COSC5510 - Project 2 - DBMS Application - US Healthcare Crisis

Name: Chao-Chin, Liu, Avery, Bai

Introduction:

NIH's Biggest Health Challenges - Opioid Addiction

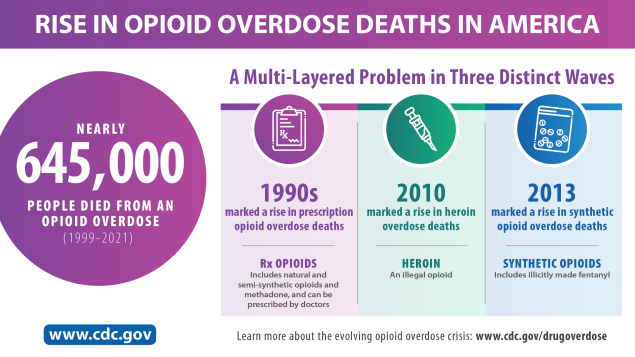
The understanding of addiction has evolved significantly from being viewed as merely a failure of willpower to a recognized chronic disease that alters the brain's structure and function. Substance abuse, especially of opioids—including prescription painkillers, heroin, and synthetics like fentanyl—has escalated into a national crisis, deeply impacting communities across the United States, including military personnel and veterans[[1]](#footnote-0)[[2]](#footnote-1). The economic toll of prescription opioid misuse alone on the U.S. economy surpasses $504 billion annually[[3]](#footnote-2). Groundbreaking NIH-supported research has led to the development of lifesaving interventions, such as NARCAN® nasal spray for reversing opioid overdoses, and the establishment of comprehensive strategies to prevent and treat opioid addiction[[4]](#footnote-3).

The opioid epidemic is multifaceted, not only presenting challenges in addiction treatment but also in managing chronic pain without resorting to addictive substances. The NIH, in collaboration with the FDA and the pharmaceutical industry, is at the forefront of research to find safe and non-addictive pain management options, including exploring the effectiveness of alternative therapies like yoga, acupuncture, and behavioral therapy[[5]](#footnote-4).

The escalation of opioid overdose deaths from 1999 to 2021, claiming nearly 645,000 lives, unfolds in three waves: the surge in prescription opioid deaths in the late 1990s; the spike in heroin-related deaths in 2010; and the alarming rise in deaths involving synthetic opioids, especially illicitly manufactured fentanyl, from 2013 onwards. The complexity of the opioid market, with fentanyl found in combination with other illicit drugs, complicates the crisis further.

[[6]](#footnote-5)

In addressing the opioid epidemic, a critical area of focus is the incidence of nonfatal overdoses—understanding their prevalence, the demographic and geographical patterns they follow, and their impact on health systems and communities. This analysis aims to delve into nonfatal overdose data to shed light on the extent of opioid misuse, identify the states most affected, examine the age groups at highest risk, and ultimately contribute to the development of targeted interventions to mitigate this ongoing public health challenge.



Dataset:

| Nonfatal Overdose Syndromic Surveillance Data | <https://www.cdc.gov/nssp/overview.html> |
| --- | --- |
| Nonfatal Overdose Emergency Department and Inpatient Hospitalization Discharge Data | <https://www.cdc.gov/drugoverdose/nonfatal/dose/about.html> |
| Federal Information Processing System (FIPS) Codes for States and Counties | <https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt> |
| State Population Totals: 2010-2019 | <https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html> |
| State Population Totals and Components of Change: 2020-2023 | <https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html> |

Important caveats to when interpreting this data:

* **Some data may be missing or incomplete.** Data not available by the reporting deadline may not ever be submitted, as data are typically considered final at submission.
* **Reporting facilities and the data they report can change over time.** States may receive data from new facilities, and the data they report could change over time. The average percent of ED visits/inpatient hospitalizations currently captured from states participating in DOSE discharge data sharing is 90%.
* **These overdoses may not be confirmed by toxicological testing.** These data may not be determined by toxicological testing, which is often limited in ED or hospital settings. Additionally, ED and inpatient hospitalization discharge data are collected for administrative/billing purposes; thus, surveillance for drug overdoses using these data may not accurately reflect the true overdose burden.
* **Data are included for overdoses of unintentional and undetermined intents.** Only discharge diagnosis codes for overdoses of unintentional and undetermined intent are included in the data presented on this dashboard.
* **Overdose visit numbers are not mutually exclusive** but rather reflect nesting of drug categories: numbers of opioid-, heroin-, and stimulant-involved overdose visits are included in the numbers of all drug overdose visits; suspected heroin-involved overdose visits are included in the numbers of opioid-involved overdose visits; and some overdose visits involved multiple substances (e.g., a given overdose ED visit could have involved both opioids and stimulants).
* **Rates beginning in 2021 may be different from prior years.** The U.S. Census Bureau instituted new methodology to calculate population estimates beginning with 2021 data. The new methodology, referred to as differential privacy, ensures that data from individuals and individual households remain confidential.
* **For some states, counts and rates for all drug overdoses during October 2021–September 2022 are not directly comparable to prior months and years.** Data from January 2018–September 2021 included discharge diagnosis codes for cannabis poisonings in the all drug overdose category. In October 2021, the prior cannabis poisoning diagnosis code (T40.7X) was retired, and new codes (T40.71, T40.72) for cannabis poisoning and synthetic cannabinoid poisoning were implemented ([FY2022 ICD-10-CM](https://www.cms.gov/medicare/icd-10/2022-icd-10-cm)). However, these new codes were not queried in DOSE ED visit and inpatient hospitalization discharge data until October 2022. Thus, there may be underestimates in the counts and rates of all drug overdoses during October 2021–September 2022.

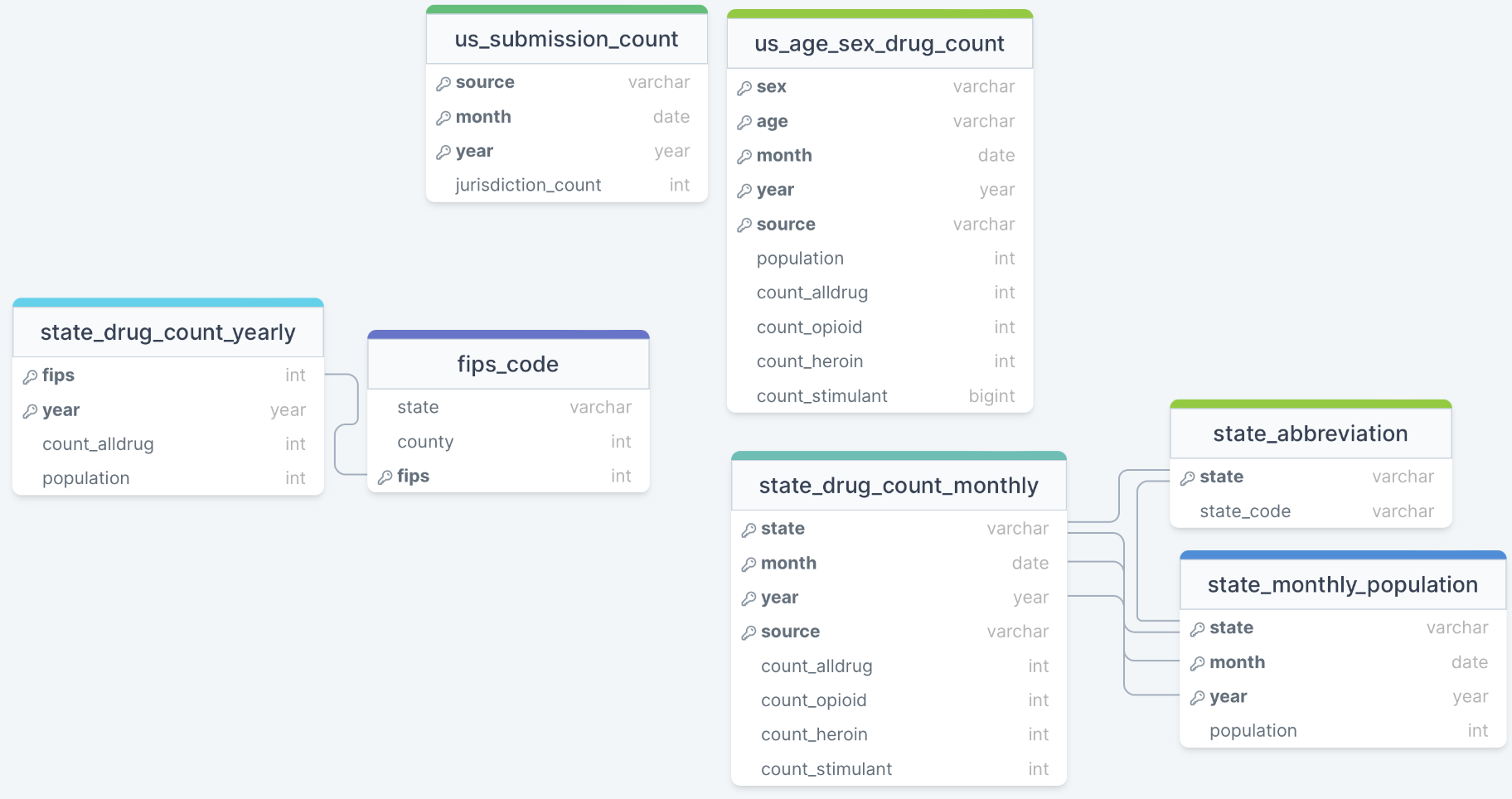
Data cleaning process with Excel

1. Remove duplicate or irrelevant observations.
2. Removing yearly data - We can calulate yearly data ourself
3. Fix structural errors
4. Handle missing data.
5. When the statement mentions that counts are "suppressed when based on 1-9 overdoses" and rates are "suppressed when based on 1-19 overdoses," it means that detailed data or specific rates of overdoses are not disclosed or made publicly available if they are derived from a very small number of cases (between 1 to 9 for counts and 1 to 19 for rates). -> Drop the data

Normalization Process:

* US state submission counts
  + primary key: dataset, month, year
  + 1NF: unique rows
  + 2NF: jurisdiction\_count fully depends on the dataset, month and year.
  + 3NF: no transitive partial dependent
  + Notes: should we change the attribute name dataset to sth else? -> Rename to "source"
* State counts rates\_monthly
  + primary key: State name, month, year, dataset
  + 1NF: unique rows
  + 2NF: count\_alldrug, count\_opioid, count\_haroin, count\_stimulant depend on state, month, year, and dataset
  + 3NF: no transitive partial dependent
  + Notes: should probably change the name of dataset to something else -> Rename to "source"
* monthly\_population
  + primary key: state, year, month
  + 1NF: unique rows
  + 2NF: population depends on state, month, year
  + 3NF: no transitive partial dependent
* yearly\_population
  + primary key: state, year, month
  + 1NF: unique rows
  + 2NF: population depends on state, month, year
  + 3NF: no transitive partial dependent
  + notes: do we need both yearly and monthly population as two tables? Is it just the summation of the monthly population? -> Drop
* State counts rates\_monthly
  + primary key: State name, month, year, dataset
  + 1NF: unique rows
  + 2NF: count\_alldrug, count\_opioid, count\_haroin, count\_stimulant depend on state, month, year, and dataset
  + 3NF: no transitive partial dependent
  + Notes: do we need both yearly and monthly counts as two tables? Is it just the summation of the monthly counts? -> Drop
* County Counts Rates\_yearly
  + Notes: 1NF: fids:15005, year(there is one row combined all years from 2018 to 2022), drug use and rate indicates “not available”. I dropped the rate\_alldrug to keep it at 3NF. I also deleted the dataset column since the whole table only recorded ED data. Is population necessary in this table? -> Drop
  + primary key: fips and year
  + other than the above issue& some fixations:
  + 1NF: unique rows
  + 2NF: count\_alldrugs and population fully depends on fips, year.
  + 3NF: no transitive partial dependence.
* fips\_code
  + primary key: fips code
  + 1NF: unique rows
  + 2NF: state and county name depend on fips.
  + 3NF: there is no transitional partial dependence between state and county.
* County Counts Rates\_5years
  + note: delete year, dataset and rate\_alldrug since this dataset only recorded ED data between 2018 to 2022
  + primary key: fips code
  + 1NF: unique rows
  + 2NF: count\_alldrug depends on fips code
  + 3NF: no transition partial dependence
* Overall by sex and age:
  + notes: i deleted state, because this dataset did not record data in state level, instead, it used overall level. Do we need population? we do have a monthly population of all states from 2018 to 2022. -> Drop
  + primary key: sex, age, month, year and dataset.
  + 1NF: unique rows
  + 2NF: count\_alldrug, count\_opioid, count\_haroin and stimulant depend on sex, age range, month, year, dataset.
  + 3NF: no transitional partial dependence between count\_alldrug, count\_opioid, count\_haroin and count\_stimulant.

Scheme:



1. Federal Communications Commission. (n.d.). Focus on opioids - Connect2Health. FCC. https://www.fcc.gov/reports-research/maps/connect2health/focus-on-opioids.htm [↑](#footnote-ref-0)
2. Compton, W. M., & Jones, C. M. (2019). Epidemiology of the U.S. opioid crisis: the importance of the vector. Annals of the New York Academy of Sciences, 1451(1), 130–143. https://doi.org/10.1111/nyas.14209 [↑](#footnote-ref-1)
3. U.S. Representative Don Beyer. (2022, September 28). JEC analysis finds opioid epidemic cost U.S. nearly $1.5 trillion in 2020. https://beyer.house.gov/news/documentsingle.aspx?DocumentID=5684 [↑](#footnote-ref-2)
4. Lewis, C. R., Vo, H. T., & Fishman, M. (2017). Intranasal naloxone and related strategies for opioid overdose intervention by nonmedical personnel: a review. Substance abuse and rehabilitation, 8, 79–95. https://doi.org/10.2147/SAR.S101700 [↑](#footnote-ref-3)
5. National Institutes of Health. (2024, January 31). NIH analysis reveals a significant rise in use of complementary health approaches, especially for pain management. https://www.nih.gov/news-events/news-releases/nih-analysis-reveals-significant-rise-use-complementary-health-approaches-especially-pain-management [↑](#footnote-ref-4)
6. Centers for Disease Control and Prevention. (n.d.). Understanding the opioid overdose epidemic, from https://www.cdc.gov/opioids/basics/epidemic.html [↑](#footnote-ref-5)