Comprehensive Analysis of End-stage Renal Disease Patient Population with a Co-morbidity of Cardiovascular Disease to Evaluate Discharge Outcomes when Treated with Beta-blockers.

by Amanda Bell

B.A. in Business, May 2008, Strayer University
A.S. in Biotechnology, May 2013, Northern Virginia Community College
B.S. Integrated Science and Technology, May 2015, The George Washington University

A Thesis submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Master of Science

August 31, 2017

Thesis directed by

Raja Mazumder
Associate Professor of Biochemistry and Molecular Biology

Dedication

The author wishes to dedicate this page to John Koons Jr. for his constant encouragement in her academic endeavors and support. His passion for the pursuit of knowledge is the true force that drives those whom he engages with and keeps them on the continual path of pursuing a greater and smarter way of doing things. Thank you, John, for being a true patron for the sciences and the best grandad any child can ask for.

Acknowledgements

The author wishes to thank her mentor, Dr. Raja Mazumder for the commitment and guidance on the USRDS project. Thank you to Dr. Vahan Simonyan for reading and contributing to the project. To the program director, Dr. Jack Vanderhoek, thank you for your invaluable advice and guidance through the master's program. To Dr. Nee and Dr. Raj for consulting from a clinical perspective and advising on how to approach USRDS data. This would have been an impossible task without the help and input from Dr. Atin Basuchoudhary and Dr. Bang on how to apply statistical analyses to this data. A special thanks to Dr. Shant Ayanian who fulfilled the dual role from a clinical informatics and machine learning perspective on this project. The author would like to acknowledge Mr. Ting-Chia (T.C.) Chang and Ms. Yu Hu for the constant support, encouragement and data conversions for this project. Thank you to Dr. Robel Kahsay for giving input and suggestions on how to convert and normalize the data. Finally, the author would like to thank Ms. Krista Smith for playing such an integral role in the USRDS project, assisting with IRBs, providing a space for all collaborators to discuss the project and writing proposals.

Abstract of Thesis

Comprehensive Analysis of End-stage Renal Disease Patient Population with a Co-morbidity of Cardiovascular Disease when Treated with Beta-blockers

Medicare costs for ESRD patients in the United States costs over $30 billion a year. The leading cause of death in ESRD patients is cardiovascular disease which worsens both short and long- term in this patient population. These costs include hospitalization events and prescription drug claims. The majority of the cost resides from inpatient and outpatient hospital events and where the leading category of death is from Arrhythmia/Cardiac arrest at approximately 40% mortality. As a whole in ESRD patient population 53.6% of deaths are from cardiovascular episodes. These cardiovascular episodes are treated with beta-blockers. Previous research describes how beta- blockers can be used as a protective measure for ESRD patients who suffer from cardiovascular events. We propose to isolate an ESRD Medicare Part D patient population hospitalized for cardiovascular events as a primary diagnosis. We used IBM spss software to evaluate the outcomes of the cardiovascular patient population when they were administered beta-blockers. The patients who suffered from congestive heart failure and were treated with metoprolol succinate had a slight advantage than those who were treated with atenolol where they were discharged to go home more frequently than transferred to an extended stay in the hospital. We have also observed in the aging African American male population that they had better discharge outcomes for atenolol. More research has to be conducted in order to evaluate what factors contribute to these findings.

Table of Contents

Dedication	iii
Acknowledgements	iv
Abstract of Thesis	v
List of Figures	vii
List of Tables	viii
List of Symbols / Nomenclature	ix
Introduction	1
Results and Discussion	5
Materials and methods	9
Conclusion	11
List of Supplementary Materials	18
References	23

List of Figures

Figure 1 Distribution of beta-blockers by ethnicity 14

Figure 2 Distribution of beta-blockers by gender 15

Figure 3 Patient discharge codes 16

Figure 4 Patient primary diagnostic codes 17

List of Tables

Table 1 Summary of USRDS data received from MBOX 12

Table 2 Extracted data from the hospital patient population 13

List of Symbols / Nomenclature

1. USRDS – United States Renal Data Systems
2. ESRD – end-stage renal disease
3. CKD – chronic kidney disease
4. CVD – cardiovascular disease
5. ASHD – atherosclerotic heart disease
6. AMI – acute myocardial infarction
7. CHF – congestive heart failure
8. VHD – valvular heart disease
9. CVA/TIA – cerebrovascular accident/transitory ischemic attack
10. PAD – peripheral arterial disease
11. AFIB – arterial fibrillation
12. SCA/VA – sudden cardiac arrest/ventricular arrhythmias
13. SNF – skilled nursing facility
14. ICF – intermediate care facility
15. RLDA – regularized linear discriminant analysis

Introduction

At the end of the first quarter of 2016 a total of 711, 822 individuals were treated for ESRD compared to 688, 530 at the same time in 2015[1]. The leading cause of death in the ESRD patient population is cardiovascular disease where the majority of the cost comes from inpatient and outpatient hospital episodes. Cardiovascular events amount for 53.6% of death with Arrhythmia/Cardiac arrest alone attributes to 38.7% of deaths in this patient population[1]. Major complications in chronic kidney disease patients are cardiovascular events[1]. Despite advances in patient care, the annual mortality, while improving, is still unacceptability high. With the explosion of research in biomedical informatics, a variety of new opportunities have emerged to analyze large clinical and related administrative datasets such as the ones available through USRDS. The existing data files in USRDS contains the metadata of patients which we can use to evaluate the treatment outcomes of generic beta-blockers which have been proven to have effective and lasting patients outcomes[2] . The USRDS conduct Bayesian methods to calculate the mortality ratios of kidney disease[3, 4] where these models are useful to narrow down the specific aspects of disease to research, for instance the prevalence of cardio cardiovascular events being the leading cause of death in the ESRD patient population.

CROWNWeb is a resource developed by the Centers for Medicare and Medicaid Services to provide a mechanism to collect and share dialysis data for ESRD patients[5, 6]. The data collected from the dialysis centers through CROWNWeb can be combined with the rest of the clinical resources from USRDS to evaluate patient outcomes, quality of care statistics and provide the most successful dialysis facilities and programs as it tracks admission of the patient and discharge information[7-9].

Leveraging this type of data allows for a better and cleared understanding of patients with comorbidity and a path to effectively and efficiently treat patients diagnosed with several diseases. Clinical studies can be created where different researchers can leverage these data sources and develop research projects simultaneously by observing the interaction of drugs and different disease and patient vitals[10].

Previous research has demonstrated that cardiovascular disease is associated with the early stages of chronic kidney disease and that antihypertensive medications such as beta-blockers are used to treat these patients[11-13]. Due to high cardiovascular disease (CVD) and hypertension events, end-stage-renal disease (ESRD) patients are prescribed medications with cardio-protective properties[14]. Beta-adrenergic blocking agents (β-blockers) have been proven to have beneficial effects on the long-term outcome after myocardial infarctions[15, 16] patients in the general population.

The patients in the data set which we obtained from USRDS were administered three different beta-blockers; metoprolol succinate and tartrate, atenolol, and carvedilol. The most prescribed beta-blockers are atenolol and metoprolol for all-cause cardiovascular disease and that only 7% of the patients who were administered atenolol were more likely to die from any causes within five years compared to 13.1% who were on metoprolol[17]. Atenolol is excreted primarily by the kidney compared to other beta-blockers, however previous research has shown that there was no difference of effects between atenolol and metoprolol succinate as far as vascular endothelial function[12, 18]. In a cohort study conducted from 1987 – 2011 it was established that patients between the age of 60 and 75 had an increased survival response when treated with either statins or beta-blockers[19]. Previous studies have also demonstrated that there were no increased hospitalization rates whether a patient was treated with atenolol or

metoprolol when eGFR rates were taken into consideration. Patients who were hospitalized and given beta-blockers and were treated with them for a year did not worsen their condition compared to those were not administered beta-blockers[20]. Evaluating other research conducted it was not easy to determine whether the patients who were on beta-blockers were discharged earlier form hospital or whether the favorable outcomes of those patients of with in-hospital cardiac arrest could be attributed to beta-blockers[21, 22].

Our goal is to test that given metoprolol succinate or atenolol for the same cardiovascular event which beta-blocker will provide the best hospital discharge scenario. Patients who suffer from AMIs are prescribed either atenolol or metoprolol and there have been no significant results to show either had a better outcome as far as death or recurring AMIs were concerned[23]. By using USRDS, we can create small cohort studies based on various data submissions and package the data in a manner where it can be available for researchers to obtain and conduct their studies. Creating such small pilot studies where analyses are done and organized in a structured manner provides researchers with an overview of the specific number of patients and the USRDS data files created.

Early stage disease recording and data sharing encourage an interoperable medium where stakeholders can gain access to the data and use it to identify complication areas of the disease[24, 25]. The data that is shared and distributed aids in establishing and interoperable medium where a community of stakeholders collectively solve major challenges and mitigate the progression of the disease compared to other patients[26]. Physicians and researchers can identify pockets where the disease is prevalent and evaluate methods to curb it by communicating the findings of the

research with patients. This mechanism set the stage for evidence based and precision medicine where the optimization of patient outcomes is the goal. Stakeholder communities are enabled to establish protocols and standards on how to treatment practices can be leveraged in order to improve quality of care to the patient while also reducing the burden of cost on the patient and hospital[27, 28]. Due to the falling cost of cutting edge technologies and accessibility to more areas to provide treatment the generation of clinical data has been increased, thus requiring the use of standard ontologies and setting data benchmarks[29-31]. Leveraging the use of open source platforms and sharing spaces such as GitHub allows for broader community participation by making the data available.

Results and Discussion

Normalizing the data

The combined file contained 17, 753 and the patient population that was isolated of the patients who were hospitalized and were on the prescription drug enrollment plan. The medical evidence file[2, 32] contains the metadata of the patient such as the first diagnosis of ESRD, number of dialysis, race, ethnicity, sex, comorbidities, social and geographical features such as smoking, drug abuse, state and region of the state that they are located.

All the patients who were hospitalized with a primary diagnosis of CVD and were administered a beta-blocker were clustered together to form a patient population of 8, 660 and furthered isolated where the patients who had CVD as a primary diagnosis and administered beta-blockers were 3, 946 as indicated Table 2. We further clustered the patients by the type of beta-blocker which they were administered Figure 1 and most of the patients were administered metoprolol over atenolol. The highest recipients for the beta-blockers were males as indicated in Figure 2 and Figure 1 indicates that the largest ethnic group was African American.

Modeling the data

There were eight final cardiovascular events that were used in the evaluation of this patient population were atherosclerotic heart disease (ASHD), acute myocardial infarction (AMI), congestive heart failure (CHF), valvular heart disease (VHD), cerebrovascular accident / transitory ischemic attack (CVA/TIA), peripheral arterial disease (PAD), arterial fibrillation (AFIB), and sudden cardiac arrest/ventricular arrhythmias (SCA/VA). We isolated two beta-blockers of interest; metoprolol

succinate and atenolol and measured their outcomes based on diagnostic (DISCSTAT) code and primary diagnostic codes Figure 3 and Figure4. The statistical analysis was conducted using the three outcomes that the patient could have from the diagnostic code list in Figure 1 which were assigned numerical values 1- discharges to go home, 3 – discharged or transferred to a skilled nursing facility and 6 – discharged to a home care of home organized facility. Outcome 1 was denoted as the best outcome and outcome 3 as the worst as the patient would require extended care.

Regularized Linear Discriminant Analysis

Regularized Linear Discriminant Analysis (RLDA)[33] is an analysis to determine the optimum regularization parameter to obtain an overview of parameters to test. We used RLDA which embedded in HIVE[34]. The parameters that were set to test to see whether there were any high level correlations between variables such as age, discharge code, primary diagnostic code and GFR. RLDA used a Bayesian type of analysis to calculate the correlation. The Bayesian type analysis is useful to determine the probability of a certain drug will influence the most desirable discharge outcome for the patient and be discharged to go home a few days after admittance. The advantages of using this method is that the predictive probability analysis can be updated each time as the data is being modelled and produce a mechanism where decision making can be made directly while observing the variables that are being tested and whether they are directly influence the positive or negative result, where in this study the result will be whether the patient discharge outcome is favorable when administered a beta-blocker or not[35, 36]. There was a strong correlation between age and eGFR and based on the nature of this RLDA algorithm when age and eGFR, however we were not able to effectively test the eGFR in relation to ethnicity and gender because the dataset was not sufficient to for observing the renal function during the experimental procedure. Previous research indicates that many

elderly patients were diagnosed with CKD based solely on the GFR but neglect to monitor the renal function thus prohibiting us to further test this hypothesis due to the lack of data that monitored the renal function systematically[37].

Multiple hypothesis tests for cardiovascular disease

We did a multiple hypothesis test to compare the number of hospitalization dates for atherosclerotic heart disease, acute myocardial infarction, congestive heart failure, valvular heart disease, cerebrovascular accident / transitory ischemic attack, peripheral arterial disease, arterial fibrillation, sudden cardiac arrest. We fail to reject the null hypothesis at the alpha level of 0.05 that atenolol and metoprolol succinate has the same discharge outcomes for patient who had acute myocardial infarction primary diagnosis. The patients had a significant level of .964 that indicates that whether a patient is given either atenolol or metoprolol succinate they will have the same discharge outcomes in general. However, when it came to patients who were hospitalized for all cardiovascular events we rejected the null since the significance level was less than 0.05. This result is consistent with previous work done where treatment with beta-blockers improved all-cause mortality in the CKD patient population[38-40].

Previous research indicates that patients who are treated with beta-blockers after myocardial infarction denoted as primary diagnosis 2 have better outcomes but this could not be verified in this dataset as only 1% of the patients who had AMI were treated with metoprolol and a half of a percentage with atenolol[41]. However, patients with congestive heart failure had better outcomes when treated with beta-blockers where metoprolol has a slight advantage over atenolol with the p-value of 0.001 at the 0.05 significant level which has been indicated in previous research about congestive

heart failure patients[42-44]. Evaluating the discharge codes for only ischemic events we failed to reject the null hypothesis at the 0.05 significance level as the p-value was 0.08 that generic drug atenolol has better diagnostic outcomes or vice versa.

We separated the patients by sex to evaluate whether they would have better outcomes, Figure 2. Based on the sex of the patient, males had better outcomes on beta-blockers than the female population. There was significant evidence that the sex of the patient had a better diagnostic outcome on atenolol than metoprolol. Whether on a beta-blocker or not the outcome for males were better than their female counter parts. When observing the statistics for congestive heart failure, diagnostic code , the female population was diagnosed less with congestive heart failure, which is consistent with studies done by other researchers and thus could lead to the false positive that males do better[45, 46]. Multiple studies indicate that the progression of disease and mortality of the aging patient population worsened, however there was no significant evidence that aging patients had better diagnostic outcomes on atenolol than metoprolol[37, 47, 48]. Although, when age and ethnicity was combined there was a slight increase in the better outcomes category for atenolol in male African American patient population. This is consistent with previous research that was conducted that indicates that African American males had higher diagnosis in cardiovascular disease than other groups and were therefore the highest group in this patient population[43, 48, 49].

Materials and methods

Data Source - USRDS

We obtained the data files from United States Renal Data Systems (USRDS) in sas format. A total of 67 GB of data from USRDS through MBOX hosted at the University of Michigan. The files were packaged in SAS format in compressed zip folders see Table1. Although we have received seven different categories of the files some of the files overlapped and thus producing repeated files in several of the categories except for the CROWNWEB data which contains entries from dialysis facilities and an overall medical patient history and the prescription drug event (PDE) files which contain the prescription drug information that the patient has been on from 2010 – 2013. The Clinical Performance Measures files contain all the information of the dialysis patients' quality of care with the dialysis facility information. The ESRD Claims file contains all the claims of patients who are at the last stage of kidney disease including their hospitalization information, whereas the Medicare Claims Clinical file includes all the clinical data of the patients who are on Medicare Part D which includes hospital, transplant and dialysis information. The Hospital folder contained two hospitals files the first the entire USRDS patient hospital information up to 2009 and the second file contained patient hospitalization from 2010 onwards. The transplant file contains data on patient transplant information, those who are on the waiting list for a transplant and those who have received transplants and their status. The core folder included other supporting data files such as the medical evidence patient files that contain patient information such as gender, first diagnosis of kidney disease and other comorbidities.

Conversion of data

Used python to convert the files from SASBDAT files to CSV see appendix Table 2 for

the sizes and description of the files. The metadata used from these files contained columns variables which ranged from Patient_ID, age, race, ethnicity, etc. see appendix Table 3. The unique patient ID has multiple rows as some of the patients were hospitalized more than once and this caused an issue with the rows repeating. To overcome this, we assigned a unique internal HIVE_ID to each row and this gave each row a unique value to be assessed more accurately.

Building a working dataset

We combined the Hospital2010On file with the prescription drug files from 2010 – 2013 and the 2005 on Medical evident files, Table 2. To create the working dataset we ran a few models to using python scripts to clean out the columns and rows and also a set of ICD-9 codes on GitHub to match the rows. This produced a table that could easily be formatted in various statistical analysis software such as R, IBM spss and SAS. Although the code on GitHub was created to use R programming, the code ran for several days and we thus used a python script that ran for only six hours. We used RLDA to evaluate the most significant variables that correlates with the hypothesis that we wanted to test. IBM spss was used to run all the statistical analysis through Colonial One, a GW server.

Conclusion

We analyzed data obtained from USRDS to evaluate whether there was any significant difference in discharge outcomes in patients who has end-stage renal disease with cardiovascular disease as a primary diagnosis to hospitalization and were administered beta-blockers, atenolol and metoprolol succinate. We evaluated whether atenolol and metoprolol succinate had the same outcomes when administered to the patient population. It was determined that there was no significant difference in discharge outcomes in ESRD patients with acute myocardial infarction whether they were administered atenolol or metoprolol succinate. However, in the patients who were admitted to the hospital with congestive heart failure as a diagnosis experienced slightly better discharge outcomes when on metoprolol succinate.

Males in general had a better outcome in discharge outcomes whether they were on beta-blockers or not. The data did also indicate that there was a significant difference between atenolol and metoprolol succinate. There was no significant evidence in the data that patients had better diagnostic outcomes with age. Combining age and ethnicity showed an increase in significance level where the African American male had better outcomes on atenolol then metoprolol succinate.

Although the data indicates that there are significant information that atenolol can be more beneficial in the ESRD-CVD African American male aging population more research needs to be conducted to evaluate the factors that contribute to these results. In order to examine this data further we are building a model to compare and match all the patient comorbidities and account for age in increments of ten years versus fifteen years done in this analysis.

Tables

Table 1 – Raw USRDS data files downloaded from MBOX.
The raw data files obtained from USRDS were compressed into smaller files. The files were compressed into one or two files but when opened there were more sub files that we converted from sas to csv.

Table 1: Raw Data Summary from USRDS			
Type of files/data	**Format of**	**Size of each individual**	**Number of**
Clinical Performance	SAS	Min 800 MB-Max10 GB	2
CROWNWeb	SAS	3.5 GB	1
ESRD Claims	SAS	1 GB – 5.3 GB	46
2015 Medicare Claims_Clinical	SAS	Min 4.5 GB – Max 1.8 GB	1
2015 Hospital	SAS	Min 2.7 GB – Max 15.4GB	1
2015 Transplant	SAS	Min 381 KB – Max 9.2GB	1
2015 Core	SAS	Min 166 KB – Max 2.5 GB	1

Table 2 – The number of patients isolated from the data files in USRDS after converted.
The tabular data sets were used to isolate the desired pieces of data in order to determine whether the data was sufficient to carry on with the experiment. We started with three separate hospital, prescription drug and medical evidence patient files and extracted all the cardiovascular disease primary diagnostic codes to create a combined file which contained data from the three separated files. We then extracted the patients who were administered beta-blockers from the file and then extracted the different beta-blockers from the beta-blocker file.

Table 2: Comprehensive files	
File Name	Number of patients
Hospital 2010 from 2010 - 2013	714, 889
Prescription Drug Enrollment 2010 - 2013	638, 940
Medical Evidence	100, 954
Combined File	
Created hospital, Med evidence	17, 753
Beta Blocker file	
Number of patients	8, 660
Cardiovascular disease and beta-blocker file	
Cardiovascular disease beta-blocker	3, 311
Metoprolol Succinate	2, 342
Carvedilol	7
Atenolol	406
Metoprolol Tartrate	1,191

Figures

Figure 1 - Distribution of beta-blockers by ethnicity
The figure indicates a list of specific beta-blockers versus the ethnicity of the patient. The x-axis indicates the ethnicity of the patient. The y-axis indicates the counts of each of the beta-blockers administered to patients.

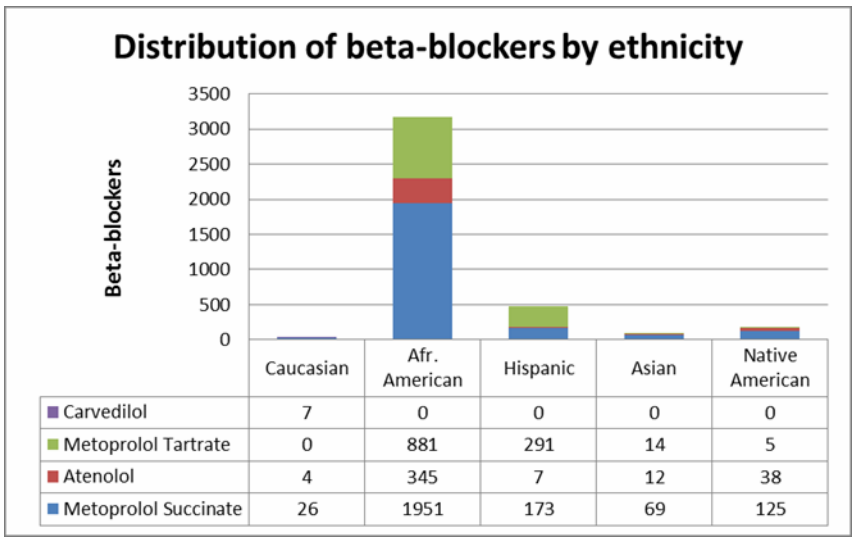


Figure 2 – Distribution of beta-blockers by gender
This figure denotes the number of beta-blockers administered to males and females. The x-axis indicates the beta-blockers administered. The y-axis indicates the counts of males and females who received the beta-blockers.

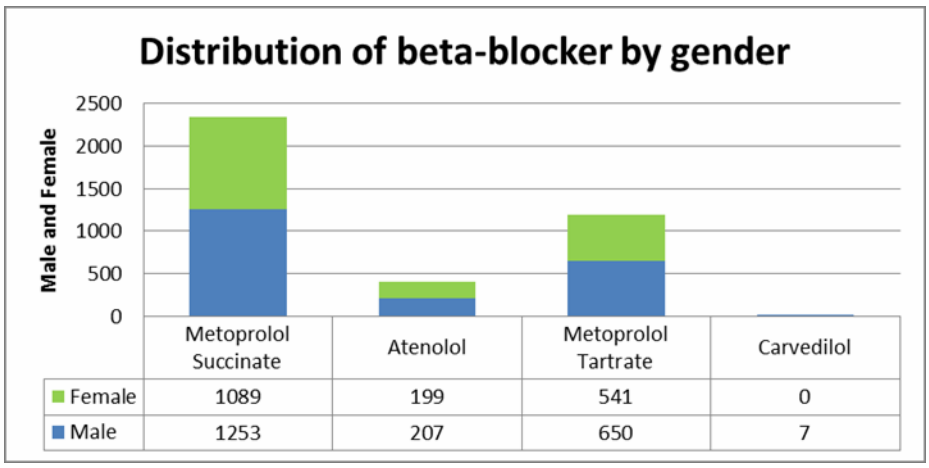
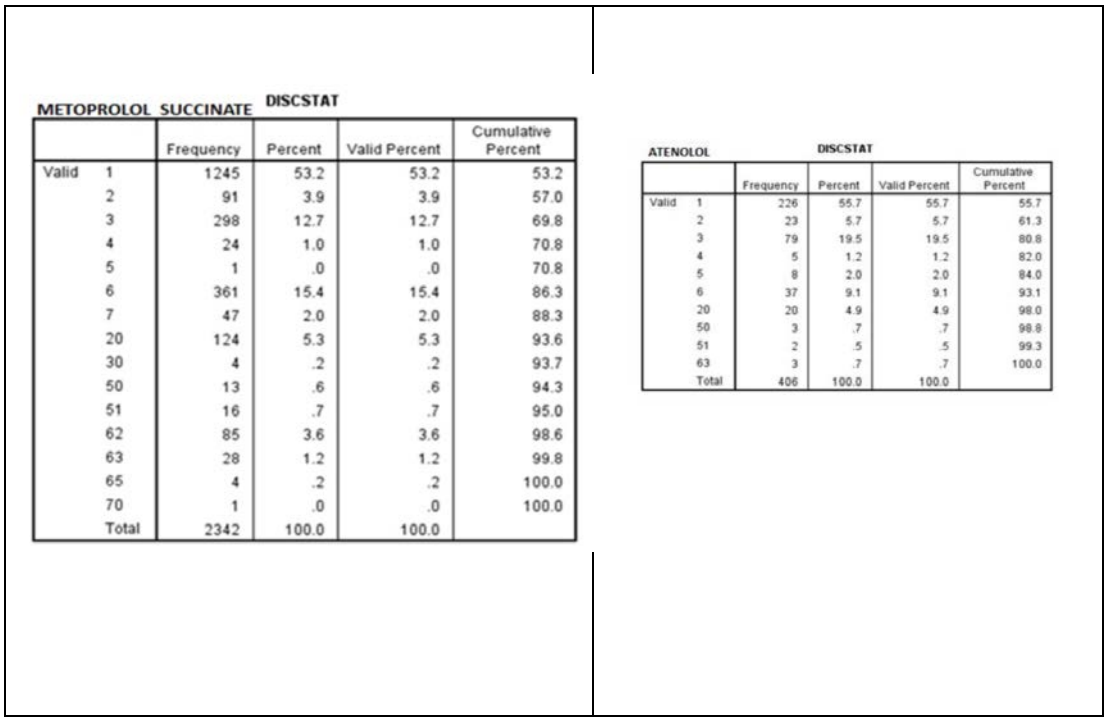


Figure 3 – Patient Discharge codes as per hospital codes
The patient discharge codes corresponding with each beta-blocker. The discharge codes are indicated in numbers where 1 denotes that the patient was discharged to go home, 2 – discharged to other short term general hospital for inpatient care, 3 – transferred to a skilled nursing facility (SNF), 4 – transferred to intermediate care facility (ICF), 5 – transferred to another institution for inpatient care, 6 – transferred to home care of organized home health service organization, 7 – left against medical advice, 20 – expired, 30 – still a patient, 50 – home hospice, 51 – medical facility hospice, 62 discharged to inpatient rehabilitation facility, 63 – transferred to long term hospital care, 65 – transferred to psychiatric hospital, 70 the patient was transferred to an undefined health care institution.



METOPROLOL SUCCINATE DISCSTAT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1245	53.2	53.2	53.2
	2	91	3.9	3.9	57.0
	3	298	12.7	12.7	69.8
	4	24	1.0	1.0	70.8
	5	1	.0	.0	70.8
	6	361	15.4	15.4	86.3
	7	47	2.0	2.0	88.3
	20	124	5.3	5.3	93.6
	30	4	.2	.2	93.7
	50	13	.6	.6	94.3
	51	16	.7	.7	95.0
	62	85	3.6	3.6	98.6
	63	28	1.2	1.2	99.8
	65	4	.2	.2	100.0
	70	1	.0	.0	100.0
	Total	2342	100.0	100.0	

ATENOLOL DISCSTAT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	226	55.7	55.7	55.7
	2	23	5.7	5.7	61.3
	3	79	19.5	19.5	80.8
	4	5	1.2	1.2	82.0
	5	8	2.0	2.0	84.0
	6	37	9.1	9.1	93.1
	20	20	4.9	4.9	98.0
	50	3	.7	.7	98.8
	51	2	.5	.5	99.3
	63	3	.7	.7	100.0
	Total	406	100.0	100.0	

Figure 4 – Patient Primary Diagnostic Codes for Cardiovascular Disease
The frequency of the cardiovascular primary diagnosis that the patients were admitted to the hospital for. The first code denoted as 1 was assigned to patients with ASHD, 2 – AMI, 3 - CHF, 4 – HD, 5 – CVA/TIA, 6 – PD, 7 – AFIB and 8 – SCA/VA.

METOPROLOL SUCCINATE PRIMDIAG

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	416	17.8	17.8	17.8
	2	31	1.3	1.3	19.1
	3	563	24.0	24.0	43.1
	4	103	4.4	4.4	47.5
	5	340	14.5	14.5	62.0
	6	311	13.3	13.3	75.3
	7	305	13.0	13.0	88.3
	8	273	11.7	11.7	100.0
	Total	2342	100.0	100.0	

ATENOLOL PRIMDIAG

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	63	15.5	15.5	15.5
	2	2	.5	.5	16.0
	3	158	38.9	38.9	54.9
	4	3	.7	.7	55.7
	5	48	11.8	11.8	67.5
	6	67	16.5	16.5	84.0
	7	25	6.2	6.2	90.1
	8	40	9.9	9.9	100.0
	Total	406	100.0	100.0	

Supplementary materials

USRDS Selective Data Summary Report

Data Source: United States Renal Data Systems (USRDS)

Experimental Unit: Study Medicare End Stage Renal Disease (ESRDS) patient population cohort.
This report contains CSV file information and column headers and number of rows for the following files:

1. Hospital files
2. PDE Part D Prescription Drug Enrollment Claims
3. IN (Institutional Claims) Files
4. PS (Physician/Supplier) Files
5. PAYHIST
6. MEDEV05

Supplementary Table 1 contains the raw information as it was downloaded from MBOX.

Table 1: Raw Data Summary from USRDS

Type of files/data	Format of files	Size of each individual file	Number of files
Clinical Performance	SAS	Min 800 MB-Max10 GB	2
CROWNWeb	SAS	3.5 GB	1
ESRD Claims	SAS	1 GB – 5.3 GB	46
2015 Medicare Claims Clinical	SAS	Min 4.5 GB – Max 1.8 GB	1
2015 Hospital	SAS	Min 2.7 GB – Max 15.4GB	1
2015 Transplant	SAS	Min 381 KB – Max 9.2GB	1
2015 Core	SAS	Min 166 KB – Max 2.5 GB	1

Supplementary Table 2 contains information of the opened files as they were in SAS format and then converted into CSV.

Table 3: Raw SAS Data Converted to CSV

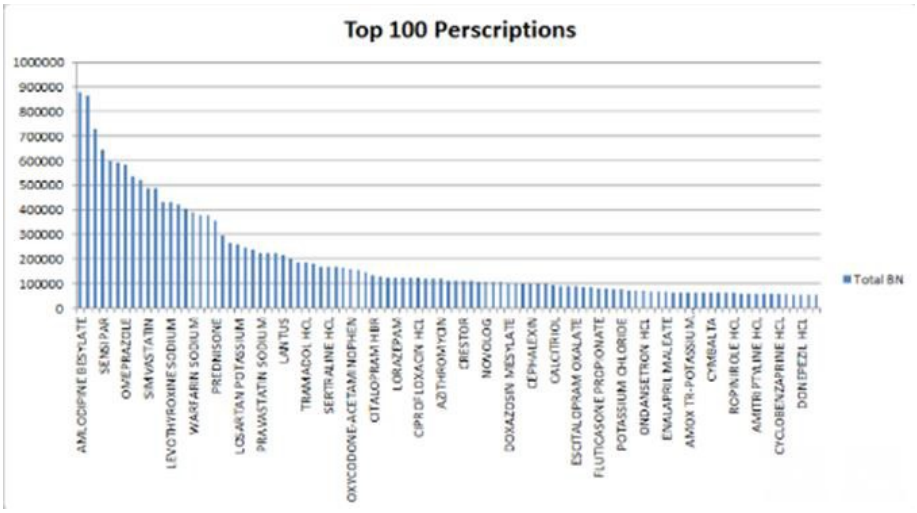
Type of files	Raw File Format	Individual Raw File Size	Converted format of file	Individual File Size	# Columns	# of Rows	Dates	Unique USRDS
Hos_to2009 + Hos2010on	SAS	654105 KB	CSV	988,989 KB	71	1048576 +	2010-2013	714889
PDE (ESRD	SAS	6532 KB	CSV	136126 KB	25	1048576 +	2010	204927
PDE (ESRD	SAS	675532 KB	CSV	745727 KB	25	1048576 +	2011	208854
PDE (ESRD	SAS	857540KB	CSV	376280KB	25	1048576 +	2012	394892
PDE (ESRD	SAS	1080230 KB	CSV	5130937 KB	25	1048576 +	2013	637174
IN(ESRD	SAS	NA	CSV	NA	NA	NA	NA	NA
IN->DET**4FILES**	SAS		CSV	576342 KB	4	1048576 +	2010	637174
IN -> REV **14**	SAS	835025KB	CSV	1175479		1048576 +		637174
IN -> INC **2**	SAS	599105 KB	CSV	878886	26	1048576 +	2010	637174
PS	SAS		CSV		19	1048576 +		162085
Payhist (Core)	SAS	218 921 KB	CSV	235,110 KB	6	1048576 +	1965-2015	324398
MEDEV05	SAS	5386 KB	CSV	8239 KB	114	14555	2005-2009	99681

Supplementary Table 3 – Statistical tests
Output data from IBM spss to test the significance of the data sets under different conditions.

Multivariate Tests[a]

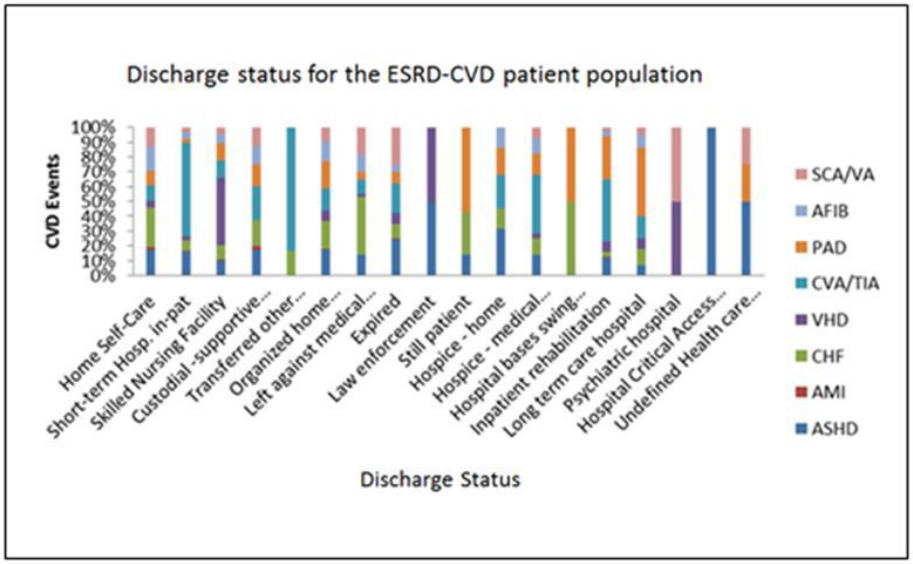
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.767	3013.841[b]	2.000	1836.000	.000	.767	6027.681	1.000
	Wilks' Lambda	.233	3013.841[b]	2.000	1836.000	.000	.767	6027.681	1.000
	Hotelling's Trace	3.283	3013.841[b]	2.000	1836.000	.000	.767	6027.681	1.000
	Roy's Largest Root	3.283	3013.841[b]	2.000	1836.000	.000	.767	6027.681	1.000
DISCSTAT	Pillai's Trace	.015	6.972	4.000	3674.000	.000	.008	27.887	.995
	Wilks' Lambda	.985	6.991[b]	4.000	3672.000	.000	.008	27.963	.995
	Hotelling's Trace	.015	7.010	4.000	3670.000	.000	.008	28.039	.995
	Roy's Largest Root	.015	13.507[c]	2.000	1837.000	.000	.014	27.014	.998
PRIMDIAG	Pillai's Trace	.028	3.682	14.000	3674.000	.000	.014	51.552	1.000
	Wilks' Lambda	.972	3.687[b]	14.000	3672.000	.000	.014	51.619	1.000
	Hotelling's Trace	.028	3.692	14.000	3670.000	.000	.014	51.685	1.000
	Roy's Largest Root	.021	5.592[c]	7.000	1837.000	.000	.021	39.145	.999
GNN	Pillai's Trace	.043	41.313[b]	2.000	1836.000	.000	.043	82.626	1.000
	Wilks' Lambda	.957	41.313[b]	2.000	1836.000	.000	.043	82.626	1.000
	Hotelling's Trace	.045	41.313[b]	2.000	1836.000	.000	.043	82.626	1.000
	Roy's Largest Root	.045	41.313[b]	2.000	1836.000	.000	.043	82.626	1.000
SEX	Pillai's Trace	.003	3.119[b]	2.000	1836.000	.044	.003	6.237	.601
	Wilks' Lambda	.997	3.119[b]	2.000	1836.000	.044	.003	6.237	.601
	Hotelling's Trace	.003	3.119[b]	2.000	1836.000	.044	.003	6.237	.601
	Roy's Largest Root	.003	3.119[b]	2.000	1836.000	.044	.003	6.237	.601
ETHN	Pillai's Trace	.024	22.643[b]	2.000	1836.000	.000	.024	45.286	1.000
	Wilks' Lambda	.976	22.643[b]	2.000	1836.000	.000	.024	45.286	1.000
	Hotelling's Trace	.025	22.643[b]	2.000	1836.000	.000	.024	45.286	1.000
	Roy's Largest Root	.025	22.643[b]	2.000	1836.000	.000	.024	45.286	1.000
DISCSTAT * PRIMDIAG	Pillai's Trace	.121	9.061	26.000	3674.000	.000	.060	235.595	1.000
	Wilks' Lambda	.883	9.071[b]	26.000	3672.000	.000	.060	235.835	1.000
	Hotelling's Trace	.129	9.080	26.000	3670.000	.000	.060	236.075	1.000
	Roy's Largest Root	.079	11.154[c]	13.000	1837.000	.000	.073	145.008	1.000
DISCSTAT * GNN	Pillai's Trace	.004	2.071	4.000	3674.000	.082	.002	8.283	.621
	Wilks' Lambda	.996	2.071[b]	4.000	3672.000	.082	.002	8.286	.621
	Hotelling's Trace	.005	2.072	4.000	3670.000	.082	.002	8.289	.622
	Roy's Largest Root	.004	3.945[c]	2.000	1837.000	.020	.004	7.891	.710
DISCSTAT * SEX	Pillai's Trace	.005	2.508	4.000	3674.000	.040	.003	10.032	.717
	Wilks' Lambda	.995	2.508[b]	4.000	3672.000	.040	.003	10.033	.717
	Hotelling's Trace	.005	2.509	4.000	3670.000	.040	.003	10.034	.717
	Roy's Largest Root	.005	4.281[c]	2.000	1837.000	.014	.005	8.562	.748
DISCSTAT * ETHN	Pillai's Trace	.042	19.662	4.000	3674.000	.000	.021	78.647	1.000
	Wilks' Lambda	.958	19.814[b]	4.000	3672.000	.000	.021	79.257	1.000
	Hotelling's Trace	.044	19.967	4.000	3670.000	.000	.021	79.866	1.000
	Roy's Largest Root	.041	37.497[c]	2.000	1837.000	.000	.039	74.994	1.000
PRIMDIAG * GNN	Pillai's Trace	.049	7.751	12.000	3674.000	.000	.025	93.007	1.000
	Wilks' Lambda	.951	7.834[b]	12.000	3672.000	.000	.025	94.009	1.000
	Hotelling's Trace	.052	7.918	12.000	3670.000	.000	.025	95.010	1.000
	Roy's Largest Root	.050	15.352[c]	6.000	1837.000	.000	.048	92.111	1.000
PRIMDIAG * SEX	Pillai's Trace	.040	6.328	12.000	3674.000	.000	.020	75.935	1.000
	Wilks' Lambda	.960	6.364[b]	12.000	3672.000	.000	.020	76.370	1.000
	Hotelling's Trace	.042	6.400	12.000	3670.000	.000	.020	76.804	1.000
	Roy's Largest Root	.037	11.390[c]	6.000	1837.000	.000	.036	68.337	1.000
PRIMDIAG * ETHN	Pillai's Trace	.041	5.520	14.000	3674.000	.000	.021	77.286	1.000
	Wilks' Lambda	.959	5.542[b]	14.000	3672.000	.000	.021	77.591	1.000
	Hotelling's Trace	.042	5.564	14.000	3670.000	.000	.021	77.895	1.000
	Roy's Largest Root	.035	9.214[c]	7.000	1837.000	.000	.034	64.499	1.000
GNN * SEX	Pillai's Trace	.002	2.088[b]	2.000	1836.000	.124	.002	4.176	.431
	Wilks' Lambda	.998	2.088[b]	2.000	1836.000	.124	.002	4.176	.431
	Hotelling's Trace	.002	2.088[b]	2.000	1836.000	.124	.002	4.176	.431
	Roy's Largest Root	.002	2.088[b]	2.000	1836.000	.124	.002	4.176	.431
GNN * ETHN	Pillai's Trace	.056	54.814[b]	2.000	1836.000	.000	.056	109.628	1.000
	Wilks' Lambda	.944	54.814[b]	2.000	1836.000	.000	.056	109.628	1.000
	Hotelling's Trace	.060	54.814[b]	2.000	1836.000	.000	.056	109.628	1.000
	Roy's Largest Root	.060	54.814[b]	2.000	1836.000	.000	.056	109.628	1.000
SEX * ETHN	Pillai's Trace	.008	6.971[b]	2.000	1836.000	.001	.008	13.942	.927
	Wilks' Lambda	.992	6.971[b]	2.000	1836.000	.001	.008	13.942	.927
	Hotelling's Trace	.008	6.971[b]	2.000	1836.000	.001	.008	13.942	.927
	Roy's Largest Root	.008	6.971[b]	2.000	1836.000	.001	.008	13.942	.927

Supplementary Figure 1 – Top 100 drugs given to the CKD patient population.



Supplementary Figure 2 – Discharge status for the eight cardiovascular events analyzed.

The x-axis denotes the patient discharge status. The y-axis denotes the frequency of the cardiovascular disease.



Reference

1. USRDS in ESRD Quarterly UpdateUnited States of America; 2015).
2. USRDS 2016).
3. Collins, A.J., Foley, R.N., Gilbertson, D.T. & Chen, S.C. United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. *Kidney international supplements* **5**, 2-7 (2015).
4. Foley, R.N. & Collins, A.J. The USRDS: what you need to know about what it can and can't tell us about ESRD. *Clinical journal of the American Society of Nephrology : CJASN* **8**, 845-851 (2013).
5. McDonough, M.J. CROWNWeb--the evolution of kidney data management (Part I). *Nephrology news & issues* **22**, 42, 44, 46 (2008).
6. McDonough, M.J. CROWNWeb: the evolution of kidney data management. 2. *Nephrology news & issues* **22**, 51-54 (2008).
7. Diamond, L.H. & Howard, A.D. The ESRD Quality Incentive Program: The Current Limitations of Evidence and Data to Develop Measures, Drive Improvement, and Incentivize Outcomes. *Advances in chronic kidney disease* **23**, 377-384 (2016).
8. Delva, O. & Seaton, J. CROWNWeb: producing data to measure performance. *Nephrology news & issues* **27**, 22, 24-25 (2013).
9. Delva, O. How CROWNWeb will validate and utilize patient data. *Nephrology news & issues* **24**, 30-33 (2010).
10. Perez-Gomez, M.V. et al. Horizon 2020 in Diabetic Kidney Disease: The Clinical Trial Pipeline for Add-On Therapies on Top of Renin Angiotensin System Blockade. *Journal of clinical medicine* **4**, 1325-1347 (2015).
11. Zhang, J. & Wang, N. Prognostic significance and therapeutic option of heart rate variability in chronic kidney disease. *International urology and nephrology* **46**, 19-25 (2014).
12. Hoot, N.R., Benitez, J.G. & Palm, K.H. Hemodynamically unstable: accidental atenolol toxicity? *The Journal of emergency medicine* **45**, 355-357 (2013).
13. Fleet, J.L. et al. Kidney function and population-based outcomes of initiating oral atenolol versus metoprolol tartrate in older adults. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **64**, 883-891 (2014).
14. Shireman, T.I., Mahnken, J.D., Phadnis, M.A. & Ellerbeck, E.F. Effectiveness comparison of cardio-selective to non-selective beta-blockers and their association with mortality and morbidity in end-stage renal disease: a retrospective cohort study. *BMC cardiovascular disorders* **16**, 60 (2016).
15. Dargie, H.J. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. *Lancet* **357**, 1385-1390 (2001).
16. Packer, M. et al. Effect of carvedilol on survival in severe chronic heart failure. *The New England journal of medicine* **344**, 1651-1658 (2001).
17. Wong, M.C. et al. The effectiveness of metoprolol versus atenolol on prevention of all-cause and cardiovascular mortality in a large Chinese population: a cohort study. *International journal of cardiology* **175**, 425-432 (2014).
18. Heffernan, K.S. et al. Effect of atenolol vs metoprolol succinate on vascular function in patients with hypertension. *Clinical cardiology* **34**, 39-44 (2011).
19. Gitsels, L.A., Kulinskaya, E. & Steel, N. Survival prospects after acute myocardial

infarction in the UK: a matched cohort study 1987-2011. *BMJ open* **7**, e013570 (2017).
20. Mirault, T. et al. Impact of betablockers on general and local outcome in patients hospitalized for lower extremity peripheral artery disease: The COPART Registry. *Medicine* **96**, e5916 (2017).
21. Berg, S.K. et al. Patient-reported outcomes at hospital discharge from Heart Centres, a national cross-sectional survey with a register-based follow-up: the DenHeart study protocol. *BMJ open* **4**, e004709 (2014).
22. Chon, G.R. et al. Clinical outcomes of witnessed and monitored cases of in-hospital cardiac arrest in the general ward of a university hospital in Korea. *Respiratory care* **58**, 1937-1944 (2013).
23. DiNicolantonio, J.J. et al. beta-Blockers in hypertension, diabetes, heart failure and acute myocardial infarction: a review of the literature. *Open heart* **2**, e000230 (2015).
24. Kutner, N.G., Johansen, K.L., Zhang, R., Huang, Y. & Amaral, S. Perspectives on the new kidney disease education benefit: early awareness, race and kidney transplant access in a USRDS study. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* **12**, 1017-1023 (2012).
25. Improvements in data quality in the USRDS database: determining treatment modalities. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **20**, 89-94 (1992).
26. Cheng, X. et al. Mortality rates among prevalent hemodialysis patients in Beijing: a comparison with USRDS data. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **28**, 724-732 (2013).
27. Girbes, A.R.J. & Marik, P.E. Protocols for the obvious: Where does it start, and stop? *Annals of intensive care* **7**, 42 (2017).
28. Banala, S.R. et al. Discharge or admit? Emergency department management of incidental pulmonary embolism in patients with cancer: a retrospective study. *International journal of emergency medicine* **10**, 19 (2017).
29. Baran, J. et al. GFVO: the Genomic Feature and Variation Ontology. *PeerJ* **3**, e933 (2015).
30. Rouder, J.N. The what, why, and how of born-open data. *Behavior research methods* **48**, 1062-1069 (2016).
31. Chalk, S.J. The Open Spectral Database: an open platform for sharing and searching spectral data. *Journal of cheminformatics* **8**, 55 (2016).
32. Krishnan, M., Weinhandl, E.D., Jackson, S., Gilbertson, D.T. & Lacson, E., Jr. Comorbidity ascertainment from the ESRD Medical Evidence Report and Medicare claims around dialysis initiation: a comparison using US Renal Data System data. *American journal of kidney diseases : the official journal of the National Kidney Foundation* **66**, 802-812 (2015).
33. Bakir, D., James, A.P. & Zollanvari, A. An efficient method to estimate the optimum regularization parameter in RLDA. *Bioinformatics* **32**, 3461-3468 (2016).
34. Simonyan, V. et al. High-performance integrated virtual environment (HIVE): a robust infrastructure for next-generation sequence data analysis. *Database : the journal of biological databases and curation* **2016** (2016).
35. Yin, G., Lam, C.K. & Shi, H. Bayesian randomized clinical trials: From fixed to adaptive design. *Contemporary clinical trials* **59**, 77-86 (2017).
36. Wheeler, M.W., Bailer, A.J., Cole, T., Park, R.M. & Shao, K. Bayesian Quantile Impairment Threshold Benchmark Dose Estimation for Continuous Endpoints. *Risk analysis : an*

official publication of the Society for Risk Analysis (2017).
37. Heras, M. et al. Elderly patients with chronic kidney disease: outcomes after 5 years of follow-up. *Nefrologia : publicacion oficial de la Sociedad Espanola Nefrologia* **32**, 300-305 (2012).
38. McQuillan, R.F. & Chan, C.T. The intuitive case for beta-blockers in patients with ESRD. *Seminars in dialysis* **25**, 15-21 (2012).
39. Badve, S.V. et al. Effects of beta-adrenergic antagonists in patients with chronic kidney disease: a systematic review and meta-analysis. *Journal of the American College of Cardiology* **58**, 1152-1161 (2011).
40. Lin, T.T. et al. Primary prevention of atrial fibrillation with beta-blockers in patients with end-stage renal disease undergoing dialysis. *Scientific reports* **5**, 17731 (2015).
41. McCullough, P.A. et al. Benefits of aspirin and beta-blockade after myocardial infarction in patients with chronic kidney disease. *American heart journal* **144**, 226-232 (2002).
42. Segall, L., Nistor, I. & Covic, A. Heart failure in patients with chronic kidney disease: a systematic integrative review. *BioMed research international* **2014**, 937398 (2014).
43. Sharma, A., Colvin-Adams, M. & Yancy, C.W. Heart failure in African Americans: disparities can be overcome. *Cleveland Clinic journal of medicine* **81**, 301-311 (2014).
44. Peperstraete, B. [Management of comorbidities in heart failure]. *Revue medicale de Bruxelles* **34**, 154-162 (2013).
45. Correa-de-Araujo, R. et al. Gender differences across racial and ethnic groups in the quality of care for acute myocardial infarction and heart failure associated with comorbidities. *Women's health issues : official publication of the Jacobs Institute of Women's Health* **16**, 44-55 (2006).
46. Conde-Martel, A. et al. Gender related differences in clinical profile and outcome of patients with heart failure. Results of the RICA Registry. *Revista clinica espanola* **215**, 363-370 (2015).
47. Weir, M.A. et al. beta-Blocker dialyzability and mortality in older patients receiving hemodialysis. *Journal of the American Society of Nephrology : JASN* **26**, 987-996 (2015).
48. Tomaszuk-Kazberuk, A. et al. Pharmacologic treatment in dialysis patients hospitalized for cardiovascular reasons: do we follow the guidelines? *Renal failure* **35**, 235-242 (2013).
49. Esquitin, R. et al. Left ventricular hypertrophy by electrocardiography and echocardiography in the African American Study of Kidney Disease Cohort Study. *Journal of the American Society of Hypertension : JASH* **6**, 193-200 (2012).