12-2018

Characterization of Cancer Incidence in Chronic Kidney Disease Patients: A Single Center Study, 2008 – 2018

Jagadeesh Puvvula
University of Nebraska Medical Center

Characterization of cancer incidence in Chronic Kidney Disease patients: A single center study, 2008 – 2018

Jagadeesh Puvvula, PharmD

Capstone

Fall 2018

Table of Contents

Abstract

Background: Chronic kidney disease (CKD) is an established risk factor for cardiovascular disease, but the significance of lower Glomerular filtration rate on incidence and mortality from cancer is uncertain. There is evidence that cancer risk and cancer mortality may be increased in individuals with End Stage Renal Disease (ESRD) requiring dialysis or after renal transplantation but whether less severe kidney disease is associated with cancer remains poorly understood.

Some studies indicate that these associations do appear to be organ specific. It has been reported that reduced renal function is associated with an increased risk of cancers of the kidney and Genito-urinary system, gastrointestinal, lung and some soft tissue and hematological cancers like myeloma.

Objectives: To characterize the Chronic Kidney Disease (CKD) patients by cancer incidents. Additionally, to calculate the crude and age-adjusted cancer incidence in individuals with compromised kidney function.

Methods: A retrospective cohort study was conducted using the electronic medical record information from the patients received treatment at Nebraska Medicine from 2008 to 2018, extracted by UNMC Public Health Informatics lab. The subjects with Chronic Kidney Disease and eventually advanced to cancer were included in the study (N=869). The subjects are characterized by descriptive comparison of age, gender, race, most frequent medical conditions, body mass index, smoking status, and frequently diagnosed cancer.

Results: The cohort is predominant with Caucasian (n=702; 80.78%) male (n=486; 55.93%) population with median age 71 (IQR: 63.53). Major proportion of the individuals in the cohort are obese (n=283; 32.57%). The frequent cancer types among the cohort are digestive (n=129;

14.85%), urinary (n=105; 12.08%), respiratory (n=96; 11.05%), skin (n=95; 10.93%) and leukemia (n=90; 10.36%). The crude/age-adjusted incidence of the frequently observed cancer groups was exponentially higher than the national, state and the city estimates.

Introduction

Placement site

The Big Data to Knowledge (BD2K) lab, is a part of the UNMC Division of Public Health Informatics, located at DRC II. Nebraska Medical Center. The research at Big Data to Knowledge Lab mainly focused on mining patient information from clinical practice, rendering information into machine-readable formats and analyzing the data using computational techniques to support clinical care, population health, and analytics. The key focus areas are extracting clinical information involves capturing information generated from the interaction between the patient and the healthcare team. Capturing clinical information in a machine-readable format, using SNOMED CT, RxNORM, and LOINC. Developing machine learning algorithms to generate evidence-based insights that benefit clinical practice and population health. In response to the patient-centered outcome research initiative to provide timely answers to the research questions to draw beneficial health outcomes, the BD2K lab developed clinical data models that retrieve relevant information from Electronic Health Records.

Problem Statement

There are no evidence-based cancer screening recommendations tailored for patients with Chronic Kidney Disease. If estimated Glomerular Filtration Rate is associated with an increased risk of

certain organ-specific cancers, then could this have implications for directing cancer screening efforts in this select population with Chronic Kidney Disease.

Importance of the project to the organization/clinical community

The capstone activity is a collaboration between Nebraska Medicine and UNMC, wherein the clinical aspects of the project are initiated by Dr. Apar Ganti (Oncologist) and Dr. Ketki Tendulkar (Nephrologist), noticed a significant number of patients with compromised renal function are being diagnosed with cancer over time. So, characterizing and identifying the risks could directly impact the clinical care of Chronic Kidney Disease patients regarding the frequency and method of cancer screening aids early detection of cancer.

Purpose and Rationale of the Study:

Majority of the epidemiologic studies are based on the national/state cancer surveillance database, wherein the information is data rich to calculate the incidence and mortality for specific cancer type. The cancer surveillance databases often fail to provide patient-related clinical information. Therefore, the proposed study is utilizing the i2b2 database that contains all the clinical data that is recorded in the patient's electronic health record. This includes patient disease history, medication history, diagnostic test results (histopathology, imaging, genetic tests, organ-specific diagnostic profiles) ever conducted during the patient encounter.

The study fills the gaps/possible confounders while evaluating the epidemiologic association between cancer and Chronic Kidney Disease.

Goals and Objectives:

Aim: To characterize chronic kidney disease (CKD) patient population with different cancer types at Nebraska Medicine, Omaha, NE, from 2008 to 2018.

Hypothesis

We hypothesize that there is an increased risk of cancer incidence associated with renal impairment (dialysis dependent and patients with glomerular filtration rate less than 60 ml/min/1.73 m^3).

Literature review

The patients with renal failure are at higher risk to develop cancer and with higher cancer mortality rate than compared to individuals with optimum kidney function. Multiple population-based studies conducted to analyze the association between impaired kidney function and cancer have reported an increased risk associated with thyroid/endocrine, gastrointestinal, hepatobiliary, urinary tract, kidney cancer and for multiple myeloma and chronic myeloid leukemia.

The incidence pattern across different countries varied, a cohort from Taiwan showed a higher incidence of bladder cancer (SIR 8.2 95% CI 6.7-9.9) (Lin et al., 2012) and lung cancer compared to kidney cancer (HR 8.81 95% CI 6.62-11.72) (Chien et al., 2017). Additionally, Kidney cancer is higher in the United States for both dialysis-dependent patients (OR 2.42 95% CI 2.01-2.92) (Shebl, Warren, Eggers, & Engels, 2012) and patients with glomerular filtrate rate less than 60 (Christensson et al., 2013; Lowrance, Ordonez, Udaltsova, Russo, & Go, 2014), Australia and New Zealand (SIR 9.9 95% CI 7.7-12.3) and Europe (SIR 3.3 95% CI 3.1-3.6) (Maisonneuve et al., 1999). The cohort from Korea showed a higher incidence of urinary tract cancer (SIR 4.7 95% CI 2.42-8.19) (Yoo et al., 2017).

A study based on white Australian cohort identified that risk of cancer death increases by 29% (mean follow up 10.1 years) and 18% (mean follow up period 12.8 years) for every 10 ml/min/1.73 m^3 reduction in estimated Glomerular Filtration Rate, with higher risk associated with lung, breast and urinary tract cancer deaths among patients with reduced kidney function (Iff et al., 2014; Wong

et al., 2009). The overall risk assessment of cancer associated with renal impairment remained inconsistent among literature and geographical locations.

The underlying factor for this observation could be due to genomic instability and genetic damage in Chronic Kidney Disease patients possibly due to the accumulation of uremic toxins, oxidative stress mediators and endogenous substances with genotoxic properties. Observation of DNA damage assessed by comet assay show higher DNA damage in Chronic Kidney Disease patients (21.42 ± 0.41 %) than controls (7.74 ± 0.41%) (Corredor et al., 2015). Other assays such as alkaline comet assay and cytokine block micronucleus assay in peripheral blood lymphocytes have also been used to quantify the DNA and chromosomal damage, and conclude that there is a significant increase in micronucleus frequency and tail DNA intensity in Chronic Kidney Disease patients (Rangel-Lopez et al., 2013).

A study based on pediatric population, used comet/enzyme-modified comet assay to determine the percentage DNA intensity as a measure of genetic damage found and oxidative DNA damage in formamidopyrimidine DNA glycosylase, is significantly higher in a pediatric group with Chronic Kidney Disease (Aykanat et al., 2011). A comet assay of biomarkers of DNA damage including micronuclei, DNA strand breaks found to be higher in Chronic Kidney Disease patients undergoing dialysis (Schupp, Stopper, & Heidland, 2016).

There is a definitive relationship between kidney function and cancer, in terms of both biological assays and population-based studies that demonstrated statistically significant correlations. However, the cancer types across geographical regions varied remarkably.

Methods

Research question:

Do patients with chronic kidney disease patients have an increased risk of developing cancer?

Study Design

A single-center retrospective population-based cohort study that includes patients diagnosed/treated for Chronic Kidney Disease prior to cancer diagnosis, at Nebraska Medicine from 2008 to 2018.

Study Population/Study sample/Data sources/ Data collection methods

The cohort predominantly includes Nebraska residents with Chronic Kidney Disease prior to cancer diagnosis and treated at Nebraska Medicine. For this study, de-identified patient information retrieved from i2b2, through the BD2K lab maintained by Dr. W. Scott Campbell M.B.A., Ph.D.

The information retrieved from i2b2 is rich in clinical information as the data is exchanged directly from the patient's electronic health record. The probability of data errors is expected to be decidedly less, as the automated machine algorithms segregate the information from the electronic health record and involve minimal or no manual data entry during the data exchange.

The de-identified data contain multiple files with a pseudo-patient identification number (primary key), that is mapped to eGFR test results, clinical diagnoses (SNOMED CT concept codes), body mass index ever documented for the patient at Nebraska Medicine, date of birth, gender and race. This allowed me to develop an entity relationship scheme (by one to many or many to one relation) using MySQL Workbench (Ver. 8.0.12) and Tableau Prep (Ver. 2018.2).

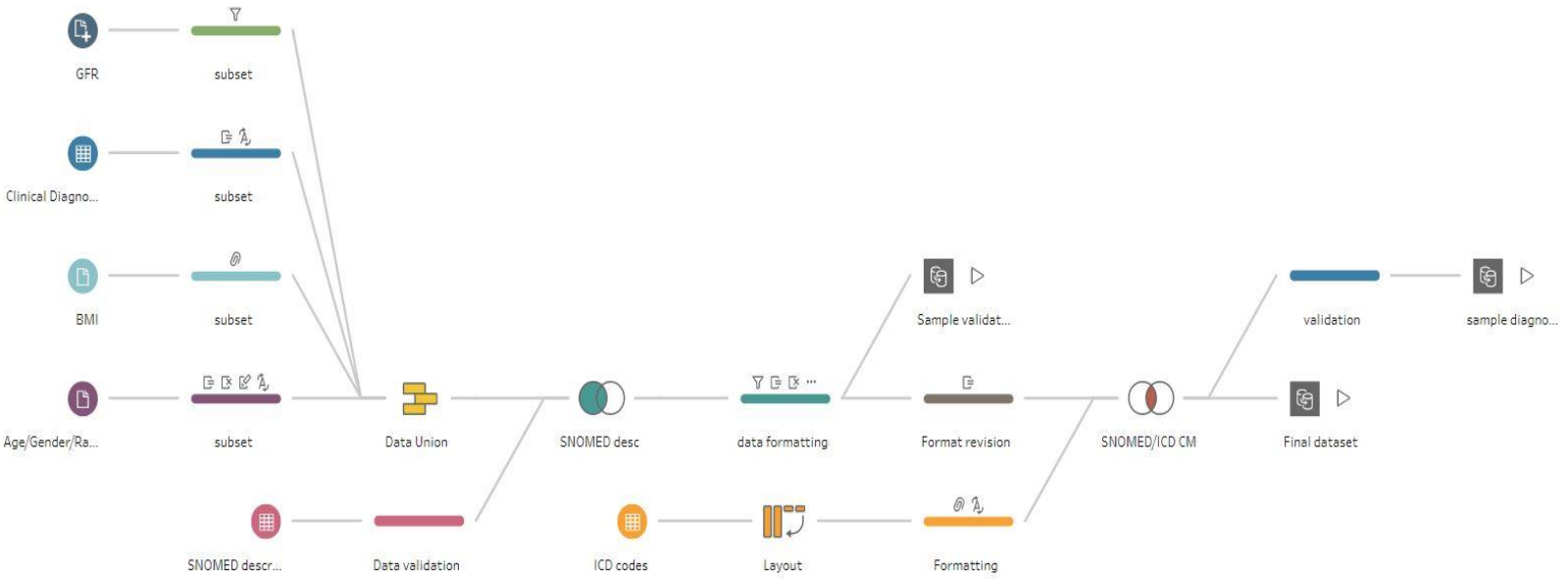

Fig 1. Extraction of analytic dataset

Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) Design:

The SNOMED CT is maintained by the International Health Terminology Standard Development Organization (IHTSDO). Unlike the International Classification of Diseases, SNOMED captures granular clinical

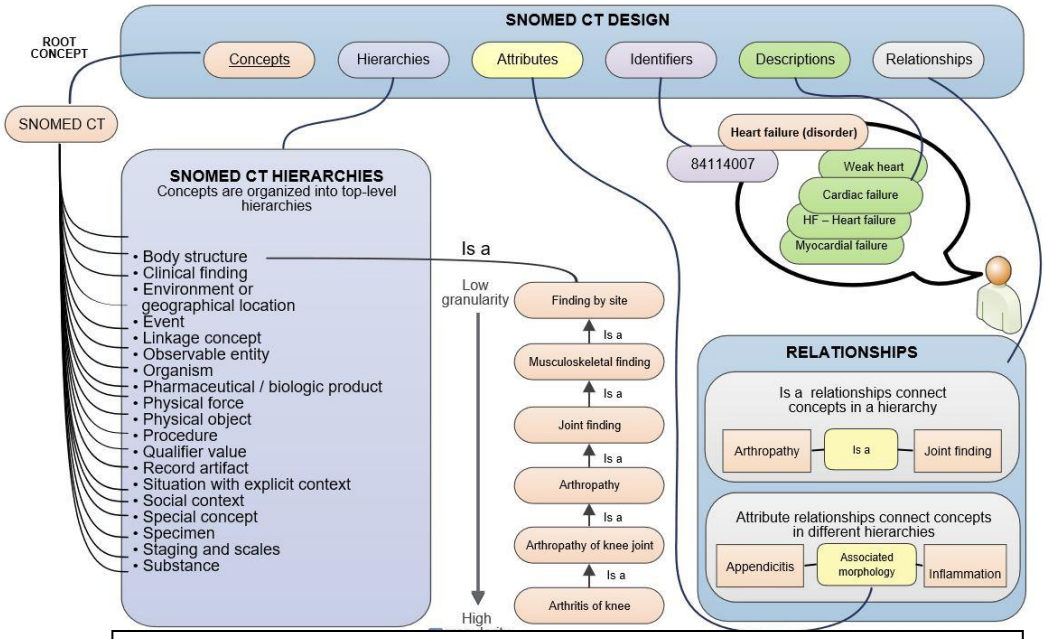

Fig 2. SNOMED CT hierarchy design

information from electronic health records. Allows medical practitioners to capture detailed ontologies and catalog the patient clinical problem list (Bhattacharyya, 2015; SNOMED, 2014).

In the current study, I included all the concept codes related to Chronic kidney disease and cancers. The SNOMED CT contains about 58 different classifications of the parent diagnosis 'Chronic Kidney Disease' (SNOMED, 2014).

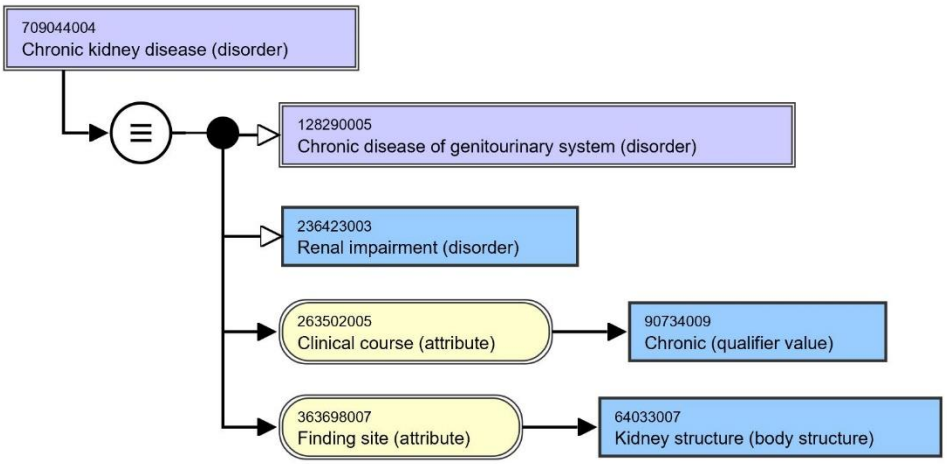

Fig 3. Chronic Kidney Disease schematic representation

Additionally, SNOMED CT code descriptions are extracted using Python (Python Version 2.7.14) Application Programming Interface (API) based on SNOMED English version 20180131 developed by SNOMED international.

Source code (Apache license Ver. 2.0, Jan 2004):

```
from urllib.request import urlopen
from urllib.parse import quote
import json

baseUrl = 'http://browser.ihtsdotools.org/api/v1/snomed/'
edition = 'en-edition'
version = 'v20180131'
#Prints description of a concept ID

def getConceptById(id):
   url = baseUrl + edition + '/' + version + '/concepts/' + id
   response = urlopen(url).read()
   data = json.loads(response.decode('utf-8'))
```

```
  print (data['fsn'])
getConceptById('XXXXXXXXXX')
```

Eligibility criterion

Inclusion

- Subjects with chronic kidney disease and eventually developed cancer.
- Age of 18 and above, received care at Nebraska Medicine between 2008 and 2018
- Subjects with at least three estimated Glomerular Filtration Rate results available in a year before cancer diagnosis.

Exclusion

- All the subjects diagnosed with secondary and metastatic cancer were excluded.
- All the individuals underwent a kidney transplant before the date of cancer diagnosis are excluded.

Case Ascertainment

The i2b2 database contains detailed de-identified patient information who are treated at Nebraska Medicine.

Measurement of Outcomes

Primary outcome: The primary outcome is to characterize the cohort by age, gender, race by cancer type and eGFR group.

Abstracted Variables

Age, gender, race, the clinical information includes co-morbidities, month and year of Chronic Kidney Disease/cancer diagnosis, cancer type, glomerular filtration rate.

Statistical Methods

The subjects are classified into four estimated Glomerular Filtration Rate groups based on the average of the latest three eGFR estimate results based on individual patient's cancer diagnosis date.

The eGFR grouping is based on Kidney Disease Improving Global Outcomes (KDIGO) 2012 classification and due to smaller sample size, the KIDGO classification is slightly modified (Bargman & Skorecki, 2018).

The cancer related SnomedCT concept code descriptions are mapped to International Classification of Disease for Oncology Version 3.0 Topographical codes and the patients are grouped by primary site of cancer (*International Classification of Diseases for Oncology*, 2013).

The crude and age-adjusted incidence are calculated for the chronic kidney disease population who received treatment from Nebraska Medicine. The population proportions for respective age groups were extracted from US Census Bureau - 2010 Decennial estimates for Omaha (*2010 Census Summary File 1*, 2012). Additionally, the age-adjusted incidence is compared with the US, Nebraska and Omaha-Council Bluffs region. The age-adjusted cancer incidence for US, Nebraska and Omaha-Council Bluffs was extracted from CDC Wonder – cancer incidence database for the year 2006 – 2015 including the population of age 20 years and older for comparison (Control & Prevention, 2008).

The descriptive statistics for the variables age, gender, ethnicity, medical history, body mass index, smoking history by cancer groups and estimated Glomerular Filtration Rate groups are generated using SAS software, version 9.4 ("SAS 9.4," 2012). The descriptive statistics and comparison of proportions for categorical data using Chi-Square or Fisher Exact test, categorical and continuous data using Kruskal-Wallis test (Burlew, 2005; Cody, 2011).

Results

A total of 869 subjects are enrolled in the study: 468 male and 383 females with a median age of 71 years (IQR 63.53, 79.52). The subjects are all chronic kidney disease patients who developed cancer and received patient care at Nebraska Medicine from the year 2008 to 2018.

The cohort is predominant with male (55.93%) Caucasians (80.78%) with a median age of 71 years (IQR 63.53, 79.52). The median estimated Glomerular filtration rate is 51 ml/min/ 1.73 m3 (IQR 37.66, 60) and majority (36.48%) of the individuals within estimated glomerular filtration rate between 45 to 50 ml/min/1.73 m3, which is clinically considered as mild to moderately decreased kidney function according to the Kidney Disease Improving Global Outcome (KDIGO) classification of chronic kidney disease (Bargman & Skorecki, 2018).

Classification	Frequency	Percent	At risk	Crude
All cancer cases	869		20400	4259.80
Digestive system	129	14.84	20400	632.35
Urinary system	105	12.08	20400	514.71
Respiratory system	96	11.05	20400	470.59
Skin	95	10.93	20400	465.69
Leukemia	90	10.36	20400	441.18
Male genital system	87	10.01	11240	774.02
Breast	66	7.59	20400	323.53
Neoplasm, uncertain whether benign or malignant	46	5.29	20400	225.49
Sarcoma, NOS	38	4.37	20400	186.27
Female genital system	30	3.45	9160	327.51
Malignant neoplasms of ill-defined, other secondary and unspecified sites	21	2.42	20400	102.94
Oral cavity and Pharynx	18	2.07	20400	88.24
Neuroendocrine carcinoma, NOS	11	1.27	20400	53.92
Carcinoma in situ, NOS	10	1.15	20400	49.02
Endocrine system	10	1.15	20400	49.02
Malignant Neoplasm of Genitourinary Organs	9	1.04	20400	44.12
Bone and joint	4	0.46	20400	19.61
Brain and other nervous system	2	0.23	20400	9.80
Mesothelioma and soft tissue	2	0.23	20400	9.80

Table 1. Crude cancer incidence per 100,000

The most frequent cancer among the cohort are digestive (14.85%), urinary (12.08%), respiratory (11.05%), skin (10.93%) and leukemia (10.36%). More than half (51.32%) of the subjects in the cohort developed cancer within a year from clinical diagnosis of chronic kidney disease. The median time to develop cancer in the cohort is 45 weeks (IQR 3.85, 175.28) post clinical diagnosis of chronic kidney disease.

	Standard Population	Age distribution of a standard population	Cohort population	cancer	Digestive System	Urinary System	Respiratory System	Skin cancer	Leukemia
Age group									
18 - 59	240074	0.783820404	5108	2286.40	245.52	322.24	214.83	306.90	306.90
60 - 69	33177	0.108319974	4192	625.32	87.85	77.52	67.18	62.02	51.68
70 - 79	18947	0.061860281	4933	342.34	57.68	36.37	42.64	38.87	33.86
80+	14089	0.04599934	6167	152.91	24.61	18.65	16.41	14.92	20.14
Total	306287	1	20400	3406.97	415.67	454.78	341.06	422.71	412.58
Standard population estimates retrieved from US Census Bureau 2010 Decennial census for Omaha									

Table 2. Age-adjusted cancer incidence per 100,000

The overall crude and age-adjusted cancer incidence among chronic kidney disease patients are exponentially high compared to the country (US), state (Nebraska) and city (Omaha / Council-Bluffs region). The overall cancer incidence in chronic kidney disease patients is about five times higher than the national average with an incidence rate of 3406.97 per 100,000 population.

The incidence of digestive cancer in chronic kidney disease patients is 3 times higher (415.67 per 100,000 population), urinary system cancers are 23 times higher (415.58 per 100,000 population), respiratory cancer 3 times (314.06 per 100,000 population), skin cancers 13 times higher (422.71 per 100,000) and leukemia 23 times higher (412.58 per 100,000 population) than the US national average.

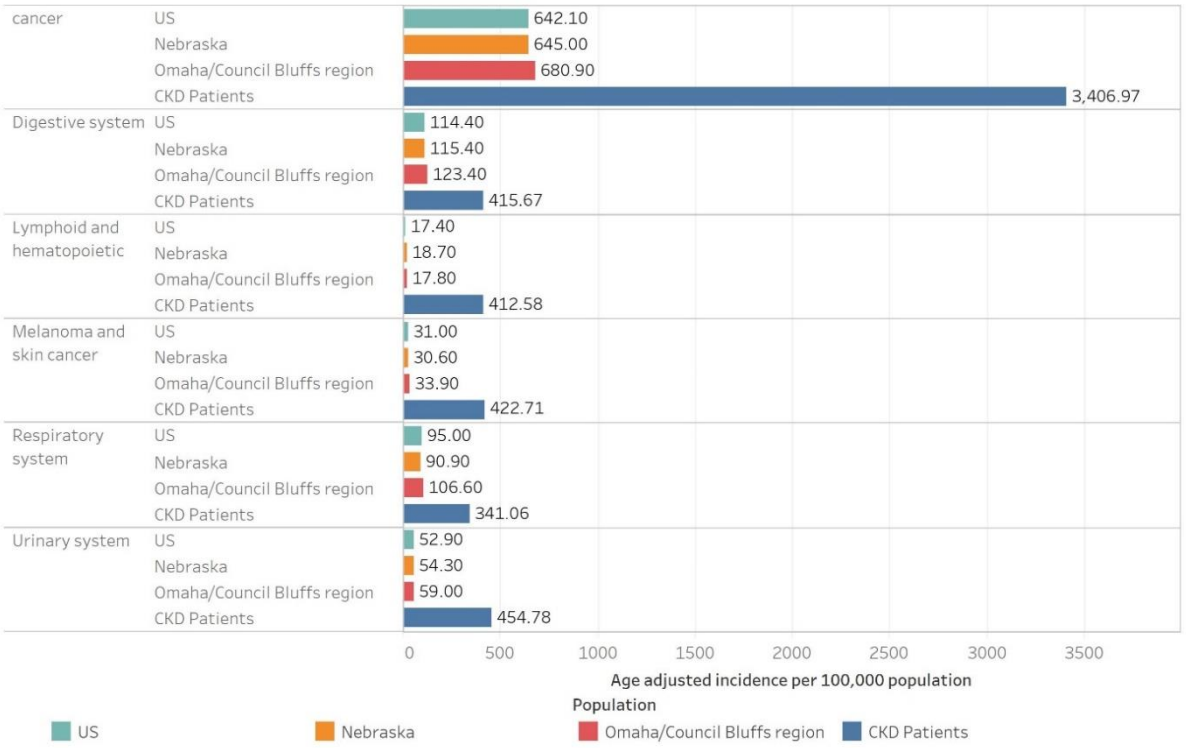

Fig 4. Age-adjusted incidence comparison

About 40 % of the of the cancer types include male genital (n=87; 10.01%), breast (n=66; 7.59%), neoplasms (uncertain whether benign or malignant) (n=46; 5.29%), sarcoma (n=38; 4.37%), female genital (n=30; 3.45%), malignant neoplasms (n=21; 2.42%), oral cavity/pharynx (n=18; 2.07%), neuroendocrine carcinoma (n=11; 1.27%), carcinoma in-situ (n=10; 1.15%), endocrine n=10; 1.15%), malignant neoplasm of genitourinary tract (n=9; 1.04%), bone/joint (n=4; 0.46%), brain/other nervous system and mesothelioma/soft tissue with 2 (0.23%) cases in each category.

All the subjects in the cohort are with at least one or more active clinical diagnosis of chronic disease. The most frequent chronic disease conditions among the cohort are anemia (44.99%), depression (22.78%), diabetes (38.09%), gastroesophageal reflux disease (GERD) (27.85%), heart failure (24.84%), hypertension (82.28%) and hyperlipidemia (46.49%).

Major proportion of the individuals in the cohort are in age group of 70 to 79 years (n=273; 31.42%) and 60 to 69 years (n=242; 27.85%), but age distribution do not have statistical differences among different cancer groups (chi-square = 22.6654, df=21, P=0.3621). Even the median age among cancer groups is not statistically different (Kruskal Wallis = 4.3668, df=7, P=0.7367). In contrast, the frequency of subjects in terms of gender and ethnicity are statistically different among cancer groups. The gender being male has higher number of cancers: digestive (n=69; 7.94%), urinary (n=75; 8.63%), respiratory (n=47; 5.41%), skin cancer (n=60; 6.9%) and leukemia (n=61; 7.02) (chi-square = 13.6324, df=7, P<0.0001). The distribution of cancer patients across different ethnic groups has statistical differences, wherein whites with higher proportion across cancer types: digestive (n=95; 10.93%), urinary (n=84; 9.67%), respiratory (n=77; 8.86%), skin (n=90; 10.36%) and leukemia (n=83; 9.55%) (chi-square = 35.1004, df=14, P=0.0014).

The information on body mass index (BMI) is missing for 36 individuals and the cohort while classified according to the National Institute of Health, National Heart, Lung and Blood Institute (NHLBI) BMI index (Kushner, 2018), majority (n=283; 32.57%) of the individuals are obese and the distribution is statistically different compared to normal weight (n=238; 27.38%), overweight (n=240; 27.62%) and extremely obese (n=72; 8.29%). While the frequencies across cancer types varied from overall cancer, major proportion of individuals with normal weight in digestive (n=49; 5.64%), obese in urinary (n=41; 4.72%) / leukemia (n=33; 3.80%) and overweight in respiratory (n=32; 3.68%) / skin (n=33; 3.80%) (chi-square = 54.2503, df=28, P=0.0021).

In contrast, when the BMI considered as a continuous variable there was no statistical difference observed among the cancer groups, the median BMI among cancer groups are: digestive 27 (IQR 23.09, 31.87), urinary 28 (25.29, 33.82), respiratory 27 (23.39, 31.84), skin 27 (24.19, 31.63) and leukemia 28 (24.56, 31.7) (Kruskal Wallis = 10.4428, df=7, P=0.1648).

There are no statistical differences in patients with anemia (chi-square = 5.8739, df=7, P=0.5545), depression (chi-square = 2.0595, df=7, P=0.9582), diabetes (chi-square = 10.462, df=7, P=0.1639), GERD (chi-square = 7.2363, df=7, P=0.4047), heart failure (chi-square = 4.1677, df=7, P=0.7603), hypertension (chi-square = 13.4567, df=7, P=0.0617) and hyperlipidemia (chi-square = 11.4824, df=7, P=0.1189) across different cancer groups.

To estimate the impact of kidney function to cause cancer, estimated Glomerular Filtration Rate is used as a surrogate, but there are no statistical differences between estimated glomerular filtration rate groups and cancer. A major proportion of the cancer incidents were noticed in patients with mild/mild to moderate kidney dysfunction (n=553; 63.63%) compared to individuals with moderate / severe kidney problem (n= 316; 36.36%) (Chi-Square = 15.4484, df=21, P=0.7998).

The median estimated Glomerular Filtration Rate did not have statistical differences among cancer types: digestive 51 (IQR 36.66, 60), urinary 47 (37.66, 55.33), respiratory 53 (40, 60), skin 53 (44, 60) and leukemia 52 (36.66, 60) (Kruskal Wallis = 9.4029, df=7, P=0.2250). The median estimated Glomerular Filtration Rate among all the cancer types could be classified as KDIGO G3a stage mild to moderate kidney dysfunction (Bargman & Skorecki, 2018).

There were statistically significant differences in the median age with change in clinical kidney function: severe dysfunction (65, IQR: 59.42, 75.38), moderate to severe dysfunction (77, IQR: 67.84, 82.85), mild to moderate dysfunction (72, IQR: 64.72, 80.65) and mild dysfunction (70, IQR: 61.69, 76.81) (Kruskal Wallis = 38.99, df=3, P < 0.001). The distribution of subjects based on kidney function among age groups was statistically different (Chi-Square= 32.72, df=9, P < 0.001). Among the individuals with severe kidney dysfunction, majority of the subjects were within the age group of 60 – 60 years (n=44; 5.06%) and the age groups 70 – 79 years was higher

among moderate to severe dysfunction (n=65; 7.48%), mild to moderate dysfunction (n=98; 11.28%) and mild dysfunction (n=80; 9.21%).

The subject's gender (Chi-Square=6.48, df=3, P = 0.09) and BMI (Chi-Square=14.89, df=12, P = 0.2473) do not have statistical differences among individuals with changes in kidney function. The median BMI among individuals with change in kidney function did not have statistical differences (Kruskal-Wallis= 1.3054, df=3, P = 0.7279). The median BMI in severe dysfunction is 27 (IQR: 22.93, 33.03), moderate to severe dysfunction 27 (IQR: 23.94, 33.36), mild to moderate dysfunction 28 (IQR: 24.27. 34.11) and mild dysfunction 28 (IQR: 23.81, 32.45). According to the NHLBI BMI scale based on the median BMI, all the subjects could be classified as overweight (Kushner, 2018). The ethnic group being Caucasians (n=702; 80.78%) was predominant in the cohort and even across different kidney function categories: severe dysfunction (n=91; 10.47%), severe to moderate dysfunction (n=153; 17.61%), mild to moderate dysfunction (n=262; 30.15%) and mild dysfunction (n=196; 22.55%) which was statistically different compared to the other ethnicities (Chi-Square= 16.98, df=6, P = 0.0093).

The chronic medical conditions anemia (Chi-Square= 22.3259, df=3, P < 0.001) and GERD (Chi-Square= 8.1441, df=3, P = 0.0431) are statistical different among kidney function groups. The spread of anemia across individuals with mild to moderate dysfunction with a higher proportion (n=137; 15.77%) and followed by moderate to severe dysfunction (n=94; 10.82%), mild dysfunction (n=84; 9.67%), severe dysfunction (n=76; 8.75%). The GERD condition among mild to moderate dysfunction (n=103; 11.85%) is higher and mild dysfunction (n=60; 6.9%), moderate to severe dysfunction (n=41; 4.72%) and severe dysfunction (n=38; 4.37%).

The remaining chronic conditions did not differ with change in kidney function: depression (Chi-Square= 0.3931, df=3, P = 0.9417), diabetes (Chi-Square= 3.9309, df=3, P = 0.269), heart failure

(Chi-Square= 3.8074, df=3, P = 0.3607), hypertension (Chi-Square= 4.0886, df=3, P = 0.521) and hyperlipidemia (Chi-Square= 1.5287, df=3, P = 0.6757).

The median time in weeks to develop cancer among individuals with a change in kidney function was not statistically different (Kruskal-Wallis= 0.0969, df=3, P = 0.9922). The median duration in weeks to develop cancer is narrow among individuals with severe dysfunction (45, IQR: 3.85, 175.28) and almost similar among moderate to severe dysfunction (49, IQR: 8.28, 164.28), mild to moderate dysfunction (48, IQR: 5.71, 137.71) and mild dysfunction (49, IQR: 3.21, 138.21).

Additionally, active smoking or smoking history is a major confounding factor associated with cancer development; there are 17 (1.96%) individuals in the cohort.

Discussion

In the current study all the Chronic Kidney Disease patients who eventually developed cancer were included, taking into consideration that the individuals could be with the underlying impaired kidney function before clinically diagnosed as Chronic Kidney Disease.

As hypothesized, the risk of cancer in Chronic Kidney Disease patients was high, and the age-adjusted incidence rate among the cohort was exponentially higher, compared to the National (US), state (Nebraska) and city (Omaha-Council Bluffs region). The cancer incidence is predominant in elderly (median age 71 years; IQR: 63.53, 79.52) male (n=486; 55.93%) Caucasians (n=702; 80.78%). The dominance in elderly could be due to the genetic changes due to aging, the changes in nuclear DNA could alter mutations, chromosomal aneuploidy, copy number variations and telomere shortening (Cabo & Couteur, 2018). In terms of gender, the alterations in genetics due to hormonal differences and molecular changes in male physiology could trigger cancer development (Kim, Lim, & Moon, 2018). Assuming that majority of the individuals receive care at Nebraska

Medicine from closer proximity, the ethnic variation could be due to the proportion of Caucasians (73.1%) in Omaha, NE (*2010 Census Summary File 1*, 2012).

The cohort contains statistical differences in BMI groups, but the median BMI across different cancer groups did not flag some differences. This could be due to the classification of individuals with a borderline BMI into specific groups. Based on the median BMI among the frequent cancer types, the subjects could be classified as overweight. The reason behind overweight consistent among all the cancer types could be due to the reason that visceral adipose tissue which plays a crucial role in producing proinflammatory cytokines. The induction of inflammatory process activates nuclear factor Kappa B (NF-kB) and signal transducer and activation of transcription 3 (STAT3). The NF-kB is a transcription factor has a potential to induce expression of protein structures associated with pro-inflammatory, proliferative and repetitive pathways (Jacob, Varghese, & Weil, 2018).

The most frequent comorbidities in the cohort, anemia is related to the underlying Chronic Kidney Disease condition. In progressive Chronic Kidney Disease condition, the patients fail to produce erythropoietin and reduction in red blood cell survival usually progress to hypo-proliferative anemia, and it is associated with the clinical stage of Chronic Kidney Disease. The chance of a Chronic Kidney Disease patient to develop anemia would multiply in individuals with diabetes (Adamson, 2018). Similarly, the subjects in the cohort have statistical differences with a change in kidney function.

Additionally, chronic medical conditions such as hypertension, hyperlipidemia, heart failure, and diabetes increase the risk of cancer incidence (Tu et al., 2018). The cardiovascular diseases alter hemodynamic and neurohormonal factors that trigger increased reactive oxygen species, altered cardiac gene expression, increases oxidative stress, myocyte necrosis and apoptosis (Helmut

Drexler, 2010). The collective factors increase mutagen sensitivity, DNA damage (Federici et al., 2015). The patients with type 2 diabetes have increased production of reactive oxygen species, reduced antioxidant withhold capacity, and susceptible to DNA damage (Lee & Chan, 2015).

The literature did not support evidence for a specific cancer type, and similarly, the results from the cohort did not have statistical differences of cardiovascular diseases or diabetes across cancer types.

As tobacco smoking is a significant factor associated with cancer incidence, the electronic medical record information on tobacco smoking is documented for 17 individuals. I feel that the lower proportion of the smokers in the cohort could be due to under-reporting of medical information in the health records.

Based on the literature, impaired kidney function is associated to cause oxidative stress due to Vitamin C deficiency, reduced Vitamin E, compromised glutathione system and increased pro-oxidant activity due to aging, chronic inflammatory status. The oxidative stress and increased pro-oxidant activity upon interaction with inflammatory stimuli, oxidant-free radicals are generated by phagocytic cells (Locatelli et al., 2003). The Chronic Kidney Disease patients often present with genomic instability, due to extensive genetic damage caused by a toxin and endogenous genotoxin accumulation (Rangel-Lopez et al., 2013). The genotoxic effect could lead to loss of chromosomal fragments or entire chromosomes, this could lead to the formation of micronuclei upon cell division. The cytokinesis-block micronucleus assay and comet assay revealed chromosomal abnormalities, reduced DNA repair and DNA lesions in Chronic Kidney Disease patients. But the population-based studies are inconsistent in stratifying the risk to a specific cancer type (Roth et al., 2008; Schupp et al., 2016).

The cancer incidence from the cohort did not have statistical differences among estimated Glomerular Kidney Function groups. Based on the median estimated Glomerular Filtration Rate among the cancer groups, the overall cohort could be classified as mild to moderate kidney dysfunction.

Limitations: The study could not be generalized to a country-wide scale, as the samples are not randomly selected. Due to age group stratification which did not precisely match the age groups from the cancer database the incidence comparison group (US, Nebraska and Omaha-Council Bluffs region) is missing with the incidence information for the individuals of age 18 years.

Recommendations for future projects:

The current work is a preliminary analysis of the planned research. The study will be extended to identify the role of drug exposures, time to cancer analysis and estimating hazard proportion ratio among different cancer groups.

Bibliography

2010 Census Summary File 1. (SF1/10-4 (RV)). (2012). Washington: U.S. Census Bureau.

Adamson, J. W. (2018). Iron Deficiency and Other Hypoproliferative Anemias. In J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, & J. Loscalzo (Eds.), *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education.

Aykanat, B., Demircigil, G. C., Fidan, K., Buyan, N., Gulleroglu, K., Baskin, E., . . . Burgaz, S. (2011). Basal damage and oxidative DNA damage in children with chronic kidney disease measured by use of the comet assay. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis, 725*(1-2), 22-28. doi:10.1016/j.mrgentox.2011.07.005

Bargman, J. M., & Skorecki, K. L. (2018). Chronic Kidney Disease. In J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, & J. Loscalzo (Eds.), *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education.

Bhattacharyya, S. (2015). *Introduction to SNOMED CT*: Springer.

Burlew, M. (2005). *SAS guide to report writing: examples*: SAS Publishing.

Cabo, R. d., & Couteur, D. G. L. (2018). The Biology of Aging. In J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, & J. Loscalzo (Eds.), *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education.

Chien, C. C., Han, M. M., Chiu, Y. H., Wang, J. J., Chu, C. C., Hung, C. Y., . . . Weng, S. F. (2017). Epidemiology of cancer in end-stage renal disease dialysis patients: a national cohort study in Taiwan. *Journal of Cancer, 8*(1), 9-18. doi:10.7150/jca.16550

Christensson, A., Savage, C., Sjoberg, D. D., Cronin, A. M., O'Brien, M. F., Lowrance, W., . . . Lilja, H. (2013). Association of cancer with moderately impaired renal function at baseline in a large, representative, population-based cohort followed for up to 30 years. *International Journal of Cancer, 133*(6), 1452-1458. doi:10.1002/ijc.28144

Cody, R. (2011). *SAS statistics by example*: SAS Institute.

Control, C. f. D., & Prevention. (2008). United States Cancer Statistics: 1999–2005 Incidence, Wonder on-Line Database. *Atlanta: United States Department of Health and Human Services, CDC, National Cancer Institute*.

Corredor, Z., Stoyanova, E., Rodriguez-Ribera, L., Coll, E., Silva, I., Diaz, J. M., . . . Pastor, S. (2015). Genomic Damage as a Biomarker of Chronic Kidney Disease Status. *Environmental and Molecular Mutagenesis, 56*(3), 301-312. doi:10.1002/em.21911

Federici, C., Drake, K. M., Rigelsky, C. M., McNelly, L. N., Meade, S. L., Comhair, S. A. A., . . . Aldred, M. A. (2015). Increased Mutagen Sensitivity and DNA Damage in Pulmonary Arterial Hypertension. *American Journal of Respiratory and Critical Care Medicine, 192*(2), 219-228. doi:10.1164/rccm.201411-2128OC

Helmut Drexler, G. H. (2010). Physiology of the Normal and Failing Heart. In M. H. Crawford, MD; DiMarco, John P., MD, PhD; Paulus, Walter J., MD, PhD, FESC (Ed.), *Cardiology* (Third ed., pp. 923-938). Philadelphia, PA: Elsevier.

Iff, S., Craig, J. C., Turner, R., Chapman, J. R., Wang, J. J., Mitchell, P., & Wong, G. (2014). Reduced Estimated GFR and Cancer Mortality. *American Journal of Kidney Diseases, 63*(1), 23-30. doi:10.1053/j.ajkd.2013.07.008

International Classification of Diseases for Oncology. (2013). In C. P. April Fritz, Andrew Jack, Kanagaratnam Shanmugaratnam, Leslie Sobin, D Max Parkin, Sharon Whelan (Ed.).

Jacob, M., Varghese, J., & Weil, P. A. (2018). Cancer: An Overview. In V. W. Rodwell, D. A. Bender, K. M. Botham, P. J. Kennelly, & P. A. Weil (Eds.), *Harper's Illustrated Biochemistry, 31e*. New York, NY: McGraw-Hill Education.

Kim, H.-I., Lim, H., & Moon, A. (2018). Sex Differences in Cancer: Epidemiology, Genetics and Therapy. *Biomolecules & therapeutics, 26*(4), 335-342. doi:10.4062/biomolther.2018.103

Kushner, R. F. (2018). Evaluation and Management of Obesity. In J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, & J. Loscalzo (Eds.), *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education.

Lee, S. C., & Chan, J. (2015). Evidence for DNA Damage as a Biological Link Between Diabetes and Cancer. *Chinese Medical Journal, 128*(11), 1543-1548. doi:10.4103/0366-6999.157693

Lin, H. F., Li, Y. H., Wang, C. H., Chou, C. L., Kuo, D. J., & Fang, T. C. (2012). Increased risk of cancer in chronic dialysis patients: a population-based cohort study in Taiwan. *Nephrology Dialysis Transplantation, 27*(4), 1585-1590. doi:10.1093/ndt/gfr464

Locatelli, F., Canaud, B., Eckardt, K. U., Stenvinkel, P., Wanner, C., & Zoccali, C. (2003). Oxidative stress in end-stage renal disease: an emerging threat to patient outcome. *Nephrology Dialysis Transplantation, 18*(7), 1272-1280. doi:10.1093/ndt/gfg074

Lowrance, W. T., Ordonez, J., Udaltsova, N., Russo, P., & Go, A. S. (2014). CKD and the Risk of Incident Cancer. *Journal of the American Society of Nephrology, 25*(10), 2327-2334. doi:10.1681/asn.2013060604

Maisonneuve, P., Agodoa, L., Gellert, R., Stewart, J. H., Buccianti, G., Lowenfels, A. B., . . . Boyle, P. (1999). Cancer in patients on dialysis for end-stage renal disease: an international collaborative study. *The Lancet, 354*(9173), 93-99. doi:10.1016/S0140-6736(99)06154-1

Rangel-Lopez, A., Paniagua-Medina, M. E., Urban-Reyes, M., Cortes-Arredondo, M., Alvarez-Aguilar, C., Lopez-Meza, J., . . . Paniagua, J. R. (2013). Genetic damage in patients with chronic kidney disease, peritoneal dialysis and haemodialysis: a comparative study. *Mutagenesis, 28*(2), 219-225. doi:10.1093/mutage/ges075

Roth, J. M., Restani, R. G., Goncalves, T. T. S., Sphor, S. L. S., Ness, A. B., Martino-Roth, M. G., & Garcias, G. L. (2008). Genotoxicity evaluation in chronic renal patients undergoing hemodialysis and peritoneal dialysis, using the micronucleus test. *Genetics and Molecular Research, 7*(2), 433-443. doi:10.4238/vol7-2gmr441

SAS 9.4. (2012). *SAS Institue Inc., Cary, NC, USA.*

Schupp, N., Stopper, H., & Heidland, A. (2016). DNA Damage in Chronic Kidney Disease: Evaluation of Clinical Biomarkers. *Oxidative Medicine and Cellular Longevity*, 10. doi:10.1155/2016/3592042

Shebl, F. M., Warren, J. L., Eggers, P. W., & Engels, E. A. (2012). Cancer risk among elderly persons with end-stage renal disease: a population-based case-control study. *Bmc Nephrology, 13*, 8. doi:10.1186/1471-2369-13-65

SNOMED, C. (2014). Starter Guide. *IHTSDO. December*.

Tu, H., Wen, C. P., Tsai, S. P., Chow, W.-H., Wen, C., Ye, Y., . . . Wu, X. (2018). Cancer risk associated with chronic diseases and disease markers: prospective cohort study. *BMJ, 360*. doi:10.1136/bmj.k134

Wong, G., Hayen, A., Chapman, J. R., Webster, A. C., Wang, J. J., Mitchell, P., & Craig, J. C. (2009). Association of CKD and Cancer Risk in Older People. *Journal of the American Society of Nephrology, 20*(6), 1341-1350. doi:10.1681/asn.2008090998

Yoo, K. D., Lee, J. P., Lee, S. M., Park, J. Y., Lee, H., Kim, D. K., . . . Kim, Y. S. (2017). Cancer in Korean patients with end-stage renal disease: A 7-year follow-up. *Plos One, 12*(7), 15. doi:10.1371/journal.pone.0178649

Capstone Experience Reflection:

- My capstone activity is a collaboration with Public Health Informatics Lab, UNMC DRC – II. At Public Health Informatics Lab, Dr. Campbell and team maintained a clinical data warehouse called i2b2 (Informatics for integrating Biology & the Bedside), based on the de-identified patient electronic medical records who received care services at Nebraska Medicine.
- The organization is multi-disciplinary, combination of health care and data science experts to maintain and execute queries to develop evidence-based medicine insights.
- The capstone opportunity is my first-hand experience to handle electronic health record information at a larger volume.
- The tasks related to the capstone allowed me to gain detailed understanding on SNOMED CT coding and cross mapping to the ICD codes.
- It's an intense exercise for me to use SQL and data mining strategies.
- The planned activity trained me to be a liaison between clinical and data science teams in parallel.
- Additionally, I gained experience on critical thinking related to epidemiological and statistical application in real time research.

Acknowledgements

I Would like to thank Dr. Campbell, for trusting my skillset and providing me an opportunity to collaborate for the capstone project.

I thank Dr. Minhas and Dr. Haynatzki, for extended support in developing the research framework and validating the results.

I thank Dr. Apar Ganti and Dr. Ketki Tendulkar, for guiding me through the clinical segments of the project.

I thank Yeshwanth Reddy and Jessica Clark for supporting me through providing data extracts and troubleshooting in data mining queries.

Additionally, I thank Jillian Koons for scheduling meetings with Dr. Campbell.