Comparing Hospitalization Rates Due to Opioid Abuse Between the U.S. and Canada

Samantha P. MacKinnon, Lori S. McNeill, Stephanie Pan & Aniket S. Shah

Harvard University

The opioid epidemic is at an all-time high. On average, 115 persons die from opioid abuse every day in the United States of America (U.S.) Between the years 1999 – 2016, the death toll from opioid overdose is estimated to be over 350,000 individuals (CDC, 2017). The rising trend over that period has been identified in three distinct waves. The first wave began in the early '90s and is associated with marketing efforts of Purdue Pharmaceuticals, manufacturer of OxyContin (Kolodny et al., 2015). The next wave began in approximately 2010 and was attributed to drug users shifting from prescribed opioids to heroin. The third wave started around 2013 and deaths were associated to synthetic opioids which include the use of fentanyl (Rudd et al., 2016). The U.S. is not the only country devastated by opioid overdose. Canada is also suffering staggeringly high rates of hospitalization and deaths attributed to opioid abuse. The death toll has more than tripled due to drug overdose in parts of Canada like British Columbia (B.C. Coroners Service, 2018).

The objective of our project is to compare emergency department visits related to opioid abuse rates between the American and Canadian population. We then plan to do a deeper dive into specific variables such as *age* and *gender* that influence opioid abuse rates. We will analyze two comparable data sets; one from the U.S. and one from Canada.

The alternative hypothesis(*Ha*) is that *Males* in the U.S. between the ages of 25 and 44 have a higher rate of opioid abuse in comparison to *Males* in Canada who belong to the same age group and that *Females* in Canada have a higher rate of opioid abuse in comparison to *Females* in the U.S. The null hypothesis (*Ho*) is that there is no difference in abuse rates between *Males* or *Females* ages 25 to 44 in both the American and Canadian population.

The following variables will be compared between datasets from the U.S. and Canada including: *rate of emergency department (ED) visits* per 100,000 individuals from 2007-2015

segregated by *age* and *gender*. The main focus of the analysis will be on the *age* category 25-44 years old. Data from the Healthcare Cost and Utilization Project (HCUP) indicate that hospitalizations due to opioid abuse are higher than any other age category between the years 2007-2015 and the rate of *emergency department visits* in that specific age category have more than doubled in the last four years of that research (CIHI, 2016).

Based on current research it appears that *Males* in the U.S. have a *higher rate of emergency department visits* due to opioid abuse than their Canadian counterparts (CIHI, 2016). Alternatively, research also indicates that *Females* in Canada appear to have a higher rate of emergency department visits due to opioid abuse in the U.S. In comparison to the female population, men in both the U.S. and Canada are twice as likely to abuse opioids, which could be because doctors prescribe higher doses of opioids to men in comparison to women. Opioid abuse could be more prevalent in the 25 to 44 age group because those prescribed opioids for the first time are assumed to be younger and could continue to use the drugs long after their injury has healed. Another assumption is that opioid abuse is highest for lower income *Males* in both the American and Canadian population groups.

Method

Participants

The data was comprised of over 100,000 male and female participants from both Canada and the U.S that were hospitalized due to opioid abuse. Participants were separated into three *age* categories, *25 to 44*, *45 to 64* and *65 and above*. This was a longitudinal observational study; therefore, patients were not recruited for the purpose of this analysis.

Materials

The data set from the U.S. was extracted from the Healthcare Cost and Utilization Project (HCUP). This project is a Federal-State-Industry partnership. It is the longest longitudinal study regarding opioids and hospital care beginning in 1988. This data set provided rates of emergency department visits due to opioid overdose on a national level from the years of interest, 2007-2015, with demographic information captured such as age, gender, healthcare status, etc. The collection of data was done through the hospital billing records and other administrative manners. Data on individual patients would only be collected once they were discharged from the hospital.

The Canadian data set was provided by the Canadian Institute for Health Information (CIHI) which aggregated data regarding opioid trends from 2007-2015. The CIHI is an independent, not-for-profit organization that collects information on healthcare and general health statistics within Canada. The two sources that provide CIHI its data is the Hospital Morbidity Database (HMDB) and the National Ambulatory Care Reporting System (NACRS). The Hospital Morbidity Database records patient hospitalizations. The National Ambulatory Care Reporting System records ambulatory care for 60% of the country. Both provide data directly to the CIHI.

Procedure

All data sets had opioid rates categorized by only one predictor variable, however, we realized that to run any statistical models on a dataset with multiple predictors, the rates needed to be stratified. Since we evaluated *age* and *gender*, the data sets needed to have a rate for each combination of groups, for example, Male & Age 25-44, Female & Age 25-44, Male & Age 45-

64, etc. Since the data sets only contained separate rates for each predictor variable, the data was stratified by taking the individual ratios of each group, and multiplying by the national rate.

Stratified Rate = (Annual rate)*(Gender specific rate/(Total rate for Gender))*(Age group specific rate/(Total Rate for Age))

An example is provided below for Canadian Males between the ages of 25-44 in the year 2007.

Stratified Rate_{M,25-44,2007, Canada} = Opioid Rate_{2007,Canada} * ((Rate_M)/(Rate_M + Rate_F)) * ((Rate₂₅₋₄₄)/(Rate_{<15} + Rate₁₅₋₂₄ + Rate₂₅₋₄₄ + Rate₄₅₋₆₄ + Rate₆₅₊)

This technique demonstrates good approximation of the actual stratified rate, given the assumptions that the proportion of *Males* and *Females* in a country is approximately equal, and the distribution of age groups amongst *Males* and *Females* is roughly similar.

The individual data sets for U.S. and Canada had to be aligned as they were collected over different time periods and contained various categorical variables. Our first step was to identify which variables to evaluate and determine which time period to include in the analysis. The variables found in both data sets were: *age, gender, year (2007-2015)* and *opioid hospitalization rate*. We then created a combined data set with both U.S. and Canada variables of interest. The age groups found in common were *25-44, 45-64, 65 and older*.

Our original plan was to run multiple regression for the combined data, however, once we started our analysis it became apparent that we needed to run ANOVA to specifically compare *rates* between the two countries. It was necessary to run an additional test because the *rates* in the U.S. were significantly higher than *rates* in Canada. Therefore, multiple regression models were built to identify patterns within each country separately, and a one-way ANOVA was run the for combined data.

To test the alternative hypothesis $H_a = \mu_{Canada,Male} < \mu_{U.S.,Male}$, a data set was created which contained stratified *opioid hospitalization rates* from 2007-2015 for *males* aged 25-44 in U.S. and Canada. A one-way ANOVA confirmed the alternative hypothesis given $\mu_{Canada,Male} =$ 1.22 & $\mu_{U.S.,Male} = 35.54$, F(1, 16)=63.122, p <0.001. $H_o = \mu_{Canada,Male} = \mu_{U.S.,Male}$

To test the alternative hypothesis $H_a = \mu_{Canada,Female} > \mu_{U.S.,Female}$, a data set was created which contained stratified *opioid hospitalization rates* from 2007-2015 for *females* aged 25-44 in U.S. and Canada. One-way ANOVA did not confirm the alternative hypothesis even though the F-Statistic was significant, $\mu_{Canada,Female} < \mu_{U.S.,Female}$, F(1, 16)=74.483, p<0.001.

ANOVA tests help test the hypotheses but did not really give us any insights about the epidemic. We explored our data further by building multiple regression models on individual data sets. For Canada, the predictors included *age* and *gender*. For the U.S., the predictors included *age, gender, income* and *patient location*. For both data sets, the criteria for normality was not met (Figures 11 and 14). We then applied two transformations (square-root and Log) on the *rate* variable and attempted to normalize the data but were unsuccessful. However, the coefficients in all the models were statistically significant, and the models gave us an interesting outlook on the opioid epidemic.

Results

We stratified and organized the opioid data to perform two types of tests; Multiple Regression and one-way ANOVA. One-way ANOVA was performed to test our hypotheses with respect to the relation of *opioid rates* per 100,000 to *age* and *gender* in the U.S. and Canada. Multiple regression was conducted to determine the impact of the categorical variables within each country. The results for one-way ANOVA on combined data (U.S. and Canada) for *Males* aged 25-44 are shown in Figures 1 and 2. The difference between the two countries is significant (F(1, 16)=63.122, p <0.001; Figure 1), and that the mean rate for *Males* aged 25-44 in U.S. (35.54) is higher than for *Males* aged 25-44 in Canada (1.22).

The results for one-way ANOVA on combined data (U.S and Canada) for *Females* aged 25-44 are shown in Figures 3 and 4. As portrayed, the difference between groups is significant (F(1, 16)=74.483, p<0.001; Figure 1), and that the mean rate for *Females* aged 25-44 in U.S. (28.18) is higher than for *Females* aged 25-44 in Canada (1.39). Multiple regression is performed individually for both Canada and U.S. We did not perform multiple regression for the combined data as the rates in Canada are grouped so closely compared to the U.S. This skews the data tremendously such that normality tests fail (see histograms in Figures 5 and 6); a square root transformation of the rates does not fix it (Figure 6 and 5).

Multiple regression for Canada included the predictor variables *gender* and *age*. Our data does not meet the criterion for normality, evident from the Durbin-Watson value of 0.548 (less than 1 is a concern), the histogram in Figure 10 and the normal P-P plot in Figure 11. Having said that, observing the high R^2 value in Figure 9 and significant coefficients in Figure 12, interpretations can still be made. The model with *age* and *gender* included has only a slight gain in R^2 (0.915 vs. 0.930) over a model that includes just the age groups. In the coefficients table, for model 2 (*age* and *gender*) are as follows: $b_{Ages 15-24}$ =0.750, t(7) = 12.17, p<.001. $b_{Ages 25-44}$ = 1.15, t(7) = 18.61, p<.001. $b_{Ages 45-64}$ = 1.16, t(7) = 26.27, p<.001. Ages 65 and older, $b_{Ages 65 and}$ older = 1.803, t(7) = 29.27, p<.001.

For gender, *Females* are coded as 1 and *Males* are coded as 2. The coefficient table show *Males* are significantly more likely to be hospitalized than *Females*, b_{Gender} = -0.167, t(7) = -

4.291, p<.001. The model was able to account for 96.5% of the variance in *rates of opioid hospitalization* in the U.S., F(1,84) = 18.409, p<.001. $R^2 = .97$ shown in Figure 12.

Multiple regression for U.S. included the predictor variables gender, age, income and patient location. Income contained four categories of income quartiles from lowest to highest. Patient Location contained the categories rural, and four metro groups from smallest to largest. Again, our data did not meet the criterion for normality exhibited in Figure 14, but the significant coefficient values in the model did help gain valuable insight. Model 4 in Figure 13 included all the categorical variables and had an *Adjusted*- R^2 of 0.883. This model was able to account for 94% variance in rates of opioid hospitalizations in Canada, F(1,1787) = 183.328, p<.001, $R^2 =$.94. In the age category, 25-44 had the highest statistically significant coefficient, $b_{Ages 15-44} =$ 1.551, t(1773) = 66.90, p<.001. In the *income* category, all the statistically significant coefficients were negative, suggesting the constant which represented the lowest income category, had the greatest impact on rate. In *patient locations*, the coefficients for small metropolitans and large fringe metropolitans were statistically insignificant, suggesting their influence on the rate is like the constant group (rural). Medium metropolitans had the highest significant coefficient amongst all location groups, $b_{\text{Medium metro}} = .125$, t(1773) = 5.4, p<.001. Gender was coded as *female* 1 and *male* 2, and the coefficient value was positive and significant, b_{Gender} =.157, t(1773) = 10.71, p<.001.

Discussion

Comparing a sample of over 100,000 men and women in the U.S. and Canada between the years 2007 to 2015, it was hypothesized that there would be a difference in opioid abuse rates between *males* and *females* aged 25-44 of the two countries. It was predicted that *males* in the U.S. from the aforementioned age group would have higher *rates* of Emergency Department visits

compared to their *male* counterparts in Canada. It was further predicted that Canadian *females* in the 25-44 age group would have higher *rates* of Emergency Department visits compared to *females* in the U.S. The H_a was found to be true for the U.S. population, in that *males* aged 25-44 had the highest *rates* of Emergency Department visits, but for Canada the H_a was not confirmed, as *females* in the U.S. had higher rates. However, it was interesting to find that in Canada, those aged 65 and *older* in both male and female groups had higher Emergency Department visits than those between the ages of 25-44. It was anticipated that in both the U.S. and Canada, *intentional overdose* would be the most common reason for hospitalization, this was found to be true in the U.S. Surprisingly, the prime reason for hospitalization related to opioid abuse in Canada was *accidental*.

Limitations

One limitation is that the Canadian data set was grouped by the fiscal cycle of April 1st to March 31st of the following year, and the U.S. data set was based on calendar year. Furthermore, once we started analysing our data, we found that although our hypothesis was formulated based on the effects of opioid abuse in regard to gender in each country, age presented itself as a much more influential factor than gender. Types of opioids were not specified, so for future studies it would be beneficial to examine which types of opioids were most commonly prescribed and used. The Durbin-Watson in our results is lower than 1 meaning there is serious concern for normalcy within our data and it hints at high positive autocorrelation. This could be attributed to the fact that our data set is a time series with high autocorrelation.

Conclusions

Opioid-related deaths are an important issue which requires attention in both the U.S. and Canada. Highlighted in this paper are important differences in the patterns that could influence these behaviors, specifically, prescribing practices between the two countries and opioid use amongst *Males* and *Females* in the varying age groups.

We were interested in examining the patterns of opioid abuse rate in the neighboring countries who rank first (U.S.) and second (Canada) in the world based on opioid abuse rates ("Understanding the epidemic," 2017). Through our research we have uncovered that in the U.S., the age group of *25 to 44* years is significant because it speaks to the pattern of drug abuse which mirrors the three waves of the rise in opioid-related deaths. The first wave was attributed to the use of prescription of opioids in the 1990's as general pain management. During this time, opioids were being marketed as "safe" and "non-addictive." Opioids acted as a gateway to stronger drugs like heroin, namely the second wave which began in 2010. It is believed that once students graduated college, they were able to support their drug habit. The third wave began in 2013 with higher rates of opioid related deaths associated with fentanyl ("Understanding the epidemic," 2017).

In comparison to the U.S., it was found in Canada that those aged 65 and older had a higher rate of hospitalization due to opioid abuse. It was noted that there was a difference in prescribing habits between Canada and the U.S. which resulted in an increase in hospitalization in the age category of *65 and older*. We believe a contributing factor could be comorbidities in that age group due to drug interactions or general decline in health. Although the rate of drug abuse is higher in the U.S., it is important to note that hospitalizations in the *65 and older* group

are proportionally higher in Canada. In summation, our research draws important findings as it relates to the opioid epidemic and provides considerations for future studies.

References

- B.C. Coroners Service General. (2018). *Illicit Drug Overdose Deaths in BC*. Burnaby, BC:Ministry of Public Safety & Solicitor General
- Canadian Institute for Public Health Information. (2016). Canadian Centre on Substance Abuse. *Hospitalizations and Emergency Department Visits Due to Opioid Poisoning in Canada.* Ottawa, ON: CIHI
- Center for Disease Control and Prevention. (2017, September 26). Retrieved from https://www.cdc.gov/vitalsigns/pdf/2017-07-vitalsigns.pdf
- Kolodny et al. (2015). The prescription opioid and heroin crisis: A public health approach to an epidemic of addiction. Annual Review of Public Health, 36, 559-74
- Rudd R.A., Aleshire N., Zibbell J.E., & Gladden R.M. (2016). Increases in Drug and Opioid
 Overdose Deaths United States 2000-2014. *Morbidity and Mortality Weekly Report*, 64(50); 1378-82.
- Understanding the Epidemic. (2017, August 30). Retrieved from https://www.cdc.gov/drugoverdose/epidemic/index.html/

APPENDICES

One-way ANOVA – Males and Ages 25-44

ANOVA

Rate					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5300.512	1	5300.512	63.122	.000
Within Groups	1343.560	16	83.973		
Total	6644.073	17			

Figure 1

Report

Rate			
Country_Code	Mean	N	Std. Deviation
Canada	1.216725763	9	.2221392409
USA	35.53715110	9	12.95745723
Total	18.37693843	18	19.76936508

Figure 2

One-way ANOVA – Females and Ages 25-44

ANOVA

Rate

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3229.209	1	3229.209	74.483	.000
Within Groups	693.682	16	43.355		
Total	3922.891	17			

Figure 3

Report

Rate			
Country_Code	Mean	N	Std. Deviation
Canada	1.393686349	9	.1645453088
USA	28.18178168	9	9.310377846
Total	14.78773401	18	15.19073037

Figure 4

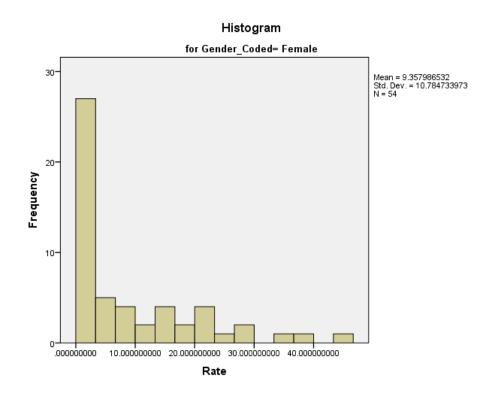


Figure 5 – Histogram of Opioid Rates for *Female* (Combined Data)

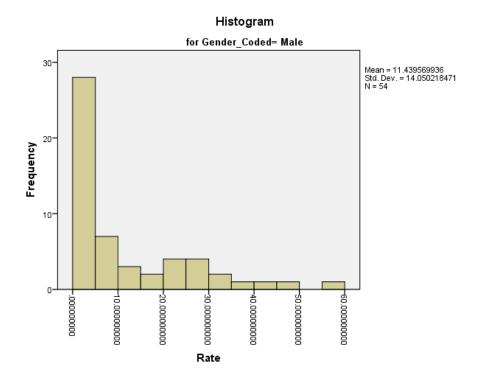


Figure 6 – Histogram of Opioid Rates for Male (Combined Data)

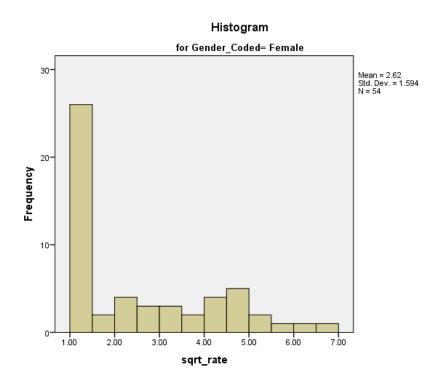


Figure 7 – Histogram of square-rooted Opioid Rates for *Female* (Combined Data)

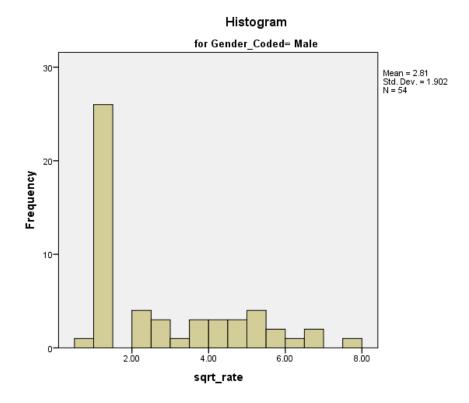


Figure 8 – Histogram of square-rooted Opioid Rates for Male (Combined Data)

Multiple Regression (Canada)

Model Summary^C

						Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin- Watson
1	.957 ^a	.915	.911	.2027844127	.915	228.722	4	85	.000	
2	.965 ^b	.930	.926	.1847461256	.015	18.409	1	84	.000	.548

a. Predictors: (Constant), Age_Coded=65+, Age_Coded=45-64, Age_Coded=25-44, Age_Coded=15-24

b. Predictors: (Constant), Age_Coded=65+, Age_Coded=45-64, Age_Coded=25-44, Age_Coded=15-24, Gender_coded

c. Dependent Variable: rateColumn

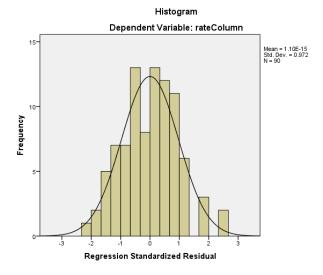


Figure 9 – Model Summary (Canada Data)

Figure 10 – Histogram of Opioid Rates (Canada Data)

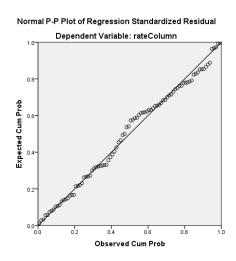


Figure 11 - Normal P-P plot for Opioid Rates (Canada Data)

		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confider	ice Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.159	.048		3.326	.001	.064	.254
	Age_Coded=15-24 .750 .068 .444		.444	11.089	.000	.615	.884	
	Age_Coded=25-44	1.146	.068	.678	16.958	.000	1.012	1.281
	Age_Coded=45-64	1.618	.068	.957	23.931	.000	1.483	1.752
	Age_Coded=65+	1.803	.068	1.067	26.672	.000	1.668	1.937
2	(Constant)	.410	.073		5.622	.000	.265	.555
	Age_Coded=15-24	.750	.062	.444	12.172	.000	.627	.872
	Age_Coded=25-44	1.146	.062	.678	18.613	.000	1.024	1.269
	Age_Coded=45-64	1.618	.062	.957	26.267	.000	1.495	1.740
	Age_Coded=65+	1.803	.062	1.067	29.276	.000	1.680	1.925
	Gender_coded	167	.039	124	-4.291	.000	245	090

Coefficients^a

a. Dependent Variable: rateColumn

Figure 12 - A table of coefficient values obtained for multiple regression models (Canada Data)

Multiple Regression (U.S.)

Model Summary^e

						Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin- Watson
1	.914 ^a	.836	.835	.15467	.836	2284.024	4	1795	.000	
2	.931 ^b	.867	.867	.13923	.031	141.110	3	1792	.000	
3	.934°	.872	.872	.13661	.005	18.318	4	1788	.000	
4	.940 ^d	.884	.883	.13014	.012	183.328	1	1787	.000	.147

a. Predictors: (Constant), Age_Groups=65+, Age_Groups=1-24, Age_Groups=25-44, Age_Groups=45-64

b. Predictors: (Constant), Age_Groups=65+, Age_Groups=1-24, Age_Groups=25-44, Age_Groups=45-64, Income_Groups=Income Quartile 4 (Highest), Income_Groups=Income Quartile 3 (2nd Highest), Income_Groups=Income Quartile 2 (2nd Lowest)

c. Predictors: (Constant), Age_Groups=65+, Age_Groups=1-24, Age_Groups=25-44, Age_Groups=45-64, Income_Groups=Income Quartile 4 (Highest), Income_Groups=Income Quartile 3 (2nd Highest), Income_Groups=Income Quartile 2 (2nd Lowest), Locations=Small Metropolitan, Locations=Large Central Metropolitan, Locations=Medium Metropolitan, Locations=Large Fringe Metropolitan

d. Predictors: (Constant), Age_Groups=65+, Age_Groups=1-24, Age_Groups=25-44, Age_Groups=45-64, Income_Groups=Income Quartile 4 (Highest), Income_Groups=Income Quartile 3 (2nd Highest), Income_Groups=Income Quartile 2 (2nd Lowest), Locations=Small Metropolitan, Locations=Large Central Metropolitan, Locations=Medium Metropolitan, Locations=Large Fringe Metropolitan, Gender_Coded

e. Dependent Variable: sqrtRate

Figure 13 – Model Summary (U.S. Data)

Normal P-P Plot of Regression Standardized Residual

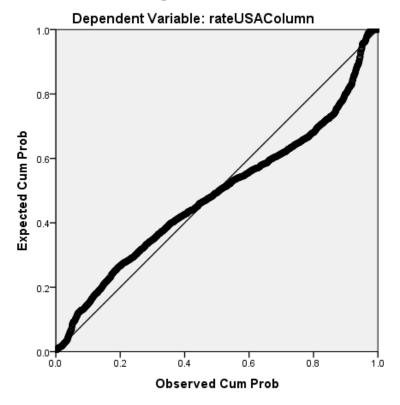


Figure 14 – Normal P-P plot for Opioid Rates (U.S. Data)

		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confider	nce Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.042	.018		2.308	.021	.006	.078
	Age_Groups=1-24	.472	.026	.295	18.187	.000	.421	.522
	Age_Groups=25-44	1.551	.026	.969	59.809	.000	1.500	1.602
	Age_Groups=45-64	.890	.026	.556	34.322	.000	.839	.941
	Age_Groups=65+	.300	.026	.187	11.571	.000	.249	.351
2	(Constant)	193	.030		-6.466	.000	252	135
	Age_Groups=1-24	.472	.025	.295	18.666	.000	.422	.521
	Age_Groups=25-44	1.551	.025	.969	61.381	.000	1.501	1.600
	Age_Groups=45-64	.890	.025	.556	35.224	.000	.840	.939
	Age_Groups=65+	.300	.025	.187	11.876	.000	.250	.350
	Gender_Coded	.157	.016	.123	9.829	.000	.126	.188
3	(Constant)	011	.031		346	.730	071	.049
	Age_Groups=1-24	.472	.023	.295	20.072	.000	.425	.518
	Age_Groups=25-44	1.551	.023	.969	66.004	.000	1.505	1.597
	Age_Groups=45-64	.890	.023	.556	37.877	.000	.844	.936
	Age_Groups=65+	.300	.023	.187	12.770	.000	.254	.346
	Gender_Coded	.157	.015	.123	10.570	.000	.128	.186
	Income_Groups=Income Quartile 2 (2nd Lowest)	154	.021	104	-7.334	.000	195	113
	Income_Groups=Income Quartile 3 (2nd Highest)	233	.021	158	-11.098	.000	274	192
	Income_Groups=Income Quartile 4 (Highest)	343	.021	232	-16.343	.000	385	302
4	(Constant)	055	.034		-1.649	.099	121	.010
	Age_Groups=1-24	.472	.023	.295	20.343	.000	.426	.517
	Age_Groups=25-44	1.551	.023	.969	66.896	.000	1.505	1.596
	Age_Groups=45-64	.890	.023	.556	38.388	.000	.844	.935
	Age_Groups=65+	.300	.023	.187	12.942	.000	.255	.345
	Gender_Coded	.157	.015	.123	10.712	.000	.128	.186
	Income_Groups=Income Quartile 2 (2nd Lowest)	154	.021	104	-7.433	.000	195	113
	Income_Groups=Income Quartile 3 (2nd Highest)	233	.021	158	-11.248	.000	274	193
	Income_Groups=Income Quartile 4 (Highest)	343	.021	232	-16.564	.000	384	303
	Locations=Large Central Metropolitan	.083	.023	.052	3.572	.000	.037	.128
	Locations=Large Fringe Metropolitan	.036	.023	.022	1.546	.122	010	.081
	Locations=Medium Metropolitan	.125	.023	.078	5.400	.000	.080	.171
	Locations=Small Metropolitan	020	.023	012	851	.395	065	.026

Coefficients^a

a. Dependent Variable: rateUSAColumn

Figure 15 – A table of coefficient values obtained for all multiple regression models (U.S. Data)