Statistical Study of Drug Overdose Fatalities: Analyzing the Influence of Multiple Contributing Factors

Shafin Ula[a,b,]

[a]Math and Science Academy, Dulles High School, Sugar Land, Texas, USA
[b]Department of Mechanical & Aerospace Engineering, University of Houston, Houston, Texas, USA

Abstract

The presence of various different drugs in society, although some with good intent, contributes greatly to the number of deaths in the United States. Drug overdose deaths, the most common form of drug death with over 100,000 deaths, has been recorded as steadily increasing in the United States every year with many states reporting up to 20% increases in number of deaths from previous years. More Research is needed on this compounding issue from the social and statistical perspectives to help mitigate this societal challenge. To aid these efforts, this study provides a thorough overview of many crucial contributing factors to the issue of drug overdose deaths through a statistical and data visualization perspective and serves baseline for further research and data analysis on this problem.

Keywords: Statistics, Drug, Data Analytics, Data Visualization

1. Introduction and Purpose

The use of drugs presents a complex balance of benefits and risks. While certain drugs offer therapeutic relief from pain, enhance energy, and alleviate stress, improper or excessive use can result in serious adverse effects, including death.

According to data from the National Institute on Drug Abuse, over 100,000 drug overdose deaths occurred in the U.S. in 2021, marking a significant 16 percent increase from 2020 and representing the highest number of overdose fatalities recorded in a single year [1]. These deaths are linked to a wide range of substances, including opioids, heroin, and sedatives. The escalating rate of overdose deaths has prompted widespread public concern, as the crisis shows no signs of abating. To address these concerns, it is essential to conduct a thorough analysis of data relating to overdose deaths to identify underlying trends and contribute towards developing effective solutions to mitigate this growing public health issue.

This study and analysis on drug overdose is important as the issue of overdose deaths impacts many families and communities in the United States [2]. This also has a huge toll on the U.S economy "through lost productivity, incarceration, and healthcare costs" due to drug treatment. In 2017, the total estimated economic cost of opioid use disorder and overdose deaths reached $1.02 trillion. This includes $100 billion in lost productivity, $34.8 billion in healthcare costs, and $14.8 billion in criminal justice costs [3].

In this study, data from various reliable sources will be collected and analyzed to understand the background and basis of the rise of drug overdose deaths in the United States. It will be determined how big of an impact the demographics and geographical aspects that may influence or play a part in this ongoing issue.

The following categories will be taken account in regard to drug overdose deaths:

1. Year
2. Gender
3. Race
4. Age

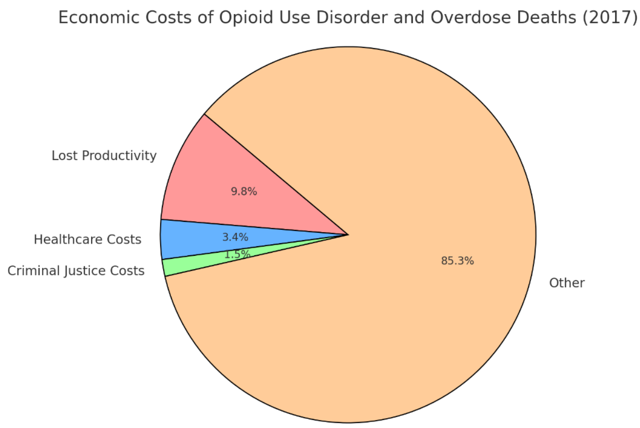


Figure 1: Percentage economic costs of the whole of seemingly insignificant sources

2. Data Collection and Structuring Methods

Secondary data collected from previous reliable data and statistics sources were utilized in this study. The datasets were collected from a dataset by the Office of the Chief Medical Examiner and Center for Disease Control and Prevention (CDC).

Data and statistics from the CDC WONDER database was utilized as a result. The year of the overdose death was limited to the years between 2013-2021 to ensure that the data would be relevant and would align to the purposes of the study.

The data from these years from various sources were extracted and sorted through systematic data collection and analysis methods through the use of APIs and PHP developed code. A snippet of such code for collecting the data to be utilized is shown below:

```php
//Shafin Ula
//CDC Wonder Drug Death Data Extraction
<?php
$ch = curl_init();
$url = "https://wonder.cdc.gov/controller/saved/D158/D390F073";
curl_setopt($ch, CURLOPT_URL, $url);
curl_setopt($ch, CURLOPT_POST, 1);


$data = http_build_query(array(
    'B_1' => 'D76.V2-level1', // Gender
    'B_2' => 'D76.V2-level2', // Race
    'B_3' => 'D76.V2-level3', // Age
    'B_4' => 'D76.V2-level4', // Year

));


curl_setopt($ch, CURLOPT_POSTFIELDS, $data);
curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);


$response = curl_exec($ch);
curl_close($ch);


if ($response) {
    file_put_contents('wonder_data.csv', $response);


    if (($handle = fopen("wonder_data.csv", "r")) !== FALSE) {
        while (($data = fgetcsv($handle, 1000, ",")) !== FALSE) {

            $year = $data[0];
            $gender = $data[1];
            $race = $data[2];
            $age = $data[3];


            echo "Year: $year, Gender: $gender, Race: $race, Age: $age\n";
        }
        fclose($handle);
    }
} else {
    echo "Failed to retrieve data from CDC WONDER";
}
?>
```

Figure 2: Data Extraction from CDC Wonder Database of Drug Overdose Deaths 2010-2022

2.1. *Sorting and Structuring*

After the collection and extraction, the CDC Wonder Data had to be sorted to fit the purpose of this study based on set conditionals. It then had to be structured to be more effective in the data analytics and output process.

To sort the extracted secondary data collected from the CDC Wonder Data, additional stages were taken to ensure that all the unnecessary data categories and years until 2013 were completely removed from the extracted data. This happened through exporting and converting the extracted data to csv format and then sorting and structuring the data.

PHP developed code that shows the process of removing unnecessary data and structuring is shown below:

```php
//Shafin Ula
//CDC Wonder Drug Death Data Sort and Structuring
<?php
// Define the path to your CSV file
$inputFile = 'path_to_your_extracted_data.csv';
$outputFile = 'path_to_your_filtered_data.csv';

// Open the CSV file for reading
if (($handle = fopen($inputFile, 'r')) !== FALSE) {
    // Read the header line
    $header = fgetcsv($handle);

    // Find the indices of the required columns
    $yearIndex = array_search('Year', $header);
    $genderIndex = array_search('Gender', $header);
    $raceIndex = array_search('Race', $header);
    $ageIndex = array_search('Age', $header);

    // Check if all required columns are present
    if ($yearIndex === FALSE || $genderIndex === FALSE || $raceIndex === FALSE || $ageIndex === FALSE) {
        die('Required columns are missing in the CSV file.');
    }

    // Open the output CSV file for writing
    $outputHandle = fopen($outputFile, 'w');

    // Write the header to the output file
    fputcsv($outputHandle, ['Year', 'Gender', 'Race', 'Age']);

    // Process each row in the CSV file
    while (($data = fgetcsv($handle)) !== FALSE) {
        // Filter rows based on year and presence of required columns
        if (isset($data[$yearIndex], $data[$genderIndex], $data[$raceIndex], $data[$ageIndex])) {
            $year = (int)$data[$yearIndex];
            if ($year < 2011 || $year > 2012) {
                // Write the row to the output file
                fputcsv($outputHandle, [
                    $data[$yearIndex],
                    $data[$genderIndex],
                    $data[$raceIndex],
                    $data[$ageIndex]
                ]);
            }
        }
    }

    // Close the file handles
    fclose($handle);
    fclose($outputHandle);

    echo "Data has been filtered and saved to '$outputFile'.";
} else {
    die('Unable to open the input CSV file.');
}
?>
```

Figure 3: Removal and Structuring of CDC Wonder Data

2.2. *Data Visualization*

With the categorized and structured data in csv format, excel graphs can be used as the most efficient method to visualize the data. Based of the circumstances of the category, a table, pie chart, or a bar graph is utilized to provide percentage distributions, counts, or percentage changes based on the specific aspect of the category that need to be emphasized.

3. Data Visualization and Results

The 4 categories (Year, Gender, Race, Age) researched in this study is visualized for interpretation and conclusion purposes

3.1. *Year*

The following visual in **figure 4** shows the number of incidents of drug overdose deaths every year using the data sorted for the years 2013-2021.

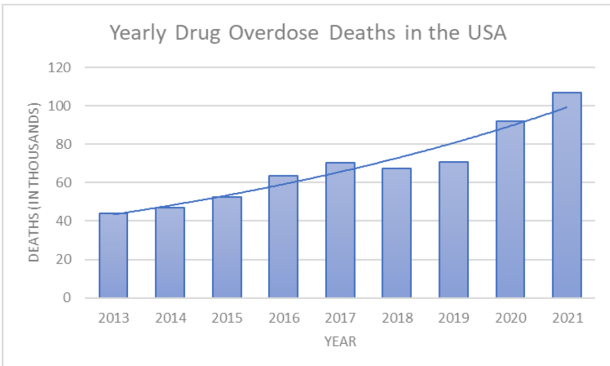


Figure 4: Histogram of Drug Overdose Deaths by the Year (2013-2021)

3.2. Gender

Using data provided by the CDC, we will determine if there is a relationship between Drug overdose deaths and gender. There is a disproportionate amount of males vs females in the USA, so counts of Deaths per 100,000 people of one gender will be utilized for our purposes.

A table shown in **Figure 5** has been created to display these counts:

Gender vs Deaths from Overdose per 100,000 popualtion from 2013-2021

	2013	2014	2015	2016	2017	2018	2019	2020	2021
Female	10.6	11.1	11.8	13.4	14.4	13.6	13.7	17.1	19.6
Male	17	18.3	20.8	26.2	29.1	27.9	29.6	39.5	45.1

Figure 5: Contingency Table of Gender vs Deaths from Overdose every year (2013-2021)

A Histogram shown in **Figure 6** is developed from the table to provide a clearer visual on the statistical difference of the genders without any bias or disproportionate representation:

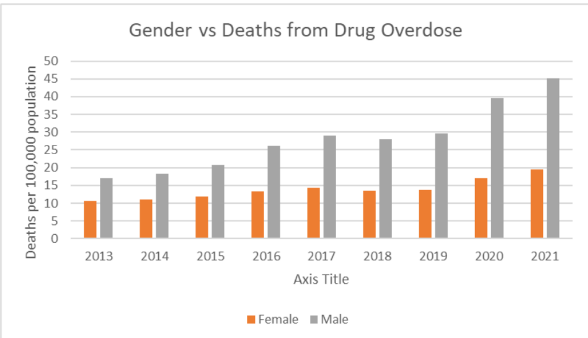


Figure 6: Cluster Column Histogram of Gender vs Deaths from Overdose every year (2013-2021)

The data visuals in **Figures 5** and **6** created from our data extracted can be used to show the difference between drug overdose deaths between Male and Female per 100,000 males or females. For example in 2021, there were about 45 male deaths from drug overdose per every 100,000 male individuals and around 20 female deaths from drug overdose for every 100,000 female individuals. Besides 2021, for every year, there is evidence that the male deaths from drug overdose per 100,000 individuals is larger than the female deaths per 100,000 individuals. Another trend given is that the male deaths per 100,000 increase every year while the female deaths stay relatively consistent throughout the years. There is a increasing gap throughout the years as there is a reported 70% more drug overdose death counts for males compared to females in 2013. On the other hand, there is a near 125% more drug overdose death count for males compared to females in 2021, indicating a upward trend.

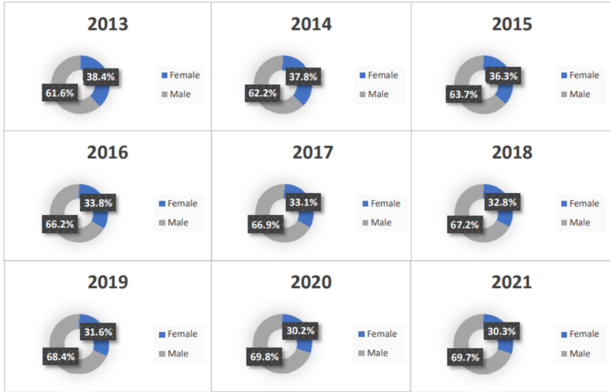


Figure 7: Donut Visual of Gender vs Proportion of Deaths from Overdose by year (2013-2021)

This visual in **Figure 7** was generated to effectively compare the percentage of female deaths from overdose to the male deaths along the years. The proportion of male deaths from overdose accrued. The only exception being from 2020 to 2021, where the proportion of male deaths decreased slightly. This general proportion difference between genders may be due to a combination of biological (e.g., men may have a greater vulnerability to the toxicity of drugs than women), behavioral (e.g. men may use these drugs in a riskier way than women), as well as other social and gender-related factors as proved in previous research studies [4].

3.3. Race

It is known that many race based groups use drugs and die drug related deaths more often than other race groups. Since there is a disproportionate amount of people in each race group in the USA, data will be based on drug overdose deaths per 100,000 people.

The data visual in **figure 8** gives some introductory insight into different race groups and the number of overdose deaths from each group per 100,000. We will be comparing results from 2018-2021, as certain race groups don't have any available reliable data before 2018.

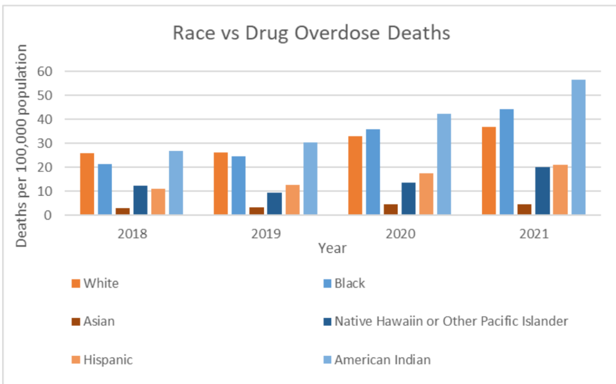
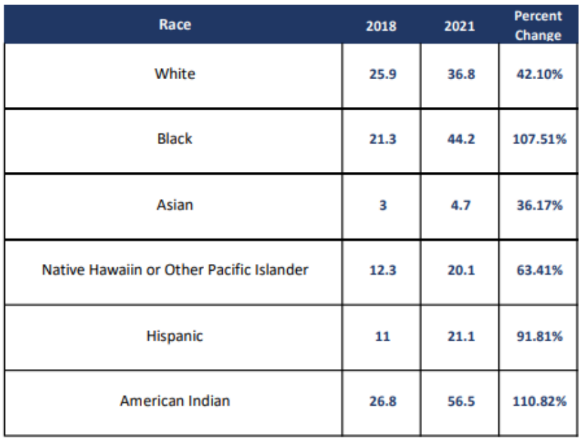


Figure 8: Cluster Column Histogram of Race vs Deaths from Drug Overdose by Year (2018-2021)

Figure 5 illustrates the change in death by drug overdose per every 100,000 population of each race group from 2018-2021. As noted before, data from before 2018 was not able to be compared as data was combined for Asians and Native Hawaiian or Other Pacific Islanders, making it unusable to compare groups with each other. A general trend observed is that the drug overdose death cases rise every year for every single race group. This suggests that this societal issue of drug overdose deaths is important to address to everyone in society regardless of race group, as the death rate is increasing for every race over the years. Additionally, we can see that the yearly rate of increase of the number of deaths by drug overdose differs by race group. The rate of increase for Asians from 2018 to 2021 is 56.7%, while it is 110.82% for American Indians. Researchers may use this info on identify the races with the greatest percentage increase so they can focus on curbing the rising rates for that specific race.

A detailed table with the rate of increase from 2018 to 2021 is shown below.



Race	2018	2021	Percent Change
White	25.9	36.8	42.10%
Black	21.3	44.2	107.51%
Asian	3	4.7	36.17%
Native Hawaiin or Other Pacific Islander	12.3	20.1	63.41%
Hispanic	11	21.1	91.81%
American Indian	26.8	56.5	110.82%

Figure 9: Table displaying the Percent Change of Deaths for Different Races from 2018 to 2021

As stated earlier, The rate of percent increase from 2018 to 2021 for Black and American Indian are over 100% while the rate for Asians and Whites are under 50% indicating a clear non-independence between proportionate drug overdose death counts and Race. A further in depth statistical analysis proves this non-independence.

3.3.1. Statistical Analysis

A Chi-Square test for independence can be used to confirm that independence doesn't seem to exist between Race groups and the counts of drug overdose deaths per 100,000 of that race.

Calculating the Expected values ($E_{i,j}$)

$$E_{ij} = \frac{(\text{Row Total}_i) \times (\text{Column Total}_j)}{\text{Overall Total}}$$

Row Total Calculations:

- White: 25.9 + 36.8 = **62.7**
- Black: 21.3 + 44.2 = **65.5**
- Asian: 3.0 + 4.7 = **7.7**
- Native Hawaiian: 12.3 + 20.1 = **32.4**
- Hispanic: 11.0 + 21.1 = **32.1**
- American Indian: 26.8 + 56.5 = **83.3**

Column Total Calculations:

- 2018: 25.9 + 21.3 + 3.0 + 12.3 + 11.0 + 26.8 = **100.3**
- 2021: 36.8 + 44.2 + 4.7 + 20.1 + 21.1 + 56.5 = **183.4**

Overall Total:

- 100.3 + 183.4 = **283.7**

$\therefore$

White (2018): $E = \frac{62.7\times100.3}{283.7} = \mathbf{22.17}$

White (2021): $E = \frac{62.7\times183.4}{283.7} = \mathbf{40.53}$

Black (2018): $E = \frac{65.5\times100.3}{283.7} = \mathbf{23.16}$

Black (2021): $E = \frac{65.5\times183.4}{283.7} = \mathbf{42.34}$

Asian (2018): $E = \frac{7.7\times100.3}{283.7} = \mathbf{2.72}$

Asian (2021): $E = \frac{7.7\times183.4}{283.7} = \mathbf{4.98}$

Native Hawaiian (2018): $E = \frac{32.4\times100.3}{283.7} = \mathbf{11.45}$

Native Hawaiian (2021): $E = \frac{32.4\times183.4}{283.7} = \mathbf{20.95}$

Hispanic (2018): $E = \frac{32.1\times100.3}{283.7} = \mathbf{11.35}$

Hispanic (2021): $E = \frac{32.1\times183.4}{283.7} = \mathbf{20.75}$

American Indian (2018): $E = \frac{83.3\times100.3}{283.7} = \mathbf{29.45}$

American Indian (2021): $E = \frac{83.3\times183.4}{283.7} = \mathbf{53.85}$

where E is the expected values of that cell

Expected values:

Race	2018	2021
White	22.17	40.53
Black	23.16	42.34
Asian	2.72	4.98
Native Hawaiian or Other Pacific Islander	11.45	20.95
Hispanic	11.35	20.75
American Indian	29.45	53.85

Chi-square Statistic Calculation (χ^2)

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where:

- O_{ij} is the observed value in the i-th row and j-th column.
- E_{ij} is the expected value in the i-th row and j-th column.

$$\text{For White (2018): } \frac{(25.9 - 22.17)^2}{22.17} = \mathbf{0.619}$$
$$\text{For White (2021): } \frac{(36.8 - 40.53)^2}{40.53} = \mathbf{0.341}$$
$$\text{For Black (2018): } \frac{(21.3 - 23.16)^2}{23.16} = \mathbf{0.145}$$
$$\text{For Black (2021): } \frac{(44.2 - 42.34)^2}{42.34} = \mathbf{0.076}$$
$$\text{For Asian (2018): } \frac{(3.0 - 2.72)^2}{2.72} = \mathbf{0.031}$$
$$\text{For Asian (2021): } \frac{(4.7 - 4.98)^2}{4.98} = \mathbf{0.012}$$
$$\text{For Native Hawaiian (2018): } \frac{(12.3 - 11.45)^2}{11.45} = \mathbf{0.068}$$
$$\text{For Native Hawaiian (2021): } \frac{(20.1 - 20.95)^2}{20.95} = \mathbf{0.037}$$
$$\text{For Hispanic (2018): } \frac{(11.0 - 11.35)^2}{11.35} = \mathbf{0.011}$$
$$\text{For Hispanic (2021): } \frac{(21.1 - 20.75)^2}{20.75} = \mathbf{0.006}$$
$$\text{For American Indian (2018): } \frac{(26.8 - 29.45)^2}{29.45} = \mathbf{0.241}$$
$$\text{For American Indian (2021): } \frac{(56.5 - 53.85)^2}{53.85} = \mathbf{0.124}$$

$$\chi^2 = 0.619 + 0.341 + 0.145 + 0.076 + 0.031 + 0.012 + 0.068 + 0.037 + 0.011 + 0.006 + 0.241 + 0.124 = \mathbf{1.73}$$

P-Value and Conclusion

$$p = P(\chi^2 \geq 1.73 \mid df = 5) \approx \mathbf{0.885}$$

With a p-value of approximately 0.885, higher than the null value or significance value of 0.05, we fail to reject the null hypothesis. This suggests that there is in fact no significant evidence of independence between race and the proportionate drug overdose death counts.

This finding suggests that there is a possible dependence on the number of drug overdose deaths based on Race. Thus, some races are statistically disproportionately affected by drug overdose deaths.

3.4. *Age*

Age is another important variable to identify groups with regard to drug overdose deaths. There are certain age groups where drug overdose deaths are more prevalent regardless of the race, gender, and other factors relating to that person. Additionally, since there is a disproportionate amount of people in each age group in the USA, we will be looking at drug overdose deaths per every 100,000 people in each age group.

For the purposes of this study, we will sort the ages into the following groups:

- under 25
- 25–44 years
- 45-64 years
- over 64

Drug Overdose death data from 2013-2019 will be used. This is due to the fact that there is no published data on CDC or any other reliable sources from 2020 onwards.

Age Group	2013	2014	2015	2016	2017	2018	2019
<25 years	8.5	8.8	9.9	12.6	12.8	11	11.5
25-44 years	43.9	48.1	55.2	69.9	77.4	73.8	76.1
45-64 years	46.7	48.5	51.8	60.1	65.7	63.6	67.3
>64 years	14.3	14.6	15.2	15.6	16.8	17.9	19.1

Figure 10: Table displaying the drug overdose deaths per 100,000 people in that age group from 2013 to 2019

Structuring and transforming the data values on the chart, a histogram which shows the drug overdose deaths per every 100,000 people for each age group from 2013-2019 can be created as shown in **Figure 11**. This will be useful in determining which age groups are most in need of concern.

From this cluster column histogram, there is a general increasing trend for the age groups of 25-44 years and 45-64

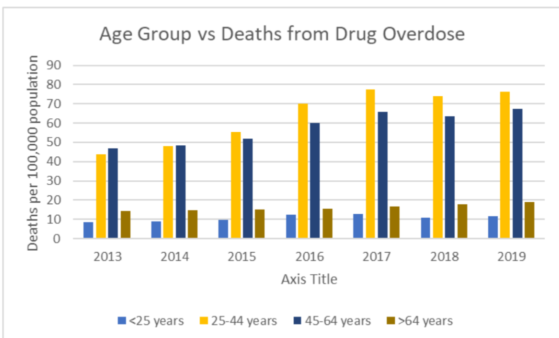


Figure 11: Cluster Column Histogram of Age Group vs Deaths from Drug Overdose from 2013 to 2021

years from 2013-2019. However, for the age groups of under 25 years and over 64 years, the deaths remain quite consistent from 2013-2019. Additionally, there are a significant number of more deaths per 100,000 people for the age groups of 25-44 years and 45-64 than those for under 25 and over 64 years.

4. Conclusion

This study aims to figure out how drug overdose deaths have changed over time and to understand which groups, based on factors such as race, gender, and age, are most affected. Through data analysis, structuring, and visualization, this study provides an analysis of 4 contributing factors towards the nationwide issue of Drug-related overdose deaths. Summary of the different factors are presented below.

4.1. Year

There is evidence that there is a rise in Drug overdose deaths every year. Though there is a small decline in deaths in 2018 and 2019, it is roughly the same as the number of deaths in 2017, so it can be generally ignored for our purposes. In addition, a regression line suggests that there is an increasing trend that is slightly exponential as the years increase.

4.2. Gender

Through our data, it can be concluded that the number of deaths with respect to the proportions of the total population of each gender is increasing as the years go on. We can also see from the models that the percentage of the proportion of male deaths from the whole increased over the years. This indicates an increase in the gender disparity in the number of deaths from drug overdose. This can possibly be explained by behavioral and social differences that exist between the two genders.

4.3. Race

Initially, we can notice the contrast between different race groups of the magnitude of deaths per 100,000 population of that race group. The Whites, Blacks, and American Indians all have the greatest number of deaths compared to the other race groups. Although the number of drug overdose deaths of every race group has increased from 2018 to 2021, Blacks and American Indians have the greatest percent increase of deaths per 100,000 population.

4.4. Age

With regards to age, there is an increase in the number of deaths per 100,000 population for those people in age groups of 25-44 years and 45-64 years from 2013-2019. However, the number of deaths per 100,000 for those age groups are greater than those for the age group for below 25 and over 64 years.

References

1. Kline, D., Hepler, S. A., Krawczyk, N., Rivera-Aguirre, A., Waller, L. A., et al. (2024). A state-level history of opioid overdose deaths in the United States: 1999-2021. *PLOS ONE*, 19(9), e0309938. `https://doi.org/10.1371/journal.pone.0309938`
2. Kneebone, E., et al. (2022). A Nation in Overdose Peril: Pinpointing the Most Impacted Communities and the Local Gaps in Care. *Brookings*, 9 Mar. 2022. `https://www.brookings.edu/articles/pinpointing-opioid-in-most-impacted-communities/`
3. Florence, C., Luo, F., Rice, K. (2021). The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. *Drug and Alcohol Dependence*, 218, 108350. `https://doi.org/10.1016/j.drugalcdep.2020.108350`
4. McHugh, R. K. (2020). The Importance of Studying Sex and Gender Differences in Opioid Misuse. *JAMA Network Open*, 3(12), e2030676. `https://doi.org/10.1001/jamanetworkopen.2020.30676`