionally, both the LDKT Pros and SE scales had >20% at the ceiling and the DDKT SE scale had nearly 20%. There could be multiple reasons why the scores of these scales are skewed toward the ceiling. First, it is possible that important content indicating concerns or barriers to transplant, both LDKT and DDKT, has not been covered in the scales. Additional studies should generate new items through focus groups with patients to help ensure that all concerns and barriers germane to patients are covered. The original items were not created with significant input from patients, so this addition would likely increase their validity.

All of these scales have 5 point response sets with Decisional Balance having "Not important," "Slightly important," "Moderately important," "Very important," and "Extremely important," and the Self-Efficacy measures having "Not at all confident," "Somewhat confident," "Moderately confident," "Very confident," and "Completely confident." Future studies should examine if these response sets could be increased to potentially increase variability and reduce the proportion of patients at the theoretical ceiling.

In the current study, the reliabilities of the LDKT Pros and Cons scales were lower than previously-reported internal consistency reliabilities from samples of patients pursuing transplant who were recruited from transplant centers. Reliabilities from these previous studies ranged from 0.78-0.86 for the LDKT Pros scale and 0.77-0.80 for the LDKT Cons scale.[74] In the present study, the LDKT Pros and Cons scales were 0.72 and 0.71, respectively, barely over the cut-off for acceptable reliability for group comparisons (0.70) and well below the cut-off for individual usage (0.90). Similar, though less severe, issues were observed for the DDKT Cons scale, which

was 0.76 in the original validation study,[73] but only 0.72 in the current study. Since these scales are employed in educational interventions for individual assessment, these scales reliabilities must be improved considerably. An approach for doing so using item response theory is discussed below.

Finally, as has been observed in previous studies of the LDKT and DDKT Decisional Balance measures (Pros and Cons), there were small, positive correlations between the Pros and Cons scales.[73,74] This result is unintuitive, since, in theory, Pros and Cons should be negatively correlated; i.e., the perception of an advantage to LDKT or DDKT should have the opposite effect of the perception of a disadvantage of the transplant options on a patient's underlying Decisional Balance. Under this theory, Pros and Cons items function similarly to positively and negatively worded items that represent a single underlying construct. This thinking is consistent with the proposition that many multi-dimensional measures are best modeled with structural models like the bifacator model, which allow for representation of a general factor for the underlying trait and multiple group factors representing direction of wording effects (vs. actual substantive traits). This hypothesis would suggest that a bifactor model for Decisional Balance would likely fit better than the two correlated factor models used in this dissertation. However, Reise and colleagues have argued that this approach is somewhat misguided, and that a more appropriate approach may be to determine the proportions of patients who are best modeled in a series of iteratively weighted least squares models as unidimensional, bifactor, or perhaps best not modeled at all due to non-sensical response patterns.[190] These approaches should be considered in future development of the LDKT and DDKT Decisional Balance measures.

MMAS-8. The mean MMAS-8 score observed in this study, 6.7 (SD=1.6), is similar to that reported in the original MMAS-8 validation study, which had a similar sample of patients.[139]

That study assessed the MMAS-8 with 1,367 hypertension patients and the mean score was 6.6 (SD=1.6).[139] In the present study, 77% of the patients are hypertensive, making these 2 studies very similar in sample composition.

The MMAS-8 scores reported in the present study were also very similar to those reported in a recent study of adherence to phosphate binders among dialysis patients. That study used alternate scoring with higher scores meaning worse adherence, so, in order to make appropriate comparisons, I have subtracted the scores reported in that study from 8. In that study, the mean MMAS-8 was 6.3 (SD=1.6).[158] That study also divided their sample into those with phosphorus levels under control (n=61) and those without phosphorus under control (n=18). Though those with their phosphorus under control had a significantly higher mean on the MMAS-8 (6.5 vs. 5.5, p<0.01), this difference was smaller than the 1.5 point difference others have cited as being clinically meaningful.[191] Further, a logistic regression model examining whether the MMAS-8, knowledge of phosphorus, and medication refill ratio found that the MMAS-8 was only marginally associated with phosphorus control [adjusted odds ratio = 0.71 (95% CI: 0.49-1.01); p=0.06].

Of all the scales considered in this study, the MMAS-8 was the only scale where the original factor structure was not supported. The factor analyses in this study suggested that the Item 5 "Yesterday," be removed since it had a very low loading (0.17) on the single factor. Polychoric correlations between this variable and the others in the scale were all <0.371 (range 0.01-0.19). For example, the correlation between this item and Item 1, "Forgetting," was -0.01. This result was surprising, since the patient sample for the present study was similar to the sample from the original validation study with hypertensive patients. The proportion of patients who responded indicating good adherence to this item was similar to the other items in the scale.

Further research is needed to determine how relevant this item is for hypertension and kidney disease medication regimens. Though many medications taken by dialysis patients, like phosphate binders, are prescribed for daily use, there may be no substantive negative health impact on the patient for skipping a single day and, therefore, no meaningful impact on adherence to the regimen. On the other hand, some medications relevant to CKD are taken more frequently than once daily, so a question about whether the medication was simply taken "yesterday" may not be sensitive enough to whether or not the regimen was followed. While the MMAS-8 is a generic scale of adherence to medications that is intended to apply to all medications used in the treatment of chronic and infectious conditions, strong consideration should be given to whether tailoring is needed for specific medications and conditions.

Additionally, the MMAS-8 showed borderline evidence of measurement inequality between Black and White patients, with a p-value of 0.01 between the Configural and Scalar models, and a ΔCFI value of 0.009 (threshold = 0.01). In an analysis of the difference in loadings between Black and White patients, the difference in loadings on Item 6 ("Under Control") was of a medium magnitude. Qualitative investigations are recommended to determine whether there is a substantive explanation for why Black and White patients may respond differently to this item, and to determine corrective changes to item wording.

KDQOL-36. The mean KDQOL-36 scale scores observed in this study were higher than those reported by the Dialysis Outcomes and Practice Patterns Study (DOPPS) for their United States sample.[172] (Table 31.) Conducted between 1996 and 1999, the DOPPS study had a sample of 2,885 patients from the U.S. on hemodialysis only.[192] Though some of the sample characteristics between the DOPPS and the current study are similar (e.g., age), all DOPPS participants were on hemodialysis while approximately 10% of participants in the current study

were on peritoneal dialysis; these two different dialysis modalities may entail different levels of quality of life.

Table 31. SF-12 Physical and Mental Health Component Scores and Kidney Disease Quality of Life (KDQOL)-36 Scales from Health Related Quality of Life Dataset Compared to Dialysis Outcomes and Practice Patterns Study (DOPPS) US Sample		
	Current Study	**DOPPS US**
SF-12 PCS	37.8	33.1
SF-12 MCS	50.7	46.6
KDQOL Burdens	52.3	40.8
KDQOL Symptoms/Prob.	77.9	71.1
KDQOL Effects	73.9	62.5

However, another explanation for the higher scores among the present study regards the potential improvement in dialysis care in the more than 20 years since the DOPPS study was conducted. Figure 14. shows the increase in 5 year survival among hemodialysis and peritoneal dialysis patients initiating treatment between 1996 and 2009. During this time, 5 year survival among hemodialysis patients increased from 33% to 42%, with even greater gains observed among peritoneal dialysis patients.

Figure 14. Increasing 5 Year Survival on Dialysis
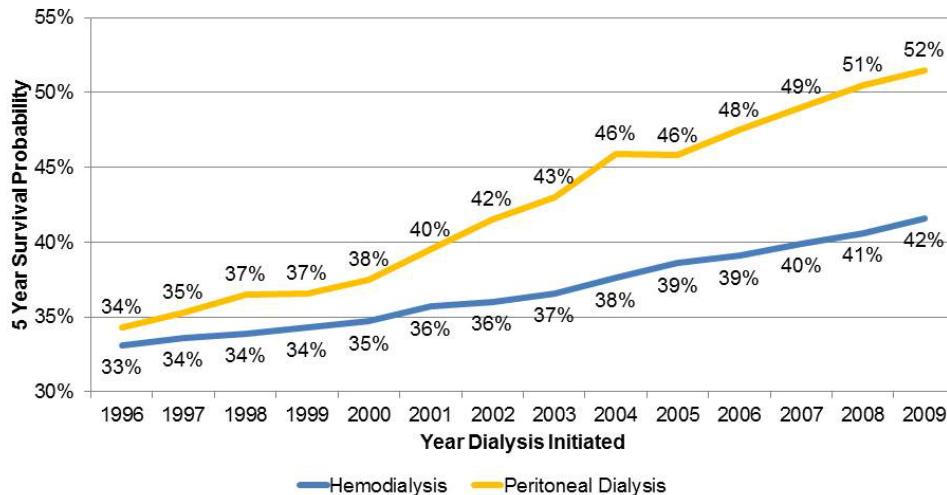
In addition to the KDQOL subscales, the SF-6D scores in the present study were very similar to those from a study of 150 ESRD patients in Singapore, which estimated a mean SF-6D of 0.70, a standard deviation of 0.14, an observed minimum of 0.37, 3% at the ceiling (score = 1).[193] A recent study reported that the mean SF-6D score in the U.S. general population of individuals reporting no medical conditions was 0.885.[194]

Support for the KDQOL-36's dimensional structure with 3 subscales was found. However, in some cases, large correlations were also observed between items and other scales. For example, item i31 (your ability to work around the house), part of the Effects scale, correlated with the Effects scale at 0.60, but is also correlated with the Burdens and Symptoms/Problems scales at 0.47 and 0.53, respectively. There were also large correlations among the scales' scores (Table 26.) These results indicate that, while the 3 KDQOL kidney-targeted scales function provide unique information, they are highly related to one another, or to a more general underlying factor. Future work should explore a bifactor model to determine if there is evidence of a more general, underlying kidney disease quality of life factor influencing these three factors.

8.2 INCREASING THE RELIABILITY OF SCALES

For many of the scales examined in this study, though acceptable or good reliability was observed, insufficient reliability was observed for individual usage (excellent reliability, ≥ 0.90). Therefore, recommendations were made in Aim 3 for future studies to create larger banks of potentially new items to attain excellent reliability. Then, starting from these item banks that have achieve high reliability, short forms that retain the item banks' reliability can be created. There are two primary statistical approaches to this research that will be discussed below, starting with the use of Item Response Theory (IRT).

As opposed to the correlational methods used in classical test theory (CTT), IRT puts

forth a series of statistical models to describe the association between an individual's underlying

trait level and an observed score on a scale. The various models used in IRT will each estimate a

different number of parameters, and most often take the form of logistic regression models. IRT

models estimate the individual's level of the latent trait directly (notated by Θ), and can

determine potentially varying item properties for respondents at different trait levels.

In IRT, item information is a critical parameter that is directly related to reliability.

Fundamentally, item information is a measure of the precision an item, or group of items, has at

various levels of the latent trait,[107] such that the information contributed by an item has an

inverse relationship with its standard error. Indeed, a key advantage to the IRT approach to

reliability is estimated for each level of the latent trait measured by a scale, not only for the scale

as a whole. Item information is calculated differently depending on the number parameters in the

model. For example, in a 2 parameter logistic model with parameters for item difficulty (*b*) and

discrimination (α), a 2PL model, the information function [*I Θ*] for item *i* is:

$$[I\ \theta] = \alpha_i^2 P_i(1 - P_i)$$

Where P_i is the proportion of respondents with a given level of trait ability who endorse *i*. In

IRT, items are most precise when their difficulty is similar to the respondent's level of trait

ability, or as Θ approaches *b*. The sum of information for each item in a scale is the total scale's

information. Another useful way to formalize item information (INF) in IRT models is as the

inverse of each item's error variance: or 1/*d*, where *d* is the error variance. Relatedly, each item's

standard error (SE) can be defined in terms of its information as: $SE = \frac{1}{\sqrt{INF}}$. This leads to the

definition of reliability in the context of IRT (for 2 parameter models):

$$Reliability = \frac{(INF - 1)}{INF} = 1 - SE^2$$

Figure 15. shows an example of a test information curve calculated from 18 transplant knowledge items. As we can see, its information peaks where Ө is approximately -1, and is lowest at the tails.

Figure 15. Example Test Information Function from 2PL IRT Model
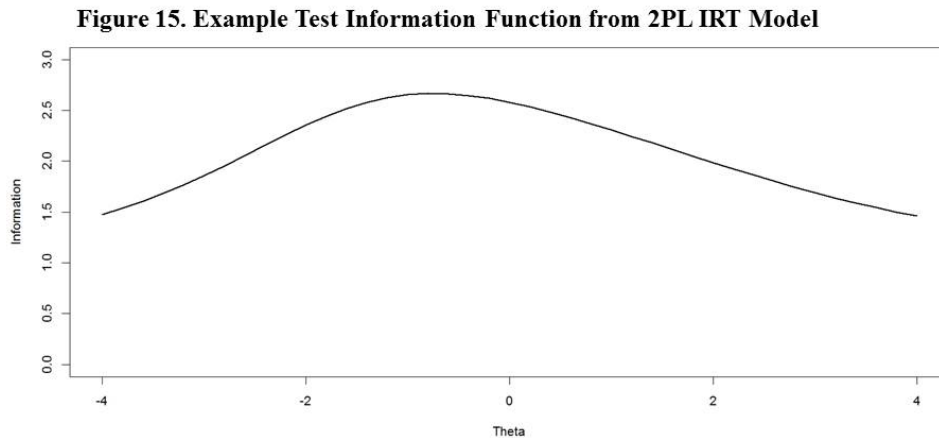
The IRT approach to creating large banks of items with high reliability then shortening these banks into briefer short forms is an approach taken by the NIH PROMIS project[24] and has multiple steps.[109] First, IRT models with item parameters estimated must first be fit so that each item's information can also be estimated. Then, a preset level of item information to achieve for the full scale is chosen for a given level of trait ability; i.e., $I(Ө) = 1/d$. For example, if we want to make sure the short form has excellent reliability at trait level = 0 (e.g., on a z distribution), we would choose the value of scale information equivalent to a reliability of 0.90 such that $I(Ө_0) = 1/d$ and $1/d$ is equivalent to a scale reliability of 0.90.

Next, the item with the highest information is chosen for the scale. Next, from the remaining items, the item with the highest information is selected. This process is continued until the smallest number of items is found that reaches $1/d$. Often, it is desirable to find the shortest scale

97

at multiple values of trait ability, so this process will need to be repeated multiple times, once at each level of Θ we desire to achieve the preset value of $1/d$ for, then we find the shortest scale that has an information of $1/d$ at each of the levels of Θ under consideration. Formally, the process can be conceived of as an optimization problem, with the computation:

$$n = \sum_{i}^{N} x_i$$

Where N is the number of items in the item bank, $x_i = 1$ if the item is included due to having $I(\Theta)$ = $1/d$ for all levels of Θ_k ($k = 1, 2, 3, 4, \ldots m$ levels of trait ability), and $x_i = 0$ if the item is not included.

The PROMIS project used a similar approach to create short forms from their initial item banks.[24] For example, in Wave 1 of their measure creation, this project created a bank of 124 items for its Physical Function domain and reduced the bank to several short forms of various lengths (e.g., 10 items). After the short forms were created, correlations between the full item banks and the short forms were examined as a measure of the extent to which the short form covers the content of the full item bank. Table 32 shows the PROMIS Wave 1 domains with the number of items in the full bank, number of items in the short form, and the correlation between the two. All correlations were ≥ 0.95.

Table 32. PROMIS Wave 1 Item Banks, Short Forms, and Correlations				
Domain	*n* of items in bank	*n* of items in short form	Difference	Correlation
Emotional distress – anger	29	8	21	0.96
Emotional distress – anxiety	29	7	22	0.96
Emotional distress – depression	28	8	20	0.96
Pain behavior	39	7	32	0.98
Pain interference	41	6	35	0.95
Physical function	124	10	114	0.96
Satisfaction with social activities	12	7	5	0.99
Satisfaction with social roles	14	7	7	0.99
Sleep disturbance	27	8	19	0.96
Sleep-related impairment	16	8	8	0.98
Median	**29**	**8**	**21**	**0.96**

Rounding-up, the median item reduction from item banks to short forms in PROMIS was 21 items, with a median correlation between the two of 0.96. The PROMIS project shows how successfully larger item banks can be reduced while retaining extensive amounts of item content from the original banks.

One drawback of using IRT to increasing reliability regards the potential complexity of scoring scales under this approach. With the exception of the simplest, 1 parameter IRT models (e.g., the Rasch model), scoring for IRT models cannot be accomplished using simple, summed scores.[109] The additional complexity of scoring IRT-based measures may create significant barriers for clinical application, as clinicians, who are often not trained to administered and score any PRMs, will likely not have the appropriate training to score IRT-based measures. Current measurement systems using IRT, like PROMIS, have made an effort to overcome this issue by creating look-up tables or encouraging the use of electronic data capture systems that automatically score measures.

8.3 IMPROVING RELIABILITY ESTIMATION

It is also necessary to note strategies within CTT to improve reliability estimation of measures. Reflecting on the Spearman-Brown analyses offered in this dissertation, it may be

clear that a straightforward approach to increasing reliability under the CTT framework involves increasing the number of items in the measure. As others have noted, one drawback of increasing items is the tendency to introduce redundant content; i.e., new items with similar content but different wording than items already in the scale.[195] While introducing redundant item content may increase reliability, it does not increase scale information. In addition, for scales used in clinical settings, brevity is needed and additional length may dissuade many users from adopting the measure. Finally, even if these constraints are not present, Spearman-Brown prophesy estimates assumed parallel tests without much heterogeneity between items, and this assumption may not be realistic in most data.

Further, it is also known that heterogeneity between items within a composite (e.g. unequal loadings, unequal variances) can result in significant underestimation of true reliability by coefficient alpha.[196] Therefore, the need to "increase reliability" under CTT may be side-stepped by improving reliability estimation with alternative approaches. One such approach is stratification of scale content.[196] This process involves creating subscales containing items of similar or correlated content. Then a stratified alpha is calculated[197] using the following formula:

$$Stratified\ alpha = 1 - \frac{\sum \sigma_i^2(1 - \alpha_i)}{\sigma_x^2}$$

Where σ_i^2 is the variance of the ith subscale and α_i is the coefficient alpha of the ith subscale. This approach will improve reliability estimation when the correlations are high within subscales and the correlations are low between the subscales.

The second approach considered is maximal reliability, or a method of weighting subscales within a composite scale.[196] Maximal reliability (R_k^*) is calculated using this formula:

$$R_k^* = \frac{A}{(K/(1 + (K - 1)\rho) + A)}$$

In this formula, K = the number of subscales and A is defined as:

$$A = n_1\rho_1/(1 - \rho_1) + n_2\rho_2/(1 - \rho_2) + \cdots n_k\rho_k/(1 - \rho_k)$$

Where n_i is the number of items in subscale i, ρ_i is the reliability of subscale i, and ρ is the common correlation between the subscales.

In a simulation study comparing the performance of multiple alternative reliability estimates under several scenarios assuming different levels of homogeneity and heterogeneity, Osburn found that as heterogeneity increases, coefficient alpha generates increasingly worse estimates of true reliability, while the stratified alpha and maximal reliability generated much better estimates.[196] Table 33. shows the estimates of reliability from coefficient alphas, stratified alpha, and maximal reliability from Osburn's study. The scenarios considered include Tau equivalence (one factor with equal factor loadings, unequal variances), Congeneric equivalence (one factor with unequal factor loadings, unequal variances), Slight heterogeneity (two factors with highest correlation between factors), Moderate heterogeneity (two factors with middle correlation between factors), and Strong heterogeneity (two factors with lowest correlations between factors).

Table 33. Comparison of Reliability Estimates from Osburn Study: Reliability Estimate (Absolute Difference from True)					
	Tau equivalence	Congeneric equivalence	Slight Hetero-geneity	Moderate Hetero-geneity	Strong Hetero-geneity
True reliability	0.798 (ref)	0.786 (ref)	0.781 (ref)	0.760 (ref)	0.703 (ref)
Coefficient alpha	0.798 (0)	0.763 (0.023)	0.694 (0.087)	0.570 (0.190)	0.234 (0.469)
Stratified alpha	0.798 (0)	0.786 (0)	0.781 (0)	0.760 (0)	0.703 (0)
Maximal reliability	0.801 (0.003)	0.801 (0.015)	0.784 (0.003)	0.764 (0.004)	0.708 (0.005)

In addition to these straightforward approaches to increasing reliability under CTT, more advanced approaches that account for specific variance and employ model-based estimates of reliability give superior estimates of reliability in comparison to coefficient alpha under circumstances where parallel tests cannot be assumed, which is likely most cases involving real data. Bentler recently argued that the failure to account for specific variance in coefficient alpha underestimates reliability, and that structural equation-based approaches to account for specific variance can overcome this bias.[198] This work even provides a specificity-enhanced version of maximal reliability, which may improve reliability estimates even further than those shown from maximal reliability in Osburn's study.

8.4 LIMITATIONS

Though multiple domains of PRMs used with CKD patients were covered in this study, several limitations should be considered. First, though a unifying conceptual model was created to include all three major sets of constructs represented in this study – Treatment Decision-Making, Medication Adherence, and Health-Related Quality of Life – a single dataset containing all of these measures could not be obtained. Rather, 3 separate datasets, one for each set of constructs, were used. Despite missing the opportunities that would have been afforded in a

unified dataset, the use of the 3 separate datasets did not detract from meeting the study aims. None of the aims were to test hypotheses about the theoretical relationships between the constructs specified in the conceptual model offered by this study, and the aims were restricted to examining and interpreting psychometric properties of specific PRMs covering each construct or set of constructs considered.

Second, and related to the first limitation, not all constructs for which important PRMs could be examined in the conceptual model offered in this study had measures to examine in the datasets obtained for this study. Several other constructs, like patients' experience with care and patient satisfaction with care, for example, are very important in the field of CKD and would have made excellent complements to the PRMs considered in this study.[199] Future expansions of this study will seek to evaluate PRMs covering these additional, important constructs.

Third, other types of data that would have added a great deal of value to the analyses conducted in this study were not available. A prime example includes the specification of medications taken by CKD patients in the Medication Adherence dataset could have been useful in understanding if the under-performance of Item 5, "Yesterday," in comparison to other studies of this measure was related to the use of medications for which a previous day's dose is not relevant to overall adherence.

Finally, all the datasets considered were not collected as probability samples, and may have some bias and not represent any larger CKD population very well. At best, the extent to which the samples in the datasets used in this study represent larger populations of CKD patients (e.g., the U.S. national population of dialysis patients) is unknown. The probability for bias varies between the datasets. The Treatment Decision-Making dataset was collected as part of a

randomized controlled trial with no consideration for external validity, and is relatively small, so this dataset has a relatively high probability of at least some bias. On the other hand, the HRQOL dataset was collected as part of a clinical activity in which all dialysis patients are mandated to participate, and is relatively large, so it has likely has a much lower probability for bias.

8.5 NEXT STEPS

In addition to the opportunities to improve the measures considered in this study that were outlined in Aim 3, a few other suggestions for the future are worth mentioning. In particular, the large dataset used for the HRQOL study makes possible an additional scope of work not covered in this dissertation. Though the dataset on its own may not be nationally representative, *post-hoc* statistical approaches can be applied to help the data better represent the U.S. national dialysis population. The PROMIS study used this approach, which uses raking to generate weights for observations in the original dataset that match the marginal distribution of key characteristics of a population of interest. In the PROMIS study, a subsample was created to match the U.S. general population's distribution of gender *x* age *x* race/ethnicity *x* educational attainment *x* marital status *x* income.[200] Using the United States Renal Data Service's (USRDS) estimates of characteristics for the U.S. national dialysis population,[148] a nationally representative dataset can be created from the one used in this study.

The nationally-representative dataset could be used to serve the critical purpose of generating updated normative scores for the KDQOL subscales. As mentioned in Section 8.1 above, the previous norms published for the KDQOL subscales may now be out of date. Since these norms are used by groups like the Medical Education Institute to tailor health-improvement interventions, having updated norms is important. In addition to national norms for the KDQOL

scales, effect sizes for differences between key clinical subgroups (e.g., dialysis type) can be created and employed in clinical trials.

For the Treatment Decision-Making and MMAS-8 measures, estimation of item parameters in future studies may be of particular interest. Individual items from these scales can be used to flag patients with different levels of key underlying traits (e.g., self-efficacy) then select educational content most relevant and appropriate for them. With this in mind, knowing these items' difficulty and discrimination could be useful. For example, knowing which particular items discriminate those with very high, moderate, and very low levels of self-efficacy to pursue LDKT so that educational messages or intervention approaches with differing levels of intensity can be incorporated into a tailored intervention. Future studies with larger datasets and more items to select from for these measures may be the most appropriate way to pursue this extension of the current research.

Finally, a CTT-based study to determine if the alternative reliability estimates described above can improve reliability estimation among common health measures in CKD has high potential to contribute to the patient-reported health measures literature. First, as advances in psychometrics applied to health measures has strongly trended toward IRT (using the rationale that IRT may be a superior way to increase reliability), determining whether or not coefficient alpha and other similar estimates of internal consistency reliability are under-estimating reliability of health measures would help determine whether or not the additional work required to apply IRT is worthwhile. Indeed, as the alternative reliability estimates described in this dissertation are easily applicable to current CTT-based health measures, using these estimates would be much easier than adopting IRT if it turns out that they estimate reliability better for

such measures. More generally, these alternative methods of estimating reliability are not well known in health measurement, so their introduction would be a benefit.

8.6 CONCLUSIONS

This dissertation found support for the current dimensional structure and at least acceptable reliability for the majority of measures considered. With this evidence in mind, the use of the living and deceased donor Treatment Decision-Making and KDQOL scales with CKD patients is recommended. Regarding the MMAS-8 scale, the 8 item one dimensional structure was not supported. Additional research is required to determine how this scale may be best used to assess medication adherence with CKD patients. For almost all the scales examined in this dissertation, increased reliability would be desirable, and future research should concentrate on this effort. The continued improvement of these already good patient-reported measures would increase their utility in clinical assessment and population-level surveillance to better understand dimensions of CKD patients' health and health-care from their own perspective.

Appendixes

Appendixes

Appendix 1. KDQOL Item Guide		
Short Name	**Item**	**Response Options**
How true or false is each of the following statements for you?		
KDQOL item13	My kidney disease interferes too much with my life	"Definitely true - "Definitely false"
KDQOL item14	Too much of my time is spent dealing with my kidney disease	"Definitely true - "Definitely false"
KDQOL item15	I feel frustrated dealing with my kidney disease	"Definitely true - "Definitely false"
KDQOL item16	I feel like a burden on my family	"Definitely true - "Definitely false"
During the past 4 weeks, to what extent were you bothered by each of the following?		
KDQOL item17	Soreness in your muscles?	"Not at all bothered"-"Extremely bothered"
KDQOL item18	Chest pain?	"Not at all bothered"-"Extremely bothered"
KDQOL item19	Cramps?	"Not at all bothered"-"Extremely bothered"
KDQOL item20	Itchy skin?	"Not at all bothered"-"Extremely bothered"
KDQOL item21	Dry skin?	"Not at all bothered"-"Extremely bothered"
KDQOL item22	Shortness of breath?	"Not at all bothered"-"Extremely bothered"
KDQOL item23	Faintness or dizziness?	"Not at all bothered"-"Extremely bothered"
KDQOL item24	Lack of appetite?	"Not at all bothered"-"Extremely bothered"
KDQOL item25	Washed out or drained?	"Not at all bothered"-"Extremely bothered"
KDQOL item26	Numbness in hands or feet?	"Not at all bothered"-"Extremely bothered"
KDQOL item27	Nausea or upset stomach?	"Not at all bothered"-"Extremely bothered"
KDQOL item28	Problems with your access/catheter site?	"Not at all bothered"-"Extremely bothered"
Some people are bothered by the effects of kidney disease on their daily life, while others are not. How much does kidney disease bother you in each of the following areas?		
KDQOL item29	Fluid restriction?	"Not at all bothered"-"Extremely bothered"
KDQOL item30	Dietary restriction?	"Not at all bothered"-"Extremely bothered"
KDQOL item31	Your ability to work around the house?	"Not at all bothered"-"Extremely bothered"
KDQOL item32	Your ability to travel?	"Not at all bothered"-"Extremely bothered"
KDQOL item33	Being dependent on doctors and other medical staff?	"Not at all bothered"-"Extremely bothered"
KDQOL item34	Stress or worries caused by kidney disease?	"Not at all bothered"-"Extremely bothered"
KDQOL item35	Your sex life?	"Not at all bothered"-"Extremely bothered"
KDQOL item36	Your personal appearance?	"Not at all bothered"-"Extremely bothered"

Appendix 2. Polychoric Correlations between KDQOL Subscale items from HRQOL Dataset

	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
item13	1.0																							
item14	**.81**	1.0																						
item15	**.70**	**.72**	1.0																					
item16	**.54**	**.54**	**.65**	1.0																				
item17	.32	.30	.34	.31	1.0																			
item18	.27	.27	.31	.28	**.42**	1.0																		
item19	.25	.24	.25	.21	**.42**	.36	1.0																	
item20	.26	.24	.26	.25	.33	.30	.32	1.0																
item21	.27	.26	.29	.27	**.37**	.30	.30	**.70**	1.0															
item22	.29	.27	.30	.28	**.37**	**.54**	.31	.33	.35	1.0														
item23	.29	.29	.33	.31	.36	**.45**	.34	.31	.31	**.45**	1.0													
item24	.26	.24	.30	.29	.32	.37	.21	.28	.32	.35	.37	1.0												
item25	**.41**	**.39**	**.44**	**.40**	**.49**	**.39**	.35	.36	**.40**	**.47**	**.48**	**.48**	1.0											
item26	.28	.27	.28	.30	**.43**	.35	.35	.34	.36	.36	**.39**	.29	**.43**	1.0										
item27	.29	.27	.33	.30	**.38**	**.43**	.31	.33	.35	**.38**	**.47**	**.50**	**.48**	**.37**	1.0									
item28	.24	.24	.27	.24	.26	.30	.25	.25	.25	.27	.29	.25	.29	.28	.30	1.0								
item29	.37	.35	.36	.31	.32	.27	.29	.26	.29	.29	.27	.19	.35	.30	.27	.25	1.0							
item30	**.39**	**.38**	**.39**	.32	.32	.28	.27	.27	.30	.27	.29	.27	.37	.28	.30	.27	**.68**	1.0						
item31	**.47**	**.43**	**.45**	**.46**	**.44**	.33	.27	.30	.36	**.39**	**.37**	.37	**.53**	**.40**	.37	.27	**.42**	**.45**	1.0					
item32	**.48**	**.45**	**.43**	**.38**	.34	.25	.25	.27	.31	.29	.30	.28	**.42**	.30	.31	.28	**.43**	**.47**	**.59**	1.0				
item33	**.48**	**.47**	**.50**	**.45**	.36	.29	.25	.28	.32	.32	.33	.31	**.45**	.32	.35	.32	**.43**	**.48**	**.55**	**.59**	1.0			
item34	**.55**	**.54**	**.63**	**.53**	**.42**	.36	.32	.32	.36	.36	**.40**	.37	**.52**	.36	**.42**	.35	**.46**	**.51**	**.55**	**.56**	**.66**	1.0		
item35	.36	.34	.36	.34	.27	.23	.22	.23	.24	.24	.25	.23	.32	.26	.26	.22	.33	.35	**.41**	**.44**	**.43**	**.49**	1.0	
item36	**.44**	**.43**	**.47**	**.46**	.36	.33	.27	.31	.36	.31	.33	.34	**.44**	.31	**.38**	.31	**.41**	**.46**	**.52**	**.51**	**.54**	**.61**	**.54**	1.0

Correlations in blue represent the KDQOL Burden scale, those in green represent the Symptoms/Problems scale, and those in orange represent the Effects scale.
Bold correlations are large according to Cohen's rule, $\geq$ .371

REFERENCES

1.   Hays RD, Kallich JD, Mapes DL, Coons SJ, Carter WB. Development of the kidney disease quality of life (KDQOL) instrument. *Qual Life Res.* 1994;3(5):329-338.

2.   Edgell ET, Coons SJ, Carter WB, et al. A review of health-related quality-of-life measures used in end-stage renal disease. *Clin Ther.* 1996;18(5):887-938.

3.   Weidmer BA, Cleary PD, Keller S, et al. Development and evaluation of the CAHPS (Consumer Assessment of Healthcare Providers and Systems) survey for in-center hemodialysis patients. *Am J Kidney Dis.* 2014;64(5):753-760.

4.   USRDS Coordinating Center. United States Renal Data System. 2016; https://www.usrds.org/. Accessed December 26th, 2016.

5.   Cohen SJ, Punt CJA, Iannotti N, et al. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients With Metastatic Colorectal Cancer. *J Clin Oncol.* 2008;26(19):3213-3221.

6.   Lin J-C, Jan J-S, Hsu C-Y, Liang W-M, Jiang R-S, Wang W-Y. Phase III Study of Concurrent Chemoradiotherapy Versus Radiotherapy Alone for Advanced Nasopharyngeal Carcinoma: Positive Effect on Overall and Progression-Free Survival. *J Clin Oncol.* 2003;21(4):631-637.

7.   Hayes DF, Cristofanilli M, Budd GT, et al. Circulating Tumor Cells at Each Follow-up Time Point during Therapy of Metastatic Breast Cancer Patients Predict Progression-Free and Overall Survival. *Clin Cancer Res.* 2006;12(14):4218-4224.

8.   Han M, Partin AW, Pound CR, Epstein JI, Walsh PC. Long-term biochemical disease-free and cancer-specific survival following anatomix radical retropublic prostatectomy: The 15-Year Johns Hopkins Experience. *Urol Clin North Am.* 2001;28(3):555-565.

9.   Massie AB, Segev DL. Rates of False Flagging Due to Statistical Artifact in CMS

        Evaluations of Transplant Programs: Results of a Stochastic Simulation. *Am J*

        Transplant. 2013;13(8):2044-2051.

10.   United States Department of Health and Human Services Food and Drug Administration.

        Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product

        Development to Support Labeling Claims. Washington DC: Center for Drug Evaluation

        and Research; December 2009.

11.   Fung CH, Hays RD. Prospects and challenges in using patient-reported outcomes in

        clinical practice. *Qual Life Res.* 2008;17(10):1297-1302.

12.   Suchman EA, Phillips BS, Streib GF. An Analysis of the Validity of Health

        Questionnaires. *Soc Forces.* 1958;36(3):223-232.

13.   Hays RD, Spritzer KL, Thompson WW, Cella D. U.S. General Population Estimate for

        "Excellent" to "Poor" Self-Rated Health Item. *J Gen Intern Med.* 2015;30(10):1511-

        1516.

14.   Dicicco L, Apple D. Health needs and opinions of older adults. *Public Health Rep.*

        1958;73(6):479-487.

15.   Brook RH, John E. Ware J, Rogers WH, et al. *The Effect of Coinsurance on the Health of*

        Adults. Santa Monica, CA: RAND;1984.

16.   Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical

        Outcomes Study. An application of methods for monitoring the results of medical care.

        JAMA. 1989;262(7):925-930.

17.   Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. *JAMA.* 1989;262(7):907-913.

18.   Hays RD, Sherbourne CD, Manzel RM. *User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life.* Santa Monica, CA: RAND;1995.

19.   Ware J, Sherbourne C. The MOS 36 -item -short-form health survey (SF-36). *Med Care.* 1992;30.

20.   Ware J, Kosinski MA, Keller SD. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales.* Boston, MA: The Health Institute, New England Medical Center;1995.

21.   Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007;45(5 Suppl 1):S3-S11.

22.   PROMIS. 2016; http://www.healthmeasures.net/explore-measurement-systems/promis. Accessed December 26th, 2016.

23.   National Institutes of Health. *A Decade of Discovery: The NIH Roadmap and Common Fund.* Bethesda, MD: National Institutes of Heath;2014.

24.   Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.

25.   Neuro-QOL. 2016; http://www.healthmeasures.net/explore-measurement-systems/neuro-qol. Accessed December 26th, 2016.

26.   Hays RD, Spritzer KL, Amtmann D, et al. Upper-Extremity and Mobility Subdomains From the Patient-Reported Outcomes Measurement Information System (PROMIS) Adult Physical Functioning Item Bank. *Arch Phys Med Rehabil.* 2013;94(11):2291-2296.

27.   ASCQ-Me. 2016; http://www.healthmeasures.net/explore-measurement-systems/ascq-me. Accessed December 26th, 2016.

28.   NIH Toolbox. 2016; http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox. Accessed December 26th, 2016.

29.   Paz SH, Slotkin J, McKean-Cowdin R, et al. Development of a vision-targeted health-related quality of life item measure. *Qual Life Res.* 2013;22(9):2477-2487.

30.   United States Renal Data System. *2015 annual data report: An overview of the epidemiology of kidney disease in the United States.* Bethesda, MD: National Institute of Health, National Institute of Diabetes and Digestive Diseases;2015.

31.   Kucirka LM, Grams ME, Balhara KS, Jaar BG, Segev DL. Disparities in provision of transplant information affect access to kidney transplantation. *Am J Transplant.* 2012;12(2):351-357.

32.   Salter ML, Kumar K, Law AH, et al. Perceptions about hemodialysis and transplantation among African American adults with end-stage renal disease: inferences from focus groups. *BMC Nephrol.* 2015;16(1):49.

33.   Waterman A, Hyland S, Goalby C, Robbins M, Dinkel K. Improving transplant education in the dialysis setting: The "Explore Transplant" initiative. *Dial Transplant.* 2010;39(6):236-241.

34.   Rodrigue JR, Paek MJ, Egbuna O, et al. Making House Calls Increases Living Donor
       Inquiries and Evaluations for Blacks on the Kidney Transplant Waiting List.
       Transplantation. 2014.

35.   Rodrigue JR, Cornell DL, Kaplan B, Howard RJ. A randomized trial of a home-based
       educational approach to increase live donor kidney transplantation: effects in blacks and
       whites. *Am J Kidney Dis.* 2008;51(4):663-670.

36.   Strigo TS, Ephraim PL, Pounds I, et al. The TALKS study to improve communication,
       logistical, and financial barriers to live donor kidney transplantation in African
       Americans: protocol of a randomized clinical trial. *BMC Nephrol.* 2015;16(1):160.

37.   Boulware LE, Hill-Briggs F, Kraus ES, et al. Effectiveness of educational and social
       worker interventions to activate patients' discussion and pursuit of preemptive living
       donor kidney transplantation: a randomized controlled trial. *Am J Kidney Dis.*
       2013;61(3):476-486.

38.   Arriola KR, Powell CL, Thompson NJ, Perryman JP, Basu M. Living donor transplant
       education for African American patients with end-stage renal disease. *Prog Transplant.*
       2014;24(4):362-370.

39.   Patzer RE, Basu M, Larsen CP, et al. iChoose Kidney: A Clinical Decision Aid for
       Kidney Transplantation Versus Dialysis Treatment. *Transplantation.* 2015.

40.   Hertz RP, Unger AN, Cornell JA, Saunders E. Racial disparities in hypertension
       prevalence, awareness, and management. *Arch Intern Med.* 2005;165(18):2098-2104.

41.   United States Renal Data System. 2014 annual data report: An overview of the
       epidemiology of kidney disease in the United States. 2014; http://www.usrds.org/.

42.   Hall EC, James NT, Grams ME, et al. Center Level Factors and Racial Disparities in Deceased Donor Kidney Transplantation. *Am J Transplant.* 2012;12:62-62.

43.   Tell GS, Mittelmark MB, Hylander B, Shumaker SA, Russell G, Burkart JM. Social support and health-related quality of life in black and white dialysis patients. *ANNA J.* 1995;22(3):301-308; discussion 309-310.

44.   Lopes AA, Bragg-Gresham JL, Satayathum S, et al. Health-related quality of life and associated outcomes among hemodialysis patients of different ethnicities in the United States: The Dialysis Outcomes and Practice Patterns Study (DOPPS). *Am J Kid Dis.* 2003;41(3):605-615.

45.   AHRQ. What is Patient Experience? 2016; https://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html. Accessed February 11, 2017.

46.   National Quality Forum. Patient-Reported Outcomes. 2017; http://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Patient-Reported_Outcomes.aspx. Accessed February 19th, 2017.

47.   Centers for Medicare & Medicaid Services. Medicare and Medicaid Programs; Conditions for Coverage for End-Stage Renal Disease Facilities; Final Rule. In: Department of Health and Human Services, ed. Vol 73. Baltimore, MD: Federal Register; 2008.

48.   Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot.* 1997;12(1):38-48.

49.   Prochaska JO. Decision making in the transtheoretical model of behavior change. *Med Decis Making.* 2008;28(6):845-849.

50.   Bandura A. *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall; 1977.

51. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice Hall; 1986.

52. Bandura A. *Self-Efficacy: The Exercise of Control.* New York: W. H. Freeman; 1997.

53. Janz NK, Becker MH. The Health Belief Model: a decade later. *Health Educ Q.* 1984;11(1):1-47.

54. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav.* 1995;36(1):1-10.

55. Bentler PM, Speckart G. Models of Attitude-Behavior Relations. *Psychol Rev.* 1979;86(5):452-464.

56. Nissenson AR. Improving outcomes for ESRD patients: shifting the quality paradigm. *Clin J Am Soc Nephrol.* 2014;9(2):430-434.

57. Burnier M, Pruijm M, Wuerzner G, Santschi V. Drug adherence in chronic kidney diseases and dialysis. *Nephrol Dial Transplant.* 2015;30(1):39-44.

58. Clark S, Farrington K, Chilcot J. Nonadherence in dialysis patients: prevalence, measurement, outcome, and psychological determinants. *Semin Dial.* 2014;27(1):42-49.

59. Denhaerynck K, Manhaeve D, Dobbels F, Garzoni D, Nolte C, De Geest S. Prevalence and consequences of nonadherence to hemodialysis regimens. *Am J Crit Care.* 2007;16(3):222-235; quiz 236.

60. Agency for Healthcare and Research Quality. CAHPS: Surveys and Tools to Advance Patient-Centered Care. 2016; https://www.ahrq.gov/cahps/index.html. Accessed January 22, 2017.

61. Finkelstein FO. Performance measures in dialysis facilities: what is the goal? *Clin J Am Soc Nephrol.* 2015;10(1):156-158.

62.   Finkelstein FO, Arsenault KL, Taveras A, Awuah K, Finkelstein SH. Assessing and improving the health-related quality of life of patients with ESRD. *Nat Rev Nephrol.* 2012;8(12):718-724.

63.   Sledge R. KDQOL-36 and the interdisciplinary team. *Nephrol News Issues.* 2010;24(7):36-38.

64.   Rubin HR, Fink NE, Plantinga LC, Sadler JH, Kliger AS, Powe NR. Patient ratings of dialysis care with peritoneal dialysis vs hemodialysis. *JAMA.* 2004;291(6):697-703.

65.   Rubin HR, Jenckes M, Fink NE, et al. Patient's view of dialysis care: development of a taxonomy and rating of importance of different aspects of care. CHOICE study. Choices for Healthy Outcomes in Caring for ESRD. *Am J Kidney Dis.* 1997;30(6):793-801.

66.   Kutner NG, Zhang R, McClellan WM. Patient-reported quality of life early in dialysis treatment: effects associated with usual exercise activity. *Nephrol Nurs J.* 2000;27(4):357-367; discussion 368, 424.

67.   Alexander GC, Sehgal AR. Dialysis patient ratings of the quality of medical care. *Am J Kid Dis.* 1998;32(2):284-289.

68.   Hays R, Arnold S. Patient and Family Satisfaction with Care for the Terminally Ill. In: Cassileth BR, Turk DC, Dush DM, eds. *Psychosocial Assessment in Terminal Care*. New York: The Hawthorn Press; 1986.

69.   Glanz K, Rimer B, Viswanath K. *Health Behavior and Health Education: Theory, Research, and Practice.* 4th ed. San Francisco, CA: Jossey-Bass; 2008.

70.   Prochaska J, Redding C, Evers K. The transtheoretical model and stages of change. In: Glanz K, Rimer B, Lewis F, eds. *Health Behavior and Health Education*. 3 ed. San Francisco, CA: Jossey-Bass; 2002:99-120.

71. Prochaska JO, Velicer WF, Rossi JS, et al. Stages of change and decisional balance for 12 problem behaviors. *Health Psychol.* 1994;13(1):39-46.

72. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Public Health.* 2010;31:399-418.

73. Waterman A, Robbins M, Paiva A, Hyland SS. Kidney patients' intention to receive a deceased donor transplant: Development of stage of change, decisional balance, and self-efficacy measures *J Health Psychol.* 2010;15(3):436-445.

74. Waterman AD, Robbins ML, Paiva AL, et al. Measuring kidney patients' motivation to pursue living donor kidney transplant: development of stage of change, decisional balance and self-efficacy measures. *J Health Psychol.* 2015;20(2):210-221.

75. Liu H, Miller LG, Hays RD, et al. Repeated measures longitudinal analyses of HIV virologic response as a function of percent adherence, dose timing, genotypic sensitivity, and other factors. *J Acquir Immune Defic Syndr.* 2006;41(3):315-322.

76. Liu H, Miller LG, Golin CE, et al. Repeated measures analyses of dose timing of antiretroviral medication and its relationship to HIV virologic outcomes. *Stat Med.* 2007;26(5):991-1007.

77. Tangkiatkumjai M, Walker D-M, Praditpornsilpa K, Boardman H. Association between medication adherence and clinical outcomes in patients with chronic kidney disease: a prospective cohort study. *Clin Exp Nephrol.* 2016:1-9.

78. Neri L, Martini A, Andreucci VE, et al. Regimen complexity and prescription adherence in dialysis patients. *Am J Nephrol.* 2011;34(1):71-76.

79.   Ghimire S, Castelino RL, Lioufas NM, Peterson GM, Zaidi STR. Nonadherence to Medication Therapy in Haemodialysis Patients: A Systematic Review. *PLOS ONE.* 2015;10(12):e0144119.

80.   Griva K, Lai AY, Lim HA, Yu Z, Foo MW, Newman SP. Non-adherence in patients on peritoneal dialysis: a systematic review. *PLoS One.* 2014;9(2):e89001.

81.   Butler JA, Peveler RC, Roderick P, Smith PW, Horne R, Mason JC. Modifiable risk factors for non-adherence to immunosuppressants in renal transplant recipients: a cross-sectional study. *Nephrol Dial Transplant.* 2004;19(12):3144-3149.

82.   Butler JA, Roderick P, Mullee M, Mason JC, Peveler RC. Frequency and impact of nonadherence to immunosuppressants after renal transplantation: a systematic review. *Transplantation.* 2004;77(5):769-776.

83.   Hardinger KL, Hutcherson T, Preston D, Murillo D. Influence of Pill Burden and Drug Cost on Renal Function After Transplantation. *Pharmacother.* 2012;32(5):427-432.

84.   Lehane E, McCarthy G. An examination of the intentional and unintentional aspects of medication non-adherence in patients diagnosed with hypertension. *J Clin Nurs.* 2007;16(4):698-706.

85.   Lehane E, McCarthy G. Intentional and unintentional medication non-adherence: a comprehensive framework for clinical research and practice? A discussion paper. *Int J Nurs Stud.* 2007;44(8):1468-1477.

86.   Becker MH. The Health Belief Model and Personal Health Behavior. *Health Educ Monogr.* 1974;2:324–473.

87.   Fishbein M, Ajzen I. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading, MA: Addison-Wesley; 1975.

88. Ajzen I, Fishbein M. *Understanding Attitudes and Predicting Social Behavior.* Englewood Cliffs, NJ: Prentice Hall; 1980.

89. Hays R. An integrated value-expectancy theory of alcohol and other drug use. *Br J Addict.* 1985;80(4):379-384.

90. Yu ZL, Lee VY, Kang AW, et al. Rates of Intentional and Unintentional Nonadherence to Peritoneal Dialysis Regimes and Associated Factors. *PLoS One.* 2016;11(2):e0149784.

91. Jordan S. Considering quality of life following kidney transplant. 2005;7(4):189-190.

92. Van Camp YPM, Vrijens B, Abraham I, Van Rompaey B, Elseviers MM. Adherence to phosphate binders in hemodialysis patients: prevalence and determinants. *J Nephrol.* 2014;27(6):673-679.

93. Wang S, Anum EA, Ramakrishnan K, Alfieri T, Braunhofer P, Newsome B. Reasons for Phosphate Binder Discontinuation Vary by Binder Type. *J Ren Nutr.*24(2):105-109.

94. DiMatteo MR, Hays RD, Gritz ER, et al. Patient Adherence to Cancer Control Regimens: Scale Development and Initial Validation. *Psychol Assess.* 1993;5(1):102-112.

95. Prochaska JM, Prochaska JO, Johnson SS. Assessing Readiness for Adherence to Treatment. In: O'Donohue WT, Levensky ER, eds. *Promoting Treatment Adherence: A Practical Handbook for Health Care Providers*. Thousand Oaks, CA: Sage; 2006:35-46.

96. Johnson SS, Driskell MM, Johnson JL, et al. Transtheoretical model intervention for adherence to lipid-lowering drugs. *Dis Manag.* 2006;9(2):102-114.

97. Evers KE, Prochaska JO, Johnson JL, Mauriello LM, Padula JA, Prochaska JM. A randomized clinical trial of a population- and transtheoretical model-based stress-management intervention. *Health Psychol.* 2006;25(4):521-529.

98.	Martins MT, Silva LF, Kraychete A, et al. Potentially modifiable factors associated with non-adherence to phosphate binder use in patients on hemodialysis. *BMC Nephrol.* 2013;14:208.

99.	Cukor D, Rosenthal DS, Jindal RM, Brown CD, Kimmel PL. Depression is an important contributor to low medication adherence in hemodialyzed patients and transplant recipients. *Kidney Int.* 2009;75(11):1223-1229.

100.	Krousel-Wood M, Islam T, Muntner P, et al. Association of depression with antihypertensive medication adherence in older adults: cross-sectional and longitudinal findings from CoSMO. *Ann Behav Med.* 2010;40(3):248-257.

101.	Pearman T, Yanez B, Peipert J, Wortman K, Beaumont J, Cella D. Ambulatory cancer and US general population reference values and cutoff scores for the functional assessment of cancer therapy. *Cancer.* 2014;120(18):2902-2909.

102.	Peipert JD, Beaumont JL, Bode R, Cella D, Garcia SF, Hahn EA. Development and validation of the functional assessment of chronic illness therapy treatment satisfaction (FACIT TS) measures. *Qual Life Res.* 2014;23(3):815-824.

103.	Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol.* 2005;23.

104.	Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med.* 1993;118(8):622-629.

105.	Cella DF, Tulsky DS. Quality of life in cancer: definition, purpose, and method of measurement. *Cancer Invest.* 1993;11(3):327-336.

106.	Cella DF. Quality of life: concepts and definition. *J Pain Symptom Manage.* 1994;9(3):186-192.

107. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther.* 2014;36(5):648-662.

108. Nunnally JC. *Psychometric theory.* 2nd ed. New York: McGraw-Hill; 1978.

109. Dato NM de Gruijter, Leo J Th van der Kamp. *Statistical Test Theory for the Behavioral Sciences.* Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group; 2008.

110. Rodrigue JR, Cornell DL, Lin JK, Kaplan B, Howard RJ. Increasing Live Donor Kidney Transplantation: A Randomized Controlled Trial of a Home-Based Educational Intervention. *Am J Transplant.* 2007;7(2):394-401.

111. Rodrigue JR, Pavlakis M, Egbuna O, Paek M, Waterman AD, Mandelbrot DA. The "House Calls" trial: a randomized controlled trial to reduce racial disparities in live donor kidney transplantation: rationale and design. *Contemp Clin Trials.* 2012;33(4):811-818.

112. Boulware LE, Hill-Briggs F, Kraus ES, et al. Protocol of a Randomized Controlled Trial of Culturally Sensitive Interventions to Improve African Americans' and Non-African Americans' Early, Shared, and Informed Consideration of Live Kidney Transplantation: The talking about Live Kidney Donation (TALK) study. *BMC Nephrol.* 2011;12(1):34.

113. Pradel FG, Suwannaprom P, Mullins CD, Sadler J, Bartlett ST. Haemodialysis patients' readiness to pursue live donor kidney transplantation. *Nephrol Dial Transplant.* 2009;24(4):1298-1305.

114. Rodrigue JR, Cornell DL, Kaplan B, Howard RJ. Patients' willingness to talk to others about living kidney donation. *Prog Transplant.* 2008;18(1):25-31.

115. Bunn H, O'Connor A. Validation of client decision-making instruments in the context of psychiatry. *Can J Nurs Res.* 1996;28(3):13-27.

116. Cranney A, O'Connor AM, Jacobsen MJ, et al. Development and pilot testing of a decision aid for postmenopausal women with osteoporosis. *Patient Educ Couns.* 2002;47(3):245-255.

117. Ismail SY, Timmerman L, Timman R, et al. A psychometric analysis of the Rotterdam Renal Replacement Knowledge-Test (R3K-T) using item response theory. *Transpl Int.* 2013;26(12):1164-1172.

118. Wright JA, Wallston KA, Elasy TA, Ikizler TA, Cavanaugh KL. Development and results of a kidney disease knowledge survey given to patients with CKD. *Am J Kidney Dis.* 2011;57(3):387-395.

119. Waterman AD, McSorley AM, Peipert JD, et al. Explore Transplant at Home: a randomized control trial of an educational intervention to increase transplant knowledge for Black and White socioeconomically disadvantaged dialysis patients. *BMC Nephrol.* 2015;16:150.

120. Waterman AD, Robbins ML, Paiva AL, et al. Your Path to Transplant: a randomized controlled trial of a tailored computer education intervention to increase living donor kidney transplant. *BMC Nephrol.* 2014;15:166.

121. Hall KL, Robbins ML, Paiva A, Knott JE, Harris L, Mattice B. Donation intentions among African American college students: decisional balance and self-efficacy measures. *J Behav Med.* 2007;30(6):483-495.

122. Burditt C, Robbins ML, Paiva A, Velicer WF, Koblin B, Kessler D. Motivation for blood donation among African Americans: developing measures for stage of change, decisional balance, and self-efficacy constructs. *J Behav Med.* 2009;32(5):429-442.

123. Brick LA, Sorensen D, Robbins ML, Paiva AL, Peipert JD, Waterman AD. Invariance of measures to understand decision-making for pursuing living donor kidney transplant. *J Health Psychol.* 2015.

124. Svarstad BL, Chewning BA, Sleath BL, Claesson C. The Brief Medication Questionnaire: a tool for screening patient adherence and barriers to adherence. *Patient Educ Couns.* 1999;37(2):113-124.

125. Cohen JL, Mann DM, Wisnivesky JP, Leventhal H, Musumeci TJ, Halm EA. Assessing the validity of self-reported medication adherence among inner city asthmatic adults. *J Gen Intern Med.* 2008;23:251-251.

126. Horne R, Buick D, Fisher M, Leake H, Cooper V, Weinman J. Doubts about necessity and concerns about adverse effects: identifying the types of beliefs that are associated with non-adherence to HAART. *Int J Std Aids.* 2004;15(1):38-44.

127. Horne R, Parham R, Driscoll R, Robinson A. Patients' attitudes to medicines and adherence to maintenance treatment in inflammatory bowel disease. *Inflamm Bowel Dis.* 2009;15(6):837-844.

128. George J, Kong DC, Thoman R, Stewart K. Factors associated with medication nonadherence in patients with COPD. *Chest.* 2005;128(5):3198-3204.

129. Home R. *The nature, determinants and effects of medication beliefs in chronic illness.* London, England: Faculty of Medicine, University of London; 1997.

130. Chater AM, Parham R, Riley S, Hutchison AJ, Horne R. Profiling patient attitudes to phosphate binding medication: a route to personalising treatment and adherence support. *Psychol Health.* 2014;29(12):1407-1420.

131. Theofilou P. Medication adherence in Greek hemodialysis patients: the contribution of depression and health cognitions. *Int J Behav Med.* 2013;20(2):311-318.

132. Wileman V, Farrington K, Wellsted D, Almond M, Davenport A, Chilcot J. Medication beliefs are associated with phosphate binder non-adherence in hyperphosphatemic haemodialysis patients. *Br J Health Psychol.* 2015;20(3):563-578.

133. Knobel H, Alonso J, Casado JL, et al. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. *AIDS.* 2002;16(4):605-613.

134. Arenas MD, Malek T, Gil MT, Moledous A, Alvarez-Ude F, Reig-Ferrer A. Challenge of phosphorus control in hemodialysis patients: a problem of adherence? *J Nephrol.* 2010;23(5):525-534.

135. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care.* 1986;24(1):67-74.

136. Toll BA, McKee SA, Martin DJ, Jatlow P, O'Malley SS. Factor Structure and Validity of the Medication Adherence Questionnaire (MAQ) with Cigarette Smokers Trying to Quit. *Nicotine Tob Res.* 2007;9(5):597-605.

137. Garcia-Llana H, Remor E, Selgas R. Adherence to treatment, emotional state and quality of life in patients with end-stage renal disease undergoing dialysis. *Psicothema.* 2013;25(1):79-86.

138. Wileman V, Chilcot J, Norton S, Hughes L, Wellsted D, Farrington K. Choosing not to take phosphate binders: the role of dialysis patients' medication beliefs. *Nephron Clin Pract.* 2011;119(3):c205-213.

139. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens (Greenwich).* 2008;10(5):348-354.

140. Neipp M, Karavul B, Jackobs S, et al. Quality of life in adult transplant recipients more than 15 years after kidney transplantation. *Transplantation.* 2006;81(12):1640-1644.

141. Valderrábano F, Jofre R, López-Gómez JM. Quality of life in end-stage renal disease patients. *Am J Kid Dis.* 2001;38(3):443-464.

142. Simmons RG, Anderson C, Kamstra L. Comparison of quality of life of patients on continuous ambulatory peritoneal dialysis, hemodialysis, and after transplantation. *Am J Kidney Dis.* 1984;4(3):253-255.

143. Evans RW, Manninen DL, Garrison LPJ, et al. The Quality of Life of Patients with End-Stage Renal Disease. *N Eng J Med.* 1985;312(9):553-559.

144. Butt Z, Yount SE, Caicedo JC, Abecassis MM, Cella D. Quality of life assessment in renal transplant: review and future directions. *Clin Transplant.* 2008;22(3):292-303.

145. Selewski DT, Massengill SF, Troost JP, et al. Gaining the Patient Reported Outcomes Measurement Information System (PROMIS) perspective in chronic kidney disease: a Midwest Pediatric Nephrology Consortium study. *Pediatr Nephrol.* 2014;29(12):2347-2356.

146. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol.* 1998;51(11):903-912.

147. Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol.* 1998;51(11):1171-1178.

148. United States Renal Data System. *2016 USRDS annual data report: Epidemiology of kidney disease in the United States.* Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases;2016.

149. Morisky DE, Levine DM, Green LW, Smith CR. Health education program effects on the management of hypertension in the elderly. *Arch Intern Med.* 1982;142(10):1835-1838.

150. Krousel-Wood M, Islam T, Webber LS, Re RN, Morisky DE, Muntner P. New medication adherence scale versus pharmacy fill rates in seniors with hypertension. *Am J Manag Care.* 2009;15(1):59-66.

151. Krousel-Wood M, Thomas S, Muntner P, Morisky D. Medication adherence: a key factor in achieving blood pressure control and good clinical outcomes in hypertensive patients. *Curr Opin Cardiol.* 2004;19(4):357-362.

152. Krousel-Wood MA, Muntner P, Islam T, Morisky DE, Webber LS. Barriers to and determinants of medication adherence in hypertension management: perspective of the cohort study of medication adherence among older adults. *Med Clin North Am.* 2009;93(3):753-769.

153. Wong MCS, Wu CHM, Wang HHX, et al. Association Between the 8-Item Morisky Medication Adherence Scale (MMAS-8) Score and Glycaemic Control Among Chinese Diabetes Patients. *J Clin Pharmacol.* 2015;55(3):279–287.

154. De las Cuevas C, Peñate W. Psychometric properties of the eight-item Morisky Medication Adherence Scale (MMAS-8) in a psychiatric outpatient setting. *Int J Clin Health Psychol.* 2015;15(2):121-129.

155. Alkatheri AM, Alyousif SM, Alshabanah N, et al. Medication adherence among adult patients on hemodialysis. *Saudi J Kidney Dis Transpl.* 2014;25(4):762-768.

156.	Morisky DE, Malotte CK, Choi P, et al. A patient education program to improve adherence rates with antituberculosis drug regimens. *Health Educ Q.* 1990;17(3):253-267.

157.	Morisky DE, DiMatteo MR. Improving the measurement of self-reported medication nonadherence: response to authors. *J Clin Epidemiol.* 2011;64(3):255-257; discussion 258-263.

158.	Joson CG, Henry SL, Kim S, et al. Patient-Reported Factors Associated With Poor Phosphorus Control in a Maintenance Hemodialysis Population. *J Ren Nutr.* 2016;26(3):141-148.

159.	Alkatheri AA, Albekairy AM, Jarab A, et al. Medication Adherence and Treatment Satisfaction Among Renal Transplant Recipients. *Ann Transplant.* 2016;21:270-278.

160.	Ricardo AC, Hacker E, Lora CM, et al. Validation of the Kidney Disease Quality of Life Short Form 36 (KDQOL-36) US Spanish and English versions in a cohort of Hispanics with chronic kidney disease. *Ethn Dis.* 2013;23(2):202-209.

161.	Chen JY, Choi EP, Wan EY, et al. Validation of the Disease-Specific Components of the Kidney Disease Quality of Life-36 (KDQOL-36) in Chinese Patients Undergoing Maintenance Dialysis. *PLoS One.* 2016;11(5):e0155188.

162.	Bentler PM. *EQS 6 Structural Equations Program Manual.* Encino, CA: Multivariate Software, Inc.; 2006.

163.	Hu LT, Bentler PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Struct Equ Modeling.* 1999;6(1):1-55.

164.   Chao S, Yen M, Lin T-C, Sung J-M, Wang M-C, Hung S-Y. Psychometric Properties of the Kidney Disease Quality of Life–36 Questionnaire (KDQOL-36™). *West J Nurs Res.* 2016;38(8):1067-1082.

165.   Yang F, Griva K, Lau T, et al. Health-related quality of life of Asian patients with end-stage renal disease (ESRD) in Singapore. *Qual Life Res.* 2015;24.

166.   Mateti UV, Nagappa AN, Attur RP, Nagaraju SP, Mayya SS, Balkrishnan R. Cross-Cultural Adaptation, Validation and Reliability of the South Indian (Kannada) Version of the Kidney Disease and Quality of Life (KDQOL-36) Instrument. *Saudi J Kidney Dis T.* 2015;26(6):1246-1252.

167.   Chow SK, Tam BM. Is the kidney disease quality of life-36 (KDQOL-36) a valid instrument for Chinese dialysis patients? *BMC Nephrol.* 2014;15:199.

168.   Tao X, Chow SK, Wong FK. Determining the validity and reliability of the Chinese version of the Kidney Disease Quality of Life Questionnaire (KDQOL-36). *BMC Nephrol.* 2014;15:115.

169.   Bayoumi M, Al Harbi A, Al Suwaida A, Al Ghonaim M, Al Wakeel J, Mishkiry A. Predictors of quality of life in hemodialysis patients. *Saudi J Kidney Dis Transpl.* 2013;24(2):254-259.

170.   Mapes DL, Lopes AA, Satayathum S, et al. Health-related quality of life as a predictor of mortality and hospitalization: the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Kidney Int.* 2003;64(1):339-349.

171.   Mapes DL, Bragg-Gresham JL, Bommer J, et al. Health-related quality of life in the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Am J Kidney Dis.* 2004;44(5 Suppl 2):54-60.

172. Fukuhara S, Lopes AA, Bragg-Gresham JL, et al. Health-related quality of life among dialysis patients on three continents: the Dialysis Outcomes and Practice Patterns Study. *Kidney Int.* 2003;64(5):1903-1910.

173. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care.* 2004;42(9):851-859.

174. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ.* 2004;13(9):873-884.

175. Medical Education Institute. KDQOL Complete. 2016; https://www.kdqol-complete.org/. Accessed December 28th, 2016.

176. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press; 1988.

177. Bentler PM, Wu EJC. *EQS 6 for Windows User's Guide* Encino, CA: Multivariate Software, Inc.; 2002.

178. Muthén LK, Muthén BO. *Mplus User's Guide.* Los Angeles, CA1998-2015.

179. Barendse MT, Oort FJ, Timmerman ME. Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Struct Equ Modeling.* 2015;22(1):87-101.

180. Hays RD, Hayashi T. Beyond internal consistency reliability; Rationale and User's Guide for Multirait Scaling Analysis Program on the microcomputer. *Behav Res Methods Instrum Comput.* 1990;22(167-175).

181. Hays RD, Wang EYI. Multitrait Scaling Program. SAS User's Group: 17th Annual Conference April 12-15, 1992; Honolulu, HI

182. Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika.* 1951;16(3):297-334.

183. Hays R, Revicki D. Reliability and validity (including responsiveness). In: Fayers P, Hays R, eds. *Assessing Quality of Life in Clinical Trials: Methods and Practice*. 2nd ed. Oxford, NY: Oxford University Press; 2005:525-539.

184. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Modeling.* 2002;9(2):233-255.

185. Yoon M, Millsap RE. Detecting violations of factorial invariance using data-based specification searches: a Monte Carlo study. *Struct Equ Modeling.* 2007;14(3):435-463.

186. de Winter JCF, Dodou D. Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *J Appl Stat.* 2012;39(4):695-710.

187. Hays RD, Spritzer K. *KDQOL-36™ Scoring Program (v1.0).* Santa Monica, CA: RAND;2000.

188. Waterman AD, Stanley SL, Covelli T, Hazel E, Hong BA, Brennan DC. Living donation decision making: recipients' concerns and educational needs. *Prog Transplant.* 2006;16(1):17-23.

189. Salter ML, Gupta N, King E, et al. Health-related and psychosocial concerns about transplantation among patients initiating dialysis. *Clin J Am Soc Nephrol.* 2014;9(11):1940-1948.

190. Reise SP, Kim DS, Mansolf M, Widaman KF. Is the Bifactor Model a Better Model or Is It Just Better at Modeling Implausible Responses? Application of Iteratively Reweighted

Least Squares to the Rosenberg Self-Esteem Scale. *Multivar Behav Res.* 2016;51(6):818-838.

191.   Muntner P, Joyce C, Holt E, et al. Defining the minimal detectable change in scores on the eight-item Morisky Medication Adherence Scale. *Ann Pharmacother.* 2011;45(5):569-575.

192.   Young EW, Goodkin DA, Mapes DL, et al. The Dialysis Outcomes and Practice Patterns Study (DOPPS): An international hemodialysis study. *Kidney Int.*57:S74-S81.

193.   Yang F, Lau T, Lee E, Vathsala A, Chia KS, Luo N. Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). *Eur J Health Econ.* 2015;16(9):1019-1026.

194.   Bharmal M, Thomas J. Comparing the EQ-5D and the SF-6D Descriptive Systems to Assess Their Ceiling Effects in the US General Population. *Value Health.* 2006;9(4):262-271.

195.   Reeve BB, Fayers P. Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers P, Hays RD, eds. *Assessing Quality of Life in Clinical Trials.* Vol 2. New York: Oxford University Press; 2005:53-73.

196.   Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods.* 2000;5(3):343-355.

197.   Feldt LS, Brennan RL. Reliability. In: Linn RL, ed. *Educational Measurement.* Vol 3. Phoenix, AZ: The Oryx Press; 1989.

198.   Bentler PM. Specificity-Enhanced Reliability Coefficients. *Psychol Methods.* 2016.

199.   Peipert JD, Hays RD. Methodological Considerations in Using Patient Reported Measures in Dialysis Clinics. *Journal of Patient-Reported Outcomes.* Under Review.

200. Liu H, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemol.* 2010;63(11):1169-1178.