

### Problem Definition

Chronic kidney disease (CKD) gradually inhibits the kidney's filtration process, causing an accumulation of fluid and waste in the body.

A 2016 article by New Mexico Epidemiology found that, in 2014, there were 28,473 (14.6%) hospitalizations for chronic kidney disease in New Mexico, a 6.0% increase from the number of CKD hospitalizations in 2013 (1). Furthermore, 15% of individuals across the United States struggle with chronic kidney disease, which may increase the risk of malignancy, or cancer (2)

Previous studies have postulated the association of CKD with kidney and genitourinary cancers and found that CKD is a risk factor for mortality in cancer patients (3). Studies have also found that exposure to heavy metals (such as cadmium, lead, and arsenic) can result in CKD and, concurrently, cancer (4). Our question is as follows: How are cancer, chronic kidney disease, and heavy metals exposure correlated? Our problem lies in the fact that there are seldom any computational studies or experiments that have analyzed the statistical correlation between the aforementioned three conditions.

### Computational Plan

Our computational plan to solve this problem is multifaceted. Firstly, we will analyze the National Health and Nutrition Examination Survey (NHANES) to isolate the variables that are specifically related to cancer incidence. The NHANES has various questionnaires, datasets, and documentation about different surveys. At the time of our study, the most complete NHANES dataset was the 2017-2018 one, likely because the coronavirus pandemic restricted the collection

of subsequent data. After isolating pertinent NHANES variables, we will isolate features of interest related to the hypothesis and combine datasets, create graphics of cancer vs. non-cancer patient data, calculate statistics, determine correlations using “seaborn” on Python, create predictive models using Scikit-Learn, and consult ChatGPT for help throughout the process.

### Current Progress

Thus far, we have analyzed the NHANES 2017-2018 questionnaire and laboratory data. The questionnaires are mainly comprised of yes-and-no questions, while the laboratory data has specific values measuring biological markers of different diseases per survey respondent. The information in these datasets is based on a “SEQN” number, which correlates with an individual survey respondent. We are currently researching different biological markers for cancer, chronic kidney disease, and heavy metal exposure. The NHANES datasets do not explicitly stipulate whether respondents were diagnosed with CKD. As such, from our literature searches, we have found that biological markers like albuminuria (the increased presence of albumin, a family of globular proteins, in urine) and a reduced glomerular filtration rate (GFR) are related to CKD (5).

We have also discovered that, in the NHANES dataset, different combinations of letters align with specific biological markers. For instance, “UM\_J” correlates with the classification “Metals - Urine.” We will use the existing NHANES dictionaries to guide our research and understanding of different abbreviations.

### Expected Results

Regarding artificial intelligence, we are using ChatGPT as a tool for teaching programming and computational analysis. We expect to find that ChatGPT enables rapid analysis of the NHANES dataset in conjunction with Python Programming and machine

learning. We also expect to see that chronic kidney disease and heavy metals exposure will be correlated with a higher cancer incidence in the 2017-2018 NHANES dataset. We hope to create a correlation matrix using Python and ChatGPT, one that reveals how positively or negatively correlated certain biological markers are as they pertain to cancer, heavy metals exposure, and chronic kidney disease.

### Citations

1. Dirymer, V. (2016). *Chronic Kidney Disease in Persons with Multiple Chronic Diseases, NM, 2014*. New Mexico Epidemiology - nmhealth.org.  
<https://www.nmhealth.org/data/view/report/1925/>
2. Tendulkar, K. K., Cope, B., Dong, J., Plumb, T. J., Campbell, W. S., & Ganti, A. K. (2022). Risk of malignancy in patients with chronic kidney disease. *PloS one*, 17(8), e0272910. <https://doi.org/10.1371/journal.pone.0272910>
3. Guo, Kanglin, et al. “Association between Chronic Kidney Disease and Cancer Including the Mortality of Cancer Patients: National Health and Nutrition Examination Survey 1999-2014.” *American Journal of Translational Research*, vol. 14, no. 4, 2022, pp. 2356–66.
4. Jalili, C., Kazemi, M., Cheng, H., Mohammadi, H., Babaei, A., Taheri, E., & Moradi, S. (2021). Associations between exposure to heavy metals and the risk of chronic kidney disease: A systematic review and meta-analysis. *Critical Reviews in Toxicology*, 1–30.  
<https://doi.org/10.1080/10408444.2021.1891196>
5. Lopez-Giacoman, S., & Madero, M. (2015). Biomarkers in chronic kidney disease, from kidney function to kidney damage. *World journal of nephrology*, 4(1), 57–73.  
<https://doi.org/10.5527/wjn.v4.i1.57>