N741 Final Project: Urbanization and Suddent Infant Death

Taylor Thul

2017-04-19

## Abstract

*TYPE IN YOUR TEXT HERE - DELETE THIS FOR FINAL REPORT: Please provide an abstract with only a sentence or two each on these items: the purpose of your project, your research questions, analysis methods, results, and conclusions (approximately 300 words).*

## Introduction

\_As a nurse, my background is in the neonatal intensive care unit (NICU). In my capacity as a bedside nurse I often discharge patients to home, a process that includes extensive education regarding prevention of sudden infant death syndrome (SIDS). SIDS, also known as crib or cot death, is the sudden, unexplained death of an infant younger than 1 year old after thorough investigation (CDC, 2017). SIDS is the leading cause of death in children one month to one year of age. It is also included as a cause of death in sudden unexpected infant death (SUID), a category that also includes accidental suffocation and strangulation in bed as well as those deaths where the cause remains unknown without a complete investigation (CDC, 2017). Of the 3,7000 SIUD in the United States in 2015, 43% were attributed to SIDS and 32% of unknown cause (CDC, 2017). SUID is a significant health concern and has been the focus of much research and major public health efforts. Many risk factors have been identified including: prone sleeping, bed sharing, loose bedding, maternal smoking, and premature birth (Carlin & Moon, 2017). A public education campaign, Safe to Sleep, launched in 1994 had great success in lowering the rate of SIDS death by more than 50% (NIH, 2017). However since that time the rates have remained relatively unchanged and there are no clear answers regarding the etiology of the disease. \_As a researcher, my interests focus on maternal/ infant health so I often encounter vital statistics and other data on SIDS reporting on a variety of variables. However, I often wonder what other variables not reported may influence SIDS rates. One is particular is if urbanization is at all associated with SIDS rates. I would postulate more urban settings would facilitate greater exposure to education campaigns and other resources but also could increase exposure to air pollution and communicable diseases. There could, of course, be many confounding variables in the proposed relationship but the first step is establishing if any association exists. \_Therefore the purpose of this paper is to explore any association between urbanization and sudden infant death syndrome (SIDS).

## Dataset Description and Summary

\_To explore the relationship between SIDS and urbanization I will utilize the CDC Multiple Cause of Death database. It contains mortality for all US counties based on the death certificate and includes demographic data as well as ICD-10 codes for cause of death and urbanization level. Unfortunately the CDC Wonder website does not allow access to the entire dataset rather some variables must be selected and a subset of data will be generated for use. To have access to the 2013 urbanization level I was not also able to have the 2006 urbanization levels or demographic variables such as ethnicity. My final subset included:

\*2013 Urbanization Levels +Large Central Metro +Large Fringe Metro +Medium Metro +Small Metro +Micropolitan +NonCore (non-metro)

\*Age Groups +< 1 day (Group 1) +1-6 days (Group 2) +7-27 days (Group 3) +28- 364 days (Group 4)

*ICD-10 Cause of Death* Number of Deaths *Population* Crude Death Rate

\_2013 urbanization levels were selected as they were the most current and the age category was chosen to be consistent with the definition of SIDS. The urbanization variable is categorical and includes large central metro, large fringe metro, medium metro, small metro, and micropolitan. The multilevel categorization is well suited for health related analyses because it differentiates counties within large metropolitan areas into “central” and “fringe” to capture the possible differences between a true urban environment compared to highly populous suburban areas. The ICD-10 cause of death will give the most detailed way to extract death by SIDS as opposed to other causes. I then did some data cleaning for the entire dataset including creating categorical variables for age group and urbanization level. There was some missing data (labeled as "unavailable") in the crude death rates. Thankfully there was no missing data in overall number of deaths. I chose to omit the missing data from the analysis containing the crude deaths variables since the raw numbers would still be represented in number of deaths. The deaths and crude rate variable was greatly skewed to the right so I performed a log10 transformation on both of them. Both were far more normally distributed after so the transformed variable was used for the remainder of analysis. I created a subset of data isolating SIDS as a cause of death using the ICD-10 code R95. This resulted in a small amount of observations (n=23) which seemed insufficient for the scope of this project. Therefore I chose to modify my questions and look at the association between urbanization and overall infant deaths then more specifically death as a result of SIDS. Additionally as many SUID deaths are unexplained including all causes of death may capture some of them as well, for example 'other and unspecified accidents.' \_Overall the data was fairly clean and analysis could begin. After some initial exploration it was noted there were some extreme outlines in the 'one day' age category. Returning to the data I noted multiple large, outlining variables associated with prenatal and childbirth related deaths, such as 'certain conditions originating in the perinatal period.' This cause of death is much more related to the prenatal period than infancy therefore I chose to eliminate the '1 day' age category from the overall analysis to avoid confounding with prenatal and birth related deaths.

The CDC public databases located at: [https://wonder.cdc.gov/controller/datarequest/D77;jsessionid=4DC93A9C344C5A163A529948737DB795](https://wonder.cdc.gov/controller/datarequest/D77;jsessionid=4DC93A9C344C5A163A5%2029948737DB795)

## Dataset Summary and Visualization

# Summary of Overall Dataset

## [1] "2013 Urbanization"   
## [2] "2013 Urbanization Code"   
## [3] "Infant Age Groups"   
## [4] "Infant Age Groups Code"   
## [5] "UCD - ICD-10 130 Cause List (Infants)"   
## [6] "UCD - ICD-10 130 Cause List (Infants) Code"  
## [7] "Deaths"   
## [8] "Population"   
## [9] "Crude Rate"

## 2013 Urbanization 2013 Urbanization Code Infant Age Groups   
## Length:2331 Min. :1.000 Length:2331   
## Class :character 1st Qu.:2.000 Class :character   
## Mode :character Median :3.000 Mode :character   
## Mean :3.331   
## 3rd Qu.:5.000   
## Max. :6.000   
## Infant Age Groups Code UCD - ICD-10 130 Cause List (Infants)  
## Length:2331 Length:2331   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
## UCD - ICD-10 130 Cause List (Infants) Code Deaths   
## Length:2331 Min. : 10.0   
## Class :character 1st Qu.: 30.0   
## Mode :character Median : 84.0   
## Mean : 528.1   
## 3rd Qu.: 287.0   
## Max. :50934.0   
## Population Crude Rate   
## Min. : 3837310 Length:2331   
## 1st Qu.: 5742486 Class :character   
## Median :14379025 Mode :character   
## Mean :12218310   
## 3rd Qu.:16098751   
## Max. :22969673

# Summary of SIDS Subdata

## urbanization urban.code Infant Age Groups age.group   
## Length:23 Min. :1.000 Length:23 Length:23   
## Class :character 1st Qu.:2.000 Class :character Class :character   
## Mode :character Median :3.000 Mode :character Mode :character   
## Mean :3.391   
## 3rd Qu.:5.000   
## Max. :6.000   
##   
## UCD - ICD-10 130 Cause List (Infants) icd Deaths   
## Length:23 Length:23 Min. : 14   
## Class :character Class :character 1st Qu.: 42   
## Mode :character Mode :character Median : 263   
## Mean :1566   
## 3rd Qu.:1789   
## Max. :8746   
##   
## Population crude.rate age.cat   
## Min. : 3837310 Min. :0.0000 Length:23   
## 1st Qu.: 5742486 1st Qu.:0.0000 Class :character   
## Median :14379025 Median :0.0000 Mode :character   
## Mean :11862892 Mean :0.1714   
## 3rd Qu.:16098751 3rd Qu.:0.4000   
## Max. :22969673 Max. :0.8000   
## NA's :2

The mean number of deaths is

## [1] 528.1167

The mean population is

## [1] 12218310

### Summary Statistics Tables

Table of Summary Stats for Numeric Variables in Causes of Death Data

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| item | n | missing | min | avg | SD | median | Q1 | Q3 | max |
| Crude Rate | 1989 | 342 | 0 | 4.665660e-02 | 1.665278e-01 | 0 | 0 | 0 | 2.2 |
| Deaths | 2331 | 0 | 10 | 5.281167e+02 | 2.095874e+03 | 84 | 30 | 287 | 50934.0 |
| City Population | 2331 | 0 | 3837310 | 1.221831e+07 | 6.937328e+06 | 14379025 | 5742486 | 16098751 | 22969673.0 |

Table of Summary Stats for Numeric Variables in SIDS Deaths

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| item | n | missing | min | avg | SD | median | Q1 | Q3 | max |
| Crude Rate | 21 | 2 | 0 | 1.714286e-01 | 2.704494e-01 | 0 | 0 | 0.4 | 0.8 |
| Deaths | 23 | 0 | 14 | 1.565826e+03 | 2.621897e+03 | 263 | 42 | 1789.0 | 8746.0 |
| City Population | 23 | 0 | 3837310 | 1.186289e+07 | 6.955833e+06 | 14379025 | 5742486 | 16098751.0 | 22969673.0 |

Frequency Table for Urbanization

|  |  |  |
| --- | --- | --- |
| urbanization | freq | pct |
| Large Central Metro | 451 | 19.34792 |
| Large Fringe Metro | 410 | 17.58902 |
| Medium Metro | 413 | 17.71772 |
| Micropolitan (non-metro) | 363 | 15.57272 |
| NonCore (non-metro) | 332 | 14.24281 |
| Small Metro | 362 | 15.52982 |

Frequency Table for Infant Age Groups

|  |  |  |
| --- | --- | --- |
| Infant Age Groups | freq | pct |
| < 1 day | 513 | 22.00772 |
| 1-6 days | 524 | 22.47962 |
| 28-364 days | 731 | 31.35993 |
| 7-27 days | 563 | 24.15272 |
| #For Frequency Table | of ICD | 10 Cause List of Death see Appendix 1 |

Frequency Table for Urbanization of SIDS Data

|  |  |  |
| --- | --- | --- |
| urbanization | freq | pct |
| Large Central Metro | 4 | 17.39130 |
| Large Fringe Metro | 4 | 17.39130 |
| Medium Metro | 4 | 17.39130 |
| Micropolitan (non-metro) | 4 | 17.39130 |
| NonCore (non-metro) | 3 | 13.04348 |
| Small Metro | 4 | 17.39130 |

Frequency Table for Infant Age Groups of SIDS Data

|  |  |  |
| --- | --- | --- |
| Infant Age Groups | freq | pct |
| < 1 day | 5 | 21.73913 |
| 1-6 days | 6 | 26.08696 |
| 28-364 days | 6 | 26.08696 |
| 7-27 days | 6 | 26.08696 |
| #Aggregate Data Tabl | es |  |

Deaths By Urbanization and Age Group

|  |  |  |
| --- | --- | --- |
| urbanization | age.group | Deaths |
| Large Central Metro | 1-6d | 476.1038 |
| Large Fringe Metro | 1-6d | 342.3913 |
| Medium Metro | 1-6d | 349.5714 |
| Micropolitan (non-metro) | 1-6d | 155.1481 |
| NonCore (non-metro) | 1-6d | 124.6389 |
| Small Metro | 1-6d | 164.1098 |
| Large Central Metro | 1d | 2095.6667 |
| Large Fringe Metro | 1d | 1420.9783 |
| Medium Metro | 1d | 1408.6154 |
| Micropolitan (non-metro) | 1d | 603.8718 |
| NonCore (non-metro) | 1d | 431.7297 |
| Small Metro | 1d | 666.7975 |
| Large Central Metro | 28-364d | 786.9708 |
| Large Fringe Metro | 28-364d | 513.2283 |
| Medium Metro | 28-364d | 588.8968 |
| Micropolitan (non-metro) | 28-364d | 293.6293 |
| NonCore (non-metro) | 28-364d | 224.9908 |
| Small Metro | 28-364d | 299.5517 |
| Large Central Metro | 7-27d | 442.9174 |
| Large Fringe Metro | 7-27d | 297.7879 |
| Medium Metro | 7-27d | 292.8000 |
| Micropolitan (non-metro) | 7-27d | 142.4545 |
| NonCore (non-metro) | 7-27d | 115.4416 |
| Small Metro | 7-27d | 161.1176 |

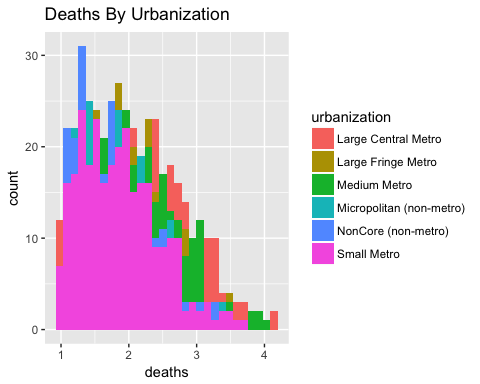
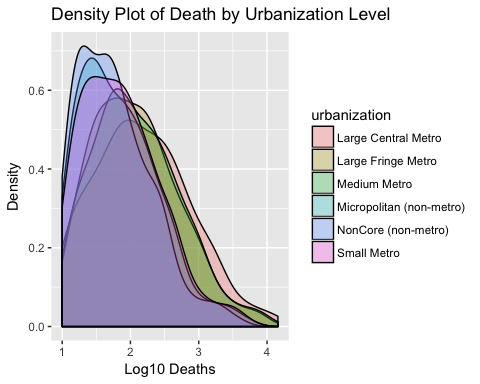
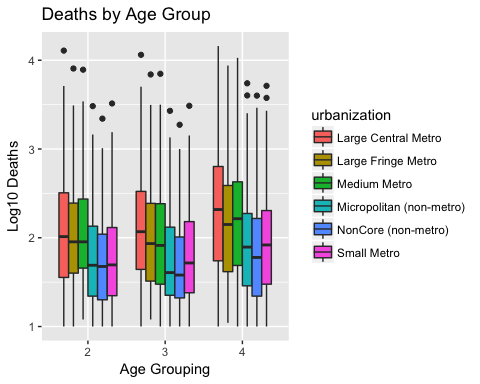
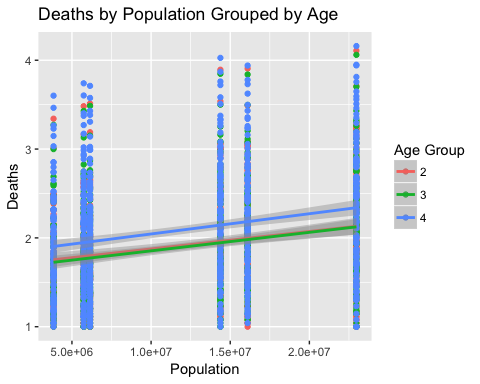
Crude Death Rate By Urbanization and Age Group

|  |  |  |
| --- | --- | --- |
| urbanization | age.group | crude.rate |
| Large Central Metro | 1-6d | 0.0193182 |
| Large Fringe Metro | 1-6d | 0.0189873 |
| Medium Metro | 1-6d | 0.0209877 |
| Micropolitan (non-metro) | 1-6d | 0.0281250 |
| NonCore (non-metro) | 1-6d | 0.0363636 |
| Small Metro | 1-6d | 0.0261538 |
| Large Central Metro | 1d | 0.0988506 |
| Large Fringe Metro | 1d | 0.0987179 |
| Medium Metro | 1d | 0.1050633 |
| Micropolitan (non-metro) | 1d | 0.1153846 |
| NonCore (non-metro) | 1d | 0.1375000 |
| Small Metro | 1d | 0.1311475 |
| Large Central Metro | 28-364d | 0.0290076 |
| Large Fringe Metro | 28-364d | 0.0254237 |
| Medium Metro | 28-364d | 0.0350427 |
| Micropolitan (non-metro) | 28-364d | 0.0525773 |
| NonCore (non-metro) | 28-364d | 0.0632184 |
| Small Metro | 28-364d | 0.0474227 |
| Large Central Metro | 7-27d | 0.0153061 |
| Large Fringe Metro | 7-27d | 0.0131868 |
| Medium Metro | 7-27d | 0.0141304 |
| Micropolitan (non-metro) | 7-27d | 0.0200000 |
| NonCore (non-metro) | 7-27d | 0.0258065 |
| Small Metro | 7-27d | 0.0211268 |

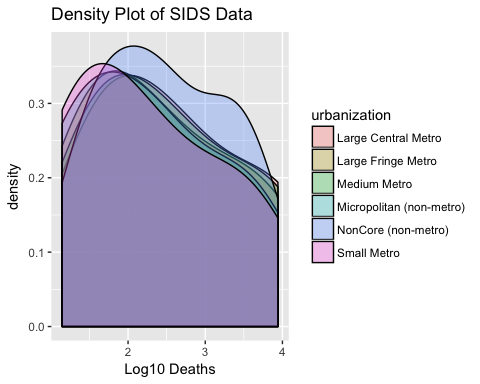
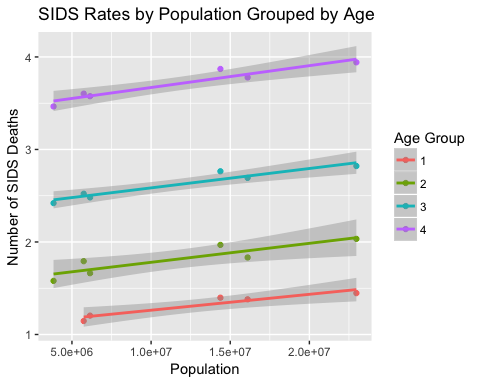
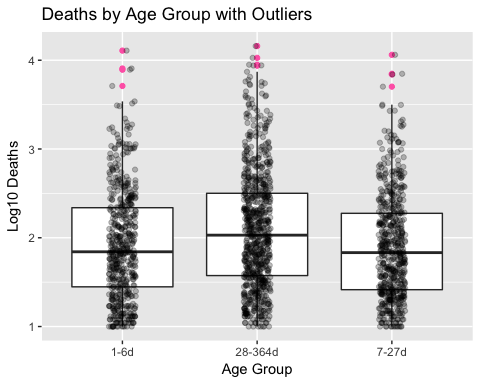
SIDS Deaths By Urbanization and Age Group

|  |  |  |
| --- | --- | --- |
| urbanization | age.group | crude.rate |
| Large Central Metro | 1-6d | 0.0 |
| Large Fringe Metro | 1-6d | 0.0 |
| Medium Metro | 1-6d | 0.0 |
| Micropolitan (non-metro) | 1-6d | 0.0 |
| NonCore (non-metro) | 1-6d | 0.0 |
| Small Metro | 1-6d | 0.0 |
| Large Central Metro | 1d | 0.0 |
| Large Fringe Metro | 1d | 0.0 |
| Medium Metro | 1d | 0.0 |
| Large Central Metro | 28-364d | 0.4 |
| Large Fringe Metro | 28-364d | 0.4 |
| Medium Metro | 28-364d | 0.5 |
| Micropolitan (non-metro) | 28-364d | 0.7 |
| NonCore (non-metro) | 28-364d | 0.8 |
| Small Metro | 28-364d | 0.6 |
| Large Central Metro | 7-27d | 0.0 |
| Large Fringe Metro | 7-27d | 0.0 |
| Medium Metro | 7-27d | 0.0 |
| Micropolitan (non-metro) | 7-27d | 0.1 |
| NonCore (non-metro) | 7-27d | 0.1 |
| Small Metro | 7-27d | 0.0 |

### Graphic Representation of Data



## $x  
## [1] "Log10 Deaths"  
##   
## attr(,"class")  
## [1] "labels"



## Methods and Analysis

Due to the smaller amount of data present when restricted to the subset of the SIDS ICD-10 code I chose to run analysis on both the overall data set and the subset. Therefore my question then became if there was any association between urbanization level and infant death and more specifically death as a result of SIDS. Since this was a continuous y and categorical x I chose to use ANOVA as the primary analysis method. I assumed the transformed data would behave as normal.

# Output of Model for SIDS Subset of Data

## Anova Table (Type II tests)  
##   
## Response: crude.rate  
## Sum Sq Df F value Pr(>F)   
## age.cat 2.482 3 31.0186 <2e-16 \*\*\*  
## urbanization 0.180 5 1.3498 0.2405   
## age.cat:urbanization 0.047 15 0.1166 1.0000   
## Residuals 52.417 1965   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Large Central Metro Large Fringe Metro Medium Metro  
## 1 0.09885057 0.09871795 0.10506329  
## 2 0.01931818 0.01898734 0.02098765  
## 3 0.01530612 0.01318681 0.01413043  
## 4 0.02900763 0.02542373 0.03504274  
## Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1 0.11538462 0.13750000 0.13114754  
## 2 0.02812500 0.03636364 0.02615385  
## 3 0.02000000 0.02580645 0.02112676  
## 4 0.05257732 0.06321839 0.04742268

## Large Central Metro Large Fringe Metro Medium Metro  
## 1 0.31047385 0.28580876 0.30377908  
## 2 0.07248932 0.06616985 0.06655083  
## 3 0.05981422 0.05208709 0.06036796  
## 4 0.08547474 0.07755499 0.10365413  
## Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1 0.29220646 0.32949686 0.32173539  
## 2 0.07861651 0.09499247 0.07761468  
## 3 0.06936566 0.07879312 0.06951977  
## 4 0.14368135 0.16075038 0.12507730

## urbanization  
## age.cat Large Central Metro Large Fringe Metro Medium Metro  
## 1 99 92 91  
## 2 106 92 91  
## 3 109 99 105  
## 4 137 127 126  
## urbanization  
## age.cat Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1 78 74 79  
## 2 81 72 82  
## 3 88 77 85  
## 4 116 109 116

With the inclusion of age, urbanization is not a statistically significant variable in SIDS death. Age is significant, which makes sense because the age categories are grouped in a very uneven manner. Next I will use ANOVA to analyze the full dataset (without day 1 death data).

# Output of Full Data Analysis

## Anova Table (Type II tests)  
##   
## Response: Deaths  
## Sum Sq Df F value Pr(>F)   
## age.group 18299932 2 9.4369 8.374e-05 \*\*\*  
## urbanization 42920048 5 8.8532 2.583e-08 \*\*\*  
## age.group:urbanization 2884598 10 0.2975 0.9819   
## Residuals 1745274447 1800   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Large Central Metro Large Fringe Metro Medium Metro  
## 1-6d 476.1038 342.3913 349.5714  
## 28-364d 786.9708 513.2283 588.8968  
## 7-27d 442.9174 297.7879 292.8000  
## Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1-6d 155.1481 124.6389 164.1098  
## 28-364d 293.6293 224.9908 299.5517  
## 7-27d 142.4545 115.4416 161.1176

## Large Central Metro Large Fringe Metro Medium Metro  
## 1-6d 1386.060 921.4762 912.7184  
## 28-364d 1849.207 1178.6038 1409.3510  
## 7-27d 1255.365 790.9188 787.8284  
## Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1-6d 377.4739 285.6727 403.1551  
## 28-364d 728.0352 543.0515 706.2416  
## 7-27d 333.1387 250.0974 377.1629

## urbanization  
## age.group Large Central Metro Large Fringe Metro Medium Metro  
## 1-6d 106 92 91  
## 28-364d 137 127 126  
## 7-27d 109 99 105  
## urbanization  
## age.group Micropolitan (non-metro) NonCore (non-metro) Small Metro  
## 1-6d 81 72 82  
## 28-364d 116 109 116  
## 7-27d 88 77 85

In this analysis both age and urban level are significant variables.

## Discussion and Conclusions

After analysis there is not enough evidence to suggest an association between SIDS and urbanization levels. However there is evidence that shows an association between all types of infant deaths and urbanization levels. As this was not the original aim of the project it could be considered an incidental finding. Perhaps the same factors hypothesizes as contributing to the theorized relationship between SIDS and urban level could have the same effect on overall infant deaths. More urban areas could have higher risk of communicable diseases and more exposure to air pollutants, perhaps increasing risk for respiratory associated deaths. However urban areas also often provide more accessibility to healthcare and public health education campaigns. I would be interested in exploring this as a primary question in further analysis and investigating the direction of the relationship.

## Limitations

There were many limitations to this analysis. Access to the dataset was limited from the beginning, forcing exclusion of possibly significant variables such as ethnicity and presence of risk factors. Furthermore once the SIDS specific data was manipulated it was far too small to do a quality analysis. Use of a more varied and larger dataset would have vastly increased the validity of the findings.

## References

Carlin RF, Moon RY. Risk Factors, Protective Factors, and Current Recommendations to Reduce Sudden Infant Death Syndrome A Review. JAMA Pediatric. 2017;171(2):175-180. <doi:10.1001/jamapediatrics.2016.3345>

Centers for Disease Control and Prevention (2017 February 1). Sudden Unexpected Infant Death and Sudden Infant Death Syndrome Data and Statistics. Retrieved from <https://www.cdc.gov/sids/data.htm>

Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

## Appendix

# Appendix 1

Frequency Table for ICD 10 Cause List of Death

|  |  |  |
| --- | --- | --- |
| UCD - ICD-10 130 Cause List (Infants) | freq | pct |
| #Sudden infant death syndrome (R95) | 23 | 100 |