Drug Poisoning Deaths in the United States

Kasen Teoh

Introduction

Drug poisoning deaths have surged over the last decade in the United States, contributing to a significant public health crisis. This research aims to explore the patterns and trends in drug poisoning deaths between 1999 and 2016, focusing on the roles of demographic variables such as gender, race, and age. This study intends to utilize the data set of 864 observations to provide robust statistical insights, such as elucidating the significant demographics most related to drug poisoning deaths.

Data Set and Cleaning

The data set utilized in this research primarily contained clean, non-missing data, making it an excellent resource for this analysis. It comprised the variables year, sex, age, race, death count, and population - each providing insightful demographic and statistical details pertinent to the occurrence of drug poisoning deaths in the United States from 1999 to 2016. In addition to these, we introduced a new column, 'id', that contains a unique numerical identifier for each unique combination of sex, age, and race in the data set. This variable is instrumental in simplifying the management and analysis of the data set, allowing for a more structured exploration of the relationship between the predictor variables and drug poisoning deaths.

Analysis and Results

Beginning with exploratory data analysis, [1] presents the average drug poisoning death rates segmented by sex over a period of 17 years. The data consistently show a higher susceptibility among males compared to females across all years. From 1999 to 2014, both genders demonstrated a relatively linear trend in death rates, but an unexpected surge is seen in males during 2015 and 2016. [2] delineates the average death rates partitioned by eight predefined age groups. The groups at the lower and upper extremities of the age spectrum, specifically 0-14, 15-24, 65-74, and 75+ displayed relatively lower death rates. The age groups 35-44 and 45-54 bore the brunt, with the latter surpassing the former in death rates from 2002 onward, despite a higher average death rate in the 35-44 group during 1999-2001. [3] underscores the death rates categorized by race, namely Hispanic, White, and Black. White emerged as the demographic with the highest death rates from 2002 onward, following Black that led from 1999 to 2001. Notably, a peak in death rates was evident in 2006 among males and most age groups [1, 2], but this pattern is only visible in the Black demographic. The empirical summary plot [4] with 95% confidence intervals reveals a gradual widening of the interval over time. The last year, 2016, exhibits an exceptionally large interval, which might be indicative of the largest increase in variance due to growing differences between the groups.

In light of the exploratory plots, a linear time trend emerged as a plausible hypothesis. To evaluate this hypothesis, three distinct models - random intercept (RI), random intercept and slope (RIAS), and continuous autoregressive of order 1 (CAR(1)) - were applied to linear, quadratic, and cubic time trends, respectively. Of the nine models considered, the CAR(1) model displayed the most favorable fit with the lowest AIC and BIC values of 3736.693 and 3760.489 respectively, as outlined in [5]. The CAR(1) model was employed to evaluate the significance of sex, age, and race predictors independently. The p-values for male and female (0.0062 and 0.0203 respectively) suggested significant roles for both sexes [6], with neither showing 0 within two standard errors of the estimate. Age also appeared to be a significant predictor; however, certain age groups (0-14, 15-24, 65-74, 75+) showed p-values above the 0.05 threshold [7], potentially due to lower death rates in these groups, as shown in [2]. Race, on the other hand, did not emerge as a significant predictor, with all p-values exceeding the

0.05 threshold [8], thus it was excluded from subsequent models.

A combined model featuring both sex and age was subsequently developed to explore whether these predictors would exhibit superior performance together rather than individually. This model demonstrated the lowest AIC and BIC among all considered models [9]. Furthermore, log-ratio tests comparing the combined model against individual sex and age models both indicated a higher log-likelihood value for the combined model, with Chi-squared tests yielding exceptionally small p-values [10]. Therefore, it appears that the combined sex and age model provides the best predictive performance.

Incorporating the year variable into the model, only the age group of 15-24, 65-74 and 75+ groups are not significant [11]. Furthermore, it is significantly lower in AIC and BIC values (3675.612 and 3737.361) compared to the model featuring only sex and age (3777.707 and 3834.720) [12]. This is further supported by a higher log-likelihood value for the model inclusive of time, and a notably small p-value, reinforcing the superiority of this model. When adding year and fitting it to various models including RI, RS, RIAS, IN, and ARMA(1,1), the ARMA(1,1) model yielded the lowest AIC and BIC values, although CAR(1) followed closely [13]. A log-ratio test was conducted to verify if the marginal difference in AIC and BIC between these two models could be disregarded. The ARMA(1,1) model returned a higher log-likelihood score than CAR, and the Chi-squared test revealed a nearly zero p-value [14], further solidifying the preferability of the ARMA(1,1) model.

Conclusion and Discussion

The ARMA(1,1) provided the most accurate fit for the linear combination of sex, age, and year, exhibiting the lowest AIC and BIC and the largest log-likelihood score. Race was found not a significant predictor, which may suggest the peak found in the Black demographic may have simply been due to chance. Analysts can further explore forecasting with this ARMA(1, 1) model or with other parameters as well as parametric and non-parametric spectral analysis.

Figures

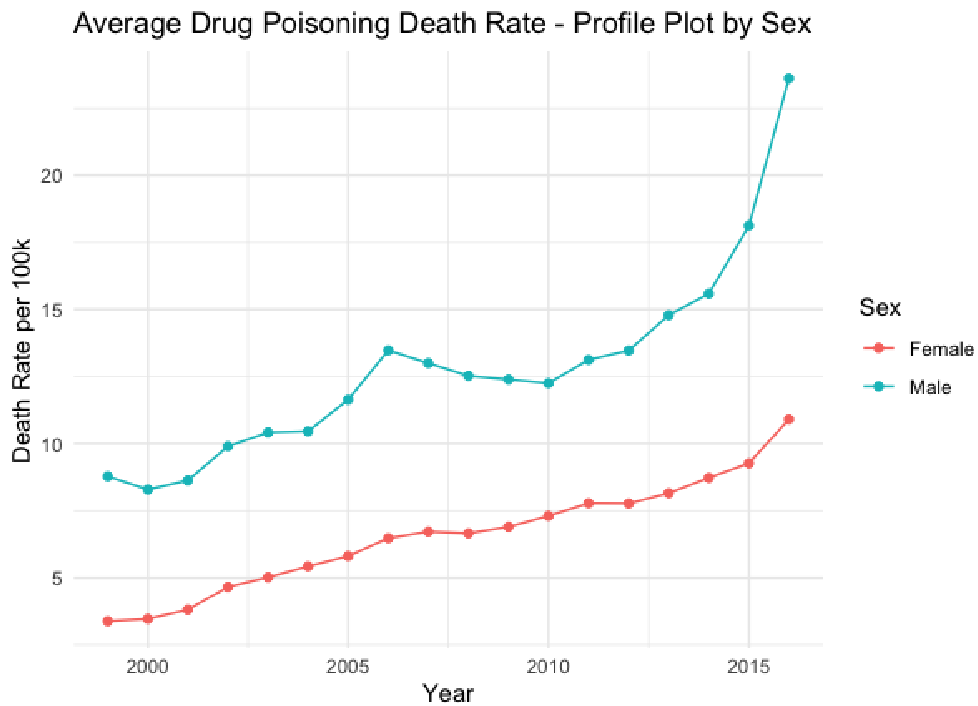

Figure 1: Average Drug Poisoning Death Rate Split By Sex

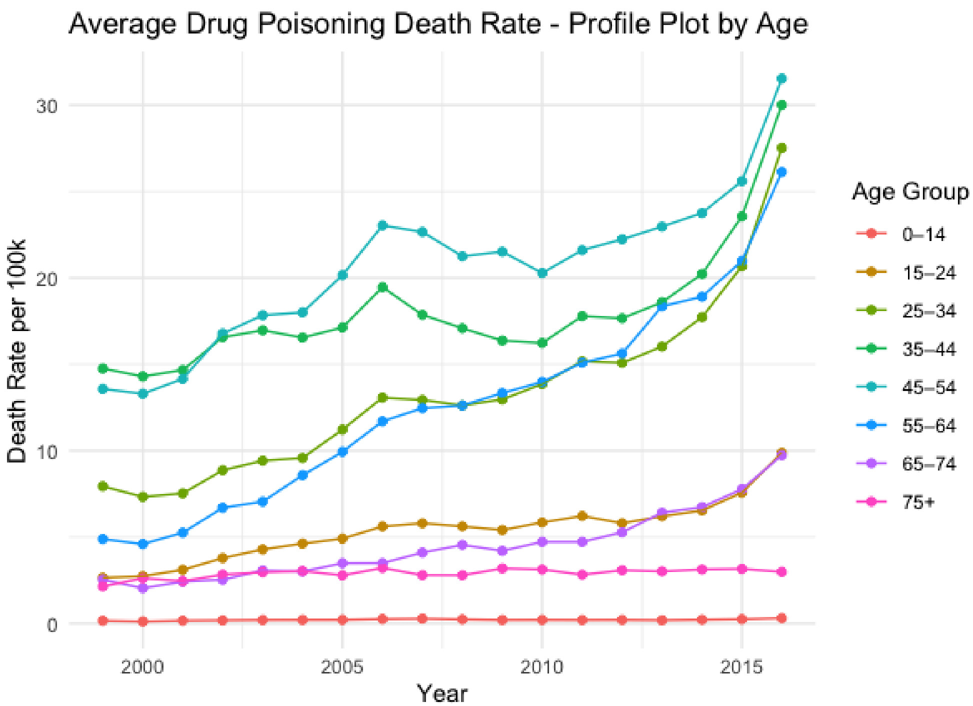

Figure 2: Average Drug Poisoning Death Rate Split By Age Group

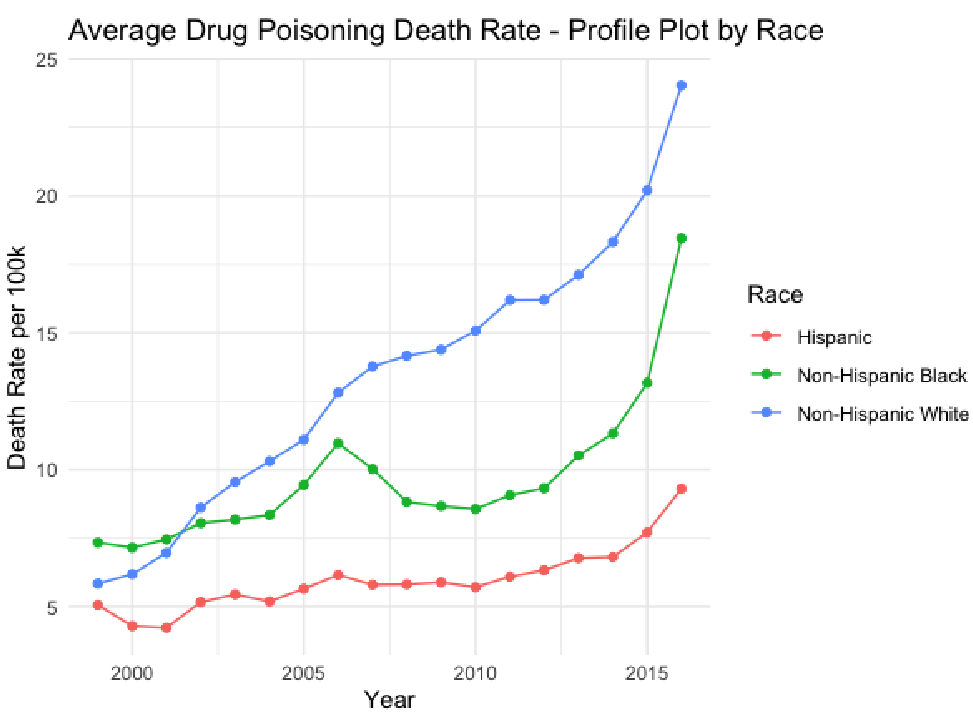

Figure 3: Average Drug Poisoning Death Rate Split By Race

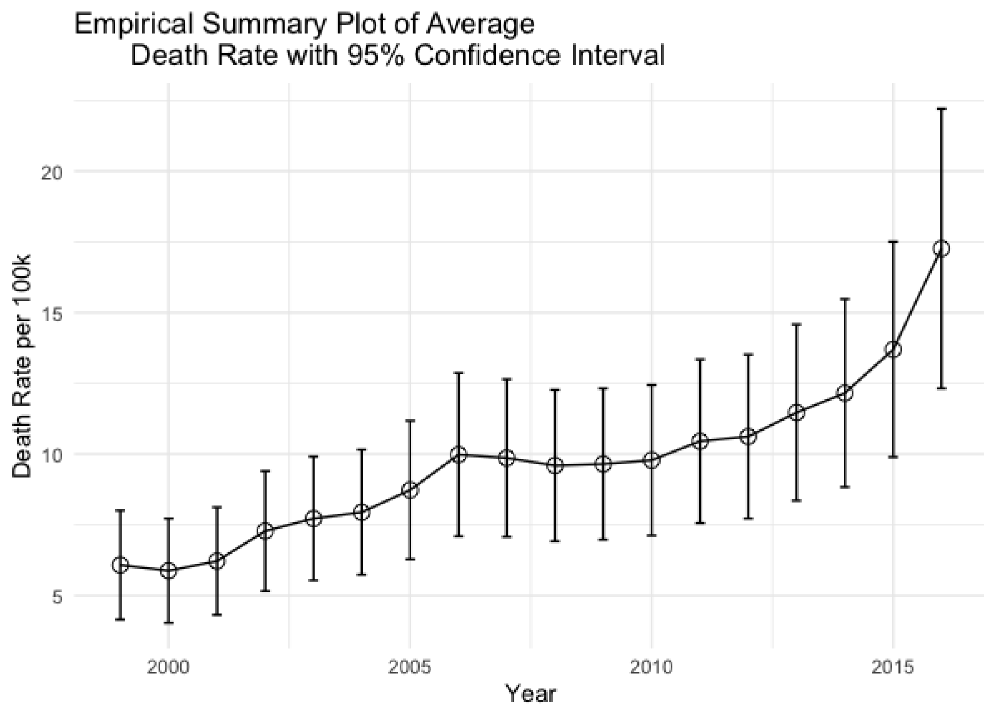

Figure 4: Average Drug Poisoning Death Rate over Time with 95% Confidence Intervals

	df	AIC	BIC
RI_l	4	5037.103	5056.149
RIAS_l	6	5032.859	5061.428
CAR_l	5	3736.693	3760.489
RI_q	4	5053.545	5072.591
RIAS_q	6	5047.010	5075.580
CAR_q	5	3752.974	3776.770
RI_c	4	5069.412	5088.458
RIAS_c	6	5062.878	5091.448
CAR_c	5	3768.680	3792.476

Figure 5: AIC and BIC Comparison of Time Trend. (*_l stands for linear, *_q for quadratic, and *_c for cubic

Fixed effects: rate ~ sex

	Value	Std.Error	DF	t-value	p-value
(Intercept)	7.094081	2.583015	816	2.746435	0.0062
sexMale	8.782266	3.652935	46	2.404167	0.0203

Figure 6: Continuous Autoregressive Model with Sex as Fixed Effects

Fixed effects: rate ~ age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.228505	4.554483	816	0.050172	0.9600
age15-24	5.930962	6.441011	40	0.920812	0.3627
age25-34	16.990811	6.441011	40	2.637910	0.0118
age35-44	21.658210	6.441011	40	3.362548	0.0017
age45-54	22.095228	6.441011	40	3.430397	0.0014
age55-64	14.939243	6.441011	40	2.319394	0.0256
age65-74	5.714862	6.441011	40	0.887262	0.3802
age75+	2.379347	6.441011	40	0.369406	0.7138

Figure 7: Continuous Autoregressive Model with Age as Fixed Effects

Fixed effects: rate ~ race

	Value	Std.Error	DF	t-value	p-value
(Intercept)	7.074097	3.288681	816	2.151044	0.0318
raceNon-Hispanic Black	5.543738	4.650897	45	1.191972	0.2395
raceNon-Hispanic White	7.725079	4.650897	45	1.660987	0.1037

Figure 8: Continuous Autoregressive Model with Race as Fixed Effects

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
RI_age	1	11	3787.764	3840.039	-1882.882			
RI_sex	2	5	3824.823	3848.619	-1907.411	1 vs 2	49.05810	<.0001
RI_sa	3	12	3777.707	3834.720	-1876.853	2 vs 3	61.11558	<.0001

Figure 9: Anova Comparison of Separate Predictors vs Combined Predictors (*_sa stands for sex + age)

```
Model 1: rate ~ sex
Model 2: rate ~ sex + age
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -1907.4
2  12 -1876.8  7 61.116  9.035e-11 ***
Model 1: rate ~ age
Model 2: rate ~ sex + age
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -1882.9
2  12 -1876.8  1 12.057  0.0005158 ***
```

Figure 10: Log-Ratio Test of Sex vs Sex + Age and Age vs Sex + Age

Fixed effects: rate ~ sex + age + year

	Value	Std.Error	DF	t-value	p-value
(Intercept)	-1321.3537	122.44038	815	-10.791813	0.0000
sexMale	8.5993	2.47866	39	3.469318	0.0013
age15-24	5.8931	4.95733	39	1.188765	0.2417
age25-34	16.8002	4.95733	39	3.388961	0.0016
age35-44	21.4736	4.95733	39	4.331692	0.0001
age45-54	22.0102	4.95733	39	4.439939	0.0001
age55-64	14.8128	4.95733	39	2.988054	0.0048
age65-74	5.6447	4.95733	39	1.138658	0.2618
age75+	2.3945	4.95733	39	0.483026	0.6318
year	0.6562	0.06096	815	10.763518	0.0000

Figure 11: Continuous Autoregressive Model with Sex + Age + Year as Fixed Effects

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
RI_sa	1	12	3777.707	3834.720	-1876.853			
RI_say	2	13	3675.612	3737.361	-1824.806	1 vs 2	104.0952	<.0001

Figure 12: Anova Comparison of Sex + Age vs Sex + Age + Year

	df	AIC	BIC
RI	12	4.969209e+03	5.026348e+03
RS	12	5.099651e+03	5.156790e+03
RIAS	14	4.966998e+03	5.033660e+03
IN	10	2.362758e+12	2.362758e+12
ARMA11	14	3.593853e+03	3.660352e+03
CAR	13	3.675612e+03	3.737361e+03

Figure 13: AIC and BIC Scores of Various Models with Sex + Age + Year as Fixed Effects

	#Df	LogLik	Df	Chisq	Pr(>Chisq)	
1	14	-1782.9				
2	13	-1824.8	-1	83.759	< 2.2e-16	***

Figure 14: Log-Ratio Test of ARMA(1, 1) vs CAR(1)