Glycohemoglobin and renal function

Diabetes mellitus is a complication of the body's glucose metabolic system that is associated with increased blood glucose (Uma et al 2001). In the US, diabetes has a high prevalence, in 2018, it was responsible for 2.3% of the Years of Potential Life Lost in people below 65 years (CDC YPLL). One of the mechanisms by which diabetes works is through nonenzymatic glycation of blood proteins causing them to alter conformation. Hemoglobin is one of the proteins that undergoes nonenzymatic glycation. Hemoglobin that undergoes glycation is called glycohemoglobin (GHb). The formation of GHb is slow, cumulative and irreversible, hence its ratio is a reliable measure of blood glucose concentration (Uma et al 2001). The kidney plays a very important role in ridding the body of toxins, maintaining blood pH, bone resorption and production, blood pressure regulation and maintaining homeostatic conditions in the body, its importance demands a proper measure to verify that it is functioning properly. Glomerular Filtration Rate is one of the most widely used measures to examine renal clearance and a normal GFR is 90ml/min/1.73 m^2 anything below can be considered a health risk factor. Serum creatinine levels are considered an endogenous marker for GFR levels. Another alternate test that indicates kidney function is the urine albumin levels test (Kimmel et al). Nephropathy which is the damage of the kidney is one of the symptoms and is associated with abnormal levels of urine albumin, 80% of subjects with type-1-diabetes with type 1 diabetes who have consistent abnormal albumin levels continue to see an increase in albumin levels and a drastic change in GFR over a period of time (Molitch et al).

In this paper we will be using statistical analysis to examine the association between glycohemoglobin levels and renal function.

Methods

In order to analyze the relationship between glycohemoglobin levels and renal function, data containing 1304 participant information was used. The relationship between sociodemographic of the individuals e.g race, gender, Household income, education level and age were tested against glycohemoglobin levels and p-values recorded with $\alpha = 0.05$ representing the cut off point for significant associations. To do this, a new categorical variable was created to specify diabetic and prediabetic populations based on their glycohemoglobin levels [<6.5% and has "Diabetes" if glycohemoglobin is >= 6.5%.] and better understand their relationship between each sociodemographic variable. Next, the association between urine albumin, serum creatinine and Hba1c levels was tested. In order to do this, new categorical variables were created out of the levels of glycohemoglobin [Normal(<5.7% glycohemoglobin), Pre-

diabetic(5.7 to 6.4% glycohemoglobin), and Diabetic (≥ 6.5% glycohemoglobin], a statistical hypothesis was defined where:

Ho = mean urine albumin levels/ serum creatinine levels for normal glycohemoglobin = mean urine albumin/ serum creatinine levels for prediabetic = mean urine albumin/serum creatinine level for diabetic.

Ha = mean of at least one group is different

The cut off for statistical significance $\alpha = 0.05$

To properly test for significance, some assumptions were assessed. The command duplicates report was used to assess the independence if the data set and make sure there were no duplicates in the data set, normality of the data set was analyzed using the Shapiro-Wilk test for normality. The ladder command was used to assess possible transformations that could be applied to ensure normality is met before conducting any parametric tests. After testing for these assumptions, a Non parametric K wallis test for multiple variables was applied over ANOVA as none of the transformations were effective enough to produce a normal distribution in the samples being analyzed.

Next, a regression analysis was carried out to assess any associations between glycohemoglobin levels and renal functions i.e. urine albumin level and serum creatinine levels. In order to properly assess these using the regression model the entire data set was analyzed for duplicates using the STATA command duplicates report id , a histogram plot of the residuals was used to check for any outliers in the dataset, a yvf plot of the residuals was also used to check for constant variability in the dataset. Scatter plots of glycohemoglobin levels and urine albumin/ serum creatinine levels were created to assess linearity and constant variability before interpreting the results of the regression model.

Results:

Statistical hypothesis testing was conducted to determine if the sociodemographic variables, age, race, gender, household income and educational level were associated with diabetes status where Ho = no significant difference between groups and Ha = there is a difference in status between groups. For the categorical variables race, gender, household income and educational level the chi-square test had p values <0.05 signifying that there may be an association between the categories of these groups. Age produced a p value of <0.01 using the Wilcoxon rank sum test, showing that there mean ages in diabetic groups differed from mean ages in non-diabetic groups.

A Kwallis test of urine albumin levels by glycohemoglobin levels showed that at least one of the groups (normal, prediabetic and diabetic) differs in their urine albumin levels. The test for serum creatinine levels also showed that at least one group had a different urine albumin level out of all three. Kwallis test was preferred because data for urine albumin and serum creatinine levels showed non- normal distributions and none of them could be appropriately transformed into normal distributions.

Further regression analysis of glycohemoglobin and urine albumin levels shows that at 0 urine albumin, glycohemoglobin levels Is expected to be 5.7 and for every 1 unit increase in urine albumin, glycohemoglobin is expected to increase by 0.0006 units. The model however has R^2 of 0.014 meaning that only 1.4% of the variability in glycohemoglobin levels Is explained by the model.

The regression model for glycohemoglobin and serum creatinine shows us that at 0 serum creatinine levels glycohemoglobin is expected to be at 5.5. and for every 1 unit increase in serum creatinine, glycohemoglobin is expected to increase by 0.27 units. This R^2 of 0.011 tells us that only 1.1% of the variability in glycohemoglobin levels is explained by the model

Discussion

The result of this statistical analysis shows that there may be a positive correlation between glycohemoglobin count and renal function. A Kwallis p-value of <0.05 showed a difference in urine albumin and serum creatinine function among three categories of diabetic status, (normal, prediabetic and diabetic). The regression analysis further went on to highlight the nature of the association possibly being positively correlated as an increase In either urine albumin or serum creatinine levels Is followed by an increase in glycohemoglobin. This agrees with previous scientific research that correlated albumin levels and GFR ratio to diabetes diagnoses (Molitch et al)

A limitation in our study could come from the weak R^2 values that show the models weakness in explaining variability in glycohemoglobin levels.

Appendix

Table 1

Statistical Analysis for independent variables

	No Diabetes <6.5 N= 1,157	Diabetes >=6.5 N= 147	Test Statistic	P value
Age (years, 49 years, SD= 18 years)	47.3 = mean SD = 17.9	62.35 = mean SD = 13.44	Wilcoxon rank sum test	[a]0.000
Gender			Chi -square	[b]0.019
Male (49.39%) n=(644)	558 = n 48.23%	86 =n 58.50%		
Female (50.61%) n= 660	599 51.77%	61 = n 41.5%		
Race			Chi square	[b] 0.00
Hispanic White (125, 9.59%)	N=106 9.16%	N= 19 12.93 %		
Non-hispanic Black (314,24%)	N= 267 23.08%	N=47 31.97%		
Non-Hispanic white (524. 40%)	489 42.26%	N= 35 23.81%		
Others (341, 26.15%)	N= 295 25.50%	N=46 31.29%		
Education			Chi square	[b] 0.000
Less than High School Grad (303, 23.24%)	246 21.26%	57 38.78%		
High School Grad(264, 20.25%)	231 19.97%	33 22.45%		
College Grad (727,56.52%)	680 58.77%	57 38.78%		
Annual household income			Chi square	[b] 0.005
<25000 (434, 33.28%)	368 31.81%	66 44.90%		
25000-54999 (376, 28.83%)	338 29.21%	38 25.85%		
>= 55000 (494,37.88%)	451 38.98%	43 29.25%		
Tried to lose weight In last year				
Yes (471, 36.12%)	410	61		

	35.44%	41.50%		
No(833, 63.88%)	747 64.56%	86 58.50%		
BMI (Mean= 28.49, SD= 6.75)	27.99= mean 6.45 =SD	32.396= mean 7.73 = SD		
Urine albumin (mean =39.49143 SD= 179.6262)	Mean= 32.3 SD = 165.78	Mean = 96.01 SD = 258.289		
Serum creatinine (mean =.9069723 SD= .3752886)	Mean = .893 SD = .375	Mean = 1.009 SD= .37		

[a]XX= Wilcoxon rank sum test

[b]XX= Chi-square test

Table 2

Regression tables for Glycohemoglobin levels with renal function. lbxgh = glycohemoglobin category, urxums = Urine albumin levels, lbxscr = serum creatinine levels.

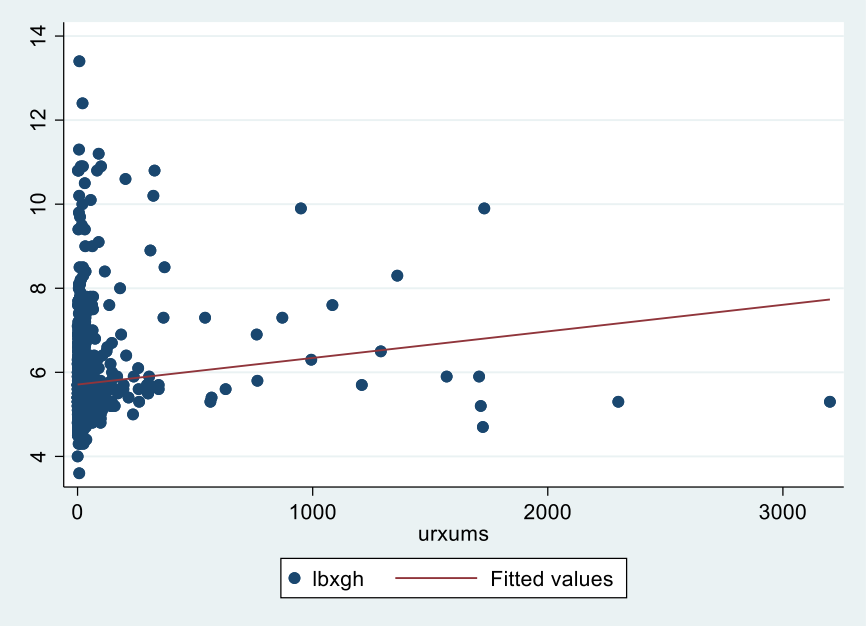
```
. regress lbxgh urxums

      Source |       SS           df       MS      Number of obs   =     1,304
-------------+----------------------------------   F(1, 1302)      =     18.95
       Model |  16.8043373         1  16.8043373   Prob > F        =    0.0000
    Residual |  1154.40103     1,302  .886636739   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0136
       Total |  1171.20537     1,303  .898852933   Root MSE        =    .94161

------------------------------------------------------------------------------
       lbxgh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      urxums |   .0006322   .0001452     4.35   0.000     .0003473    .0009171
       _cons |   5.710002   .0266988   213.87   0.000     5.657625    5.762379
------------------------------------------------------------------------------
```

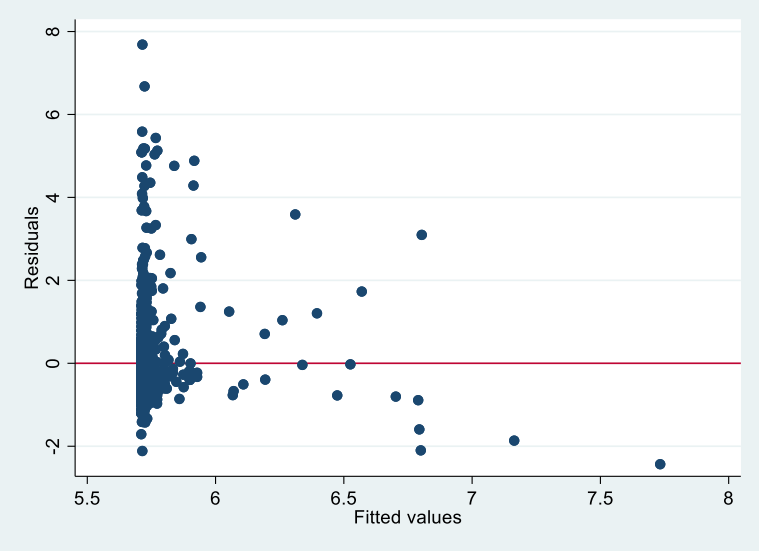
```
. regress lbxgh lbxscr
```

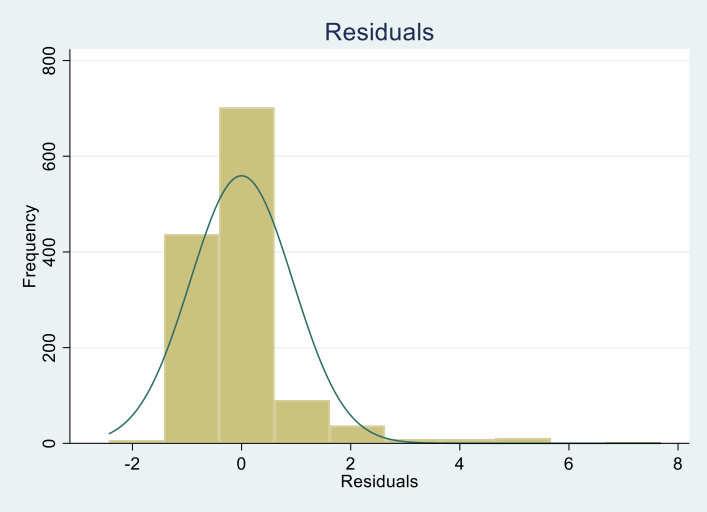
Source	SS	df	MS	Number of obs	=	1,265
				F(1, 1263)	=	14.97
Model	13.5055225	1	13.5055225	Prob > F	=	0.0001
Residual	1139.20304	1,263	.901981818	R-squared	=	0.0117
				Adj R-squared	=	0.0109
Total	1152.70856	1,264	.911952973	Root MSE	=	.94973

lbxgh	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lbxscr	.2754335	.0711803	3.87	0.000	.1357888	.4150782
_cons	5.487185	.069863	78.54	0.000	5.350125	5.624246

Checking Residual assumptions

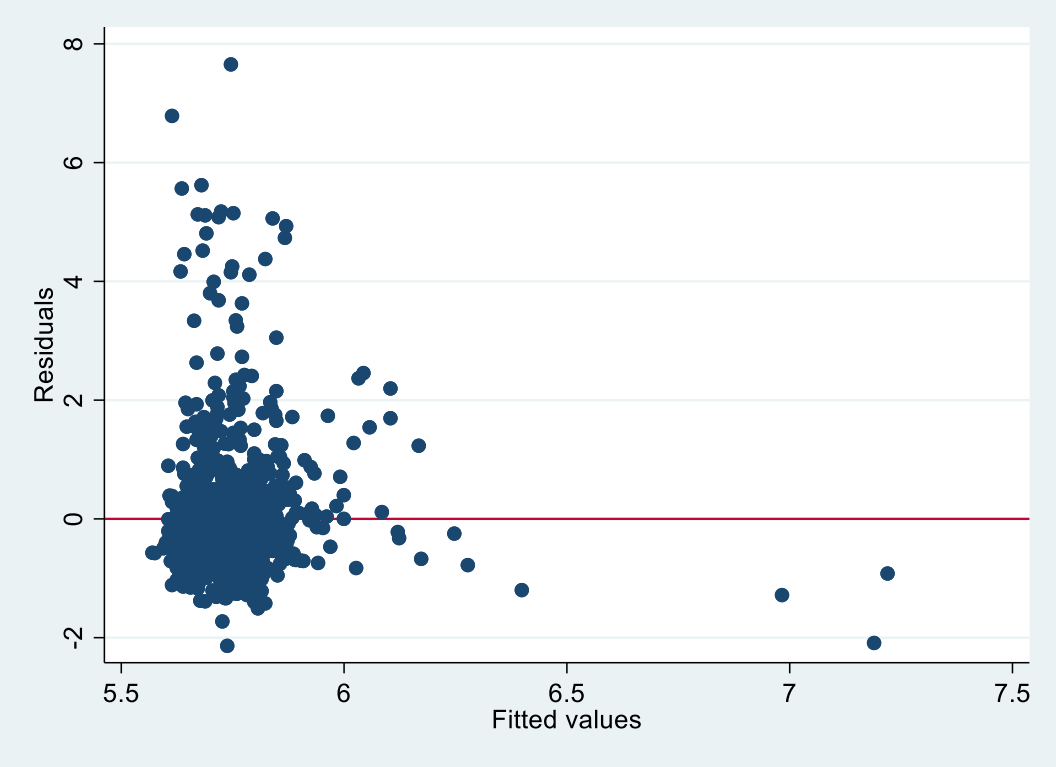


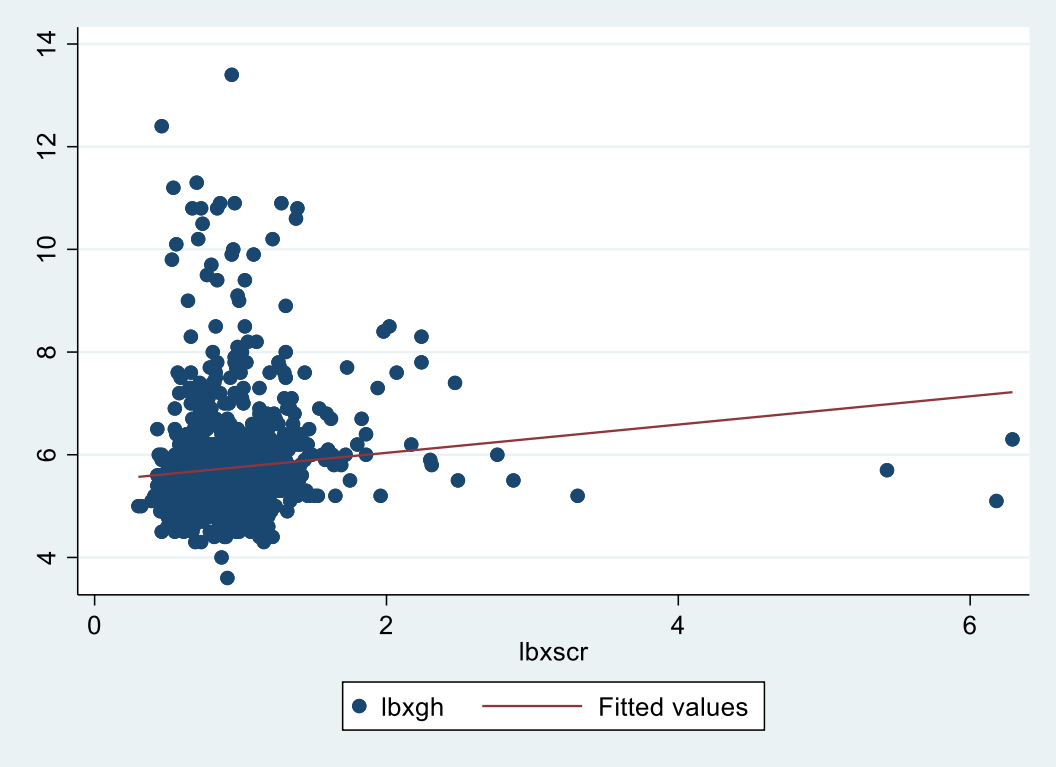





```
. ladder lbxgh

Transformation          formula              chi2(2)     P(chi2)
-----------------------------------------------------------------
cubic                   lbxgh^3                   .           .
square                  lbxgh^2              935.17       0.000
identity                lbxgh                695.75       0.000
square root             sqrt(lbxgh)          574.97       0.000
log                     log(lbxgh)           455.57       0.000
1/(square root)         1/sqrt(lbxgh)        340.42       0.000
inverse                 1/lbxgh              234.20       0.000
1/square                1/(lbxgh^2)           85.01       0.000
1/cubic                 1/(lbxgh^3)          138.20       0.000
```

For serum creatinine





Checking assumptions for hypothesis test in urine albumin

Variable	Obs	W	V	z	Prob>z
urxums	868	0.13039	482.140	15.211	0.00000

```
. swilk urxums if hba1c==2
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
urxums	289	0.18575	167.828	12.004	0.00000

```
. swilk uruxms if hba1c==3
```

```
. swilk urxums if hba1c ==3
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
urxums	147	0.38275	70.588	9.640	0.00000

```
. ladder urxums if hba1c==1
```

Transformation	formula	chi2(2)	P(chi2)
cubic	urxums^3	.	.
square	urxums^2	.	.
identity	urxums	.	.
square root	sqrt(urxums)	828.08	0.000
log	log(urxums)	123.54	0.000
1/(square root)	1/sqrt(urxums)	308.49	0.000
inverse	1/urxums	830.13	0.000
1/square	1/(urxums^2)	.	.
1/cubic	1/(urxums^3)	.	.

```
. ladder urxums if hba1c ==2
```

Transformation	formula	chi2(2)	P(chi2)
cubic	urxums^3	392.84	0.000
square	urxums^2	376.88	0.000
identity	urxums	333.80	0.000
square root	sqrt(urxums)	245.16	0.000
log	log(urxums)	37.83	0.000
1/(square root)	1/sqrt(urxums)	41.74	0.000
inverse	1/urxums	134.00	0.000
1/square	1/(urxums^2)	263.54	0.000
1/cubic	1/(urxums^3)	337.71	0.000

```
. ladder urxums if hba1c==3
```

Transformation	formula	chi2(2)	P(chi2)
cubic	urxums^3	172.12	0.000
square	urxums^2	147.62	0.000
identity	urxums	111.43	0.000
square root	sqrt(urxums)	76.54	0.000
log	log(urxums)	11.60	0.003
1/(square root)	1/sqrt(urxums)	31.51	0.000
inverse	1/urxums	96.19	0.000
1/square	1/(urxums^2)	177.42	0.000
1/cubic	1/(urxums^3)	207.92	0.000

```
. swilk lbxscr if hba1c==1
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lbxscr	842	0.56591	234.107	13.415	0.00000

```
. swilk lbxscr if hba1c==2
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lbxscr	278	0.57898	83.850	10.358	0.00000

```
. swilk lbxscr if hba1c ==3
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
lbxscr	145	0.88058	13.497	5.889	0.00000

. ladder lbxscr if hba1c==1

Transformation	formula	chi2(2)	P(chi2)
cubic	lbxscr^3	.	.
square	lbxscr^2	.	.
identity	lbxscr	914.71	0.000
square root	sqrt(lbxscr)	524.46	0.000
log	log(lbxscr)	165.91	0.000
1/(square root)	1/sqrt(lbxscr)	43.52	0.000
inverse	1/lbxscr	134.20	0.000
1/square	1/(lbxscr^2)	436.99	0.000
1/cubic	1/(lbxscr^3)	769.95	0.000

. ladder lbxscr if hba1c==2

Transformation	formula	chi2(2)	P(chi2)
cubic	lbxscr^3	444.94	0.000
square	lbxscr^2	417.03	0.000
identity	lbxscr	282.76	0.000
square root	sqrt(lbxscr)	168.93	0.000
log	log(lbxscr)	66.55	0.000
1/(square root)	1/sqrt(lbxscr)	10.90	0.004
inverse	1/lbxscr	10.54	0.005
1/square	1/(lbxscr^2)	65.47	0.000
1/cubic	1/(lbxscr^3)	126.25	0.000

. ladder lbxscr if hba1c==3

Transformation	formula	chi2(2)	P(chi2)
cubic	lbxscr^3	96.73	0.000
square	lbxscr^2	70.29	0.000
identity	lbxscr	36.30	0.000
square root	sqrt(lbxscr)	18.98	0.000
log	log(lbxscr)	5.75	0.056
1/(square root)	1/sqrt(lbxscr)	0.27	0.874
inverse	1/lbxscr	8.48	0.014
1/square	1/(lbxscr^2)	43.65	0.000
1/cubic	1/(lbxscr^3)	81.08	0.000

References

Kimmel, P. L., & Rosenberg, M. E. (2015). *Chronic renal disease*. San Diego, California ;: Academic Press.

Molitch, M. E., DeFronzo, R. A., Franz, M. J., Keane, W. F., & et al. (2004). Nephropathy in diabetes. *Diabetes Care, 27*, S79-83. doi:http://dx.doi.org/10.2337/diacare.27.2007.S79

Krishnamurti, U., & Steffes, M. W. (2001). Glycohemoglobin: A primary predictor of the development or reversal of complications of diabetes mellitus. *Clinical Chemistry (Baltimore, Md.); Clin Chem, 47*(7), 1157-1165. doi:10.1093/clinchem/47.7.1157

Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Top Ten Leading Causes of Death in the U.S. for Ages 1-44 from 1981-2018. Injuries and Violence Are Leading Causes of Death (cdc.gov)