**Proposal:** Investigate links between confirmed contaminant exposure and death by chronic diseases

**Hypothesis:** Contaminant exposure is a significant contributor to death by chronic disease.

**Context:**

Chronic disease is a significant problem in the United States. It was documented as the cause of death for more than 65% of deaths among adults as of 2019, before the Covid-19 pandemic.

(References: <https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-08-508.pdf> and

<https://www.cdc.gov/nchs/products/databriefs/db395.htm>).

It's estimated that in 2019, approximately 133 million Americans were living with one or more chronic diseases. This figure represents more than 40% of the total population of the United States at that time, and as of 2019, a rough estimate for the annual US healthcare costs for managing chronic conditions in patients with chronic diseases was around $3.31 trillion.

(Reference: <https://meps.ahrq.gov/data_files/publications/st539/stat539.shtml>)

**Criteria for success:**

Links between contaminants and chronic diseases have been demonstrated in various cases. To be successful, this examination will build a model-training data set to predict death by chronic disease from confirmed contaminant exposure, using the following steps:

* Identify and quantify contaminant exposure
* Identify and quantify chronic disease status at the time of survey
* Evaluate relationship between contaminants and chronic disease at the time of survey
* Evaluate relationship between contaminants and chronic disease as cause of death

**Scope of solution space:**

This analysis focuses on contaminants as a suspected cause of death by chronic disease.

Although the NHANES data set also provides data on other factors that affect health, such as diet, the other factors are outside the scope of this analysis.

**Constraints:**

* Patient data de-identification limits available demographic information. No location data is available, for example.
* There are no health status updates between the initial NHANES data and the vital records update of deaths.
* Data will need to be wrangled to a useful state by joining tables, applying labels, etc.

**Stakeholders:**

Many individuals and organizations have an incentive to reduce the incidence of chronic diseases.With greater knowledge of contributors to chronic disease, individuals, families, and communities can make choices that improve their lives. Government policymakers, healthcare providers, and insurers can share information and take action to reduce contributors to chronic disease in communities and workplaces.

**Data sources:**

Public datasets available from CDC NHANES “National Health and Nutrition Examination Survey”

“NHANES is unique in that it combines personal interviews with standardized physical examinations and laboratory tests.. The purpose of NHANES is to collect data about the health, nutritional status, and health behaviors of the noninstitutionalized civilian resident population of the United States.” <https://www.cdc.gov/nchs/data/series/sr_01/sr01_056.pdf>

Two survey cycles are combined in this analysis to ensure enough patient records for training a model: 1999 - 2000 and 2001 - 2002 [https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/](https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=1999)