Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ ______________

Maraia Tremarelli Date

Global Mortality Attributed to Tropical Cyclones: A Prediction Analysis of Historical Data of Australia, Taiwan, and the United States from 1987 to 2016

By

Maraia Tremarelli

Master of Science in Public Health

Global Epidemiology

A. Cecile J. W. Janssens, PhD

Committee Chair

Mark E. Keim, MD, MBA

Committee Member

Global Mortality Attributed to Tropical Cyclones: A Prediction Analysis of Historical Data of Australia, Taiwan, and the United States from 1987 to 2016

By

Maraia Tremarelli

Bachelor of Science

Rhodes College

2015

Faculty Thesis Advisor: A. Cecile J. W. Janssens, PhD

Thesis Field Advisor: Mark E. Keim, MD, MBA

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Global Epidemiology

2018

Abstract

Global Mortality Attributed to Tropical Cyclones: A Prediction Analysis of Historical Data of Australia, Taiwan, and the United States from 1987 to 2016

By Maraia Tremarelli

Introduction: Disasters are tragic events that affect populations around the world. Tropical cyclones alone have caused more than 1.3 million deaths since 1900 due to their detrimental impact. Risk assessments are the principal tools used to predict future risks of disasters, but they do not consider a great deal of quantitative data, creating insecurity in these predictions. This produces a need to better express the probability of tropical cyclone events and the estimated mortality associated using historical data of these events.

Methods: To estimate past tropical storm events, the Centre for Research on the Epidemiology of Disasters (CRED)'s Emergency Events Database (EM-DAT) was consulted. Storms that hit Australia, Taiwan and the United States from 1987 to 2006 were chosen to have a variety of locations. Values that contained magnitude of the storm, associated deaths and people affected were used as the main predictors to construct the model.

Results: After evaluating the dataset, it was found that the data was unreliable due to an inconsistency of reported data. It was apparent that the discrepancies were too great to construct prediction models due to 55% of the magnitude entries were missing and 49% of the 'Total affected' data were missing. Cross-referencing the data with other agencies' datasets delivered varying reports as well.

Discussion: When analyzing all the data further, it was found that the dataset was unable to provide sufficient results to produce a reliable prediction model. The inconsistencies among different datasets when cross-referencing proves that there is a need to have set criteria that is followed by those who report on disaster data for dependability. Definitions of data are important to better understand where the data comes from and hoe different databases report and present their data. These limitations inhibit the creation of a reliable prediction model of mortality.

Global Mortality Attributed to Tropical Cyclones: A Prediction Analysis of Historical Data of Australia, Taiwan, and the United States from 1987 to 2016

By

Maraia Tremarelli

Bachelor of Science

Rhodes College

2015

Faculty Thesis Advisor: A. Cecile J. W. Janssens, PhD

Thesis Field Advisor: Mark E. Keim, MD, MBA

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Global Epidemiology

2018

TABLE OF CONTENTS

INTRODUCTION

Disasters are tragic events that happen worldwide and are serious threats to human life and development. Their impacts can range from death and injury to structural damage and economic burden [1]. A disaster is caused by the hazard and the vulnerability of an area, and with disasters increasing worldwide due to the increase of vulnerable populations, it is important to assess the risk of disasters to better prepare and respond. However, there is little information regarding the health risk, or incidence of disease per event, of a population during disasters. Alone, tropical cyclones have been attributed to more than 1.3 million deaths since 1900 with more than 629 million people being affected [2]. By using historical data of mortality from the impact of tropical cyclones, exposure and capacity of an area can be assessed and used to determine the expected mortality rate from the impact of a cyclone.

Risk assessments can help predict future trends of these storms in order to help prevent public health consequences of this disaster, including mortality. A risk assessment's intent is prevention and reduction [3], and in this instance, of public health consequences of disasters. Currently, there are many risk assessments that are used for predicting future risks of disasters, but most use either qualitative or semi-quantitative data, leaving a great deal of uncertainty in the predictions [3]. According to a study conducted by Jonkman and Kelman, there has been a great deal of research dedicated to long-term hazardous events from disasters [4]. However, mortality studies that look at countries with frequent tropical cyclone landfalls have seldom been researched to see why deaths keep occurring, even with the advances in warnings and tracking of these storms [5].

This study will focus on quantitative data to better express the probability of a tropical cyclone occurring for a country and the estimated risk of mortality for each landfall. To assess the predictability of tropical cyclones, historical data of deaths by tropical cyclones were used to directly calculate the probability of disaster-related health impacts. If proven valid and informative, these calculations can be used for better predicting the risk of tropical cyclones and help better comprehend mortality rates that are associated and minimize or prevent mortality in countries that experience tropical cyclone landfalls.

Overview of Disasters

According the United Nations International Strategy for Disaster Research (2009), a disaster is defined as "a serious disruption of the functioning of a community or a society causing widespread human, material, economic or environmental losses and impacts, which exceed the ability of the affected community or society to cope using its own resources" [6]. An easier way to describe a disaster is a mismatch of needs and resources; the need becomes too great and the resources become too few [7]. An event becomes a disaster when there is a hazard and a vulnerability concurrently. A hazard is a "a dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage" and a vulnerability is "the characteristic and circumstance of a community, system or asset that makes it susceptible to the damaging effects of a hazard" [6]. Furthermore, Mark E. Keim (2016) writes that the impacts of disasters can include death, injury, disease and various negative effects on a human's physical, mental and social well-being [8]. Additionally, damage to property, economies and loss of services are frequently encountered. The threat of disaster is increasing worldwide due to the rise of vulnerabilities. Some vulnerabilities include more people moving to disaster prone areas, the development of toxic chemicals and increased industrial activity, and the rapid industrialization of cities in developing nations [7, 9].

Disasters can be classified based upon time, presence of advance notice, size of response or type of hazard [10]. Hazard based classifications are then broken down to natural, technological, and hybrid, which is a combination of natural and technological. Natural disasters are then classified into hydro-meteorological, geological and biological, while technological disasters are classified into thermal, technological and mechanical [7]. Man-made disasters mostly fall in the category of technological, due to the nature of disasters.

Overview of Tropical Cyclones

Tropical cyclones are one of the most devastating weather events that happen in different parts of the world. A tropical cyclone is a type of environmental disaster that is categorized as a hydro-meteorological disaster under the hazards-based classification of natural disasters. They are defined as low-pressure weather systems that develop over warm ocean water [11]. Since warm water is essential for the formation of a tropical cyclone, they typically form near the equator between the latitudes of 30° North and 30° South [12, 13]. Within these regions, there are seven distinct tropical cyclone basins: (1) Atlantic Ocean; (2) northeast Pacific Ocean; (3) northwest Pacific Ocean; (4) north Indian Ocean; (5) southwest Indian Ocean; (6) Australian/southeast Indian Ocean; and (7) Australian/southwest Pacific Ocean [13]. There must also be rotating winds that sustain a speed of at least 74 miles per hour (118 kilometers per hour) to be considered a tropical cyclone [12]. There are six characteristics that must happen concurrently for there to be a successful tropical cyclone formation, otherwise known as cyclogenesis: warm ocean water of at least 79.7° Fahrenheit with an ocean depth of at least 150 feet; an atmosphere that quickly cools so warm air continues to rise; moisture present in the mid-troposphere for cloud formation; strong Coriolis forces; a pre-existing, organized condition with a rotating system; and enough wind shear to sustain the rotating winds, but not too strong where the heat and moisture are stripped away [12, 13, 14].

The wind circulation of these cyclones is controlled by a balance between pressure gradient and the Coriolis effect, with the bigger difference between the pressures, the faster the winds [12]. The Coriolis effect is also the reason for the differing rotating winds in the northern and southern hemisphere, with winds rotating counterclockwise in the northern hemisphere and winds rotating clockwise in the southern hemisphere [12, 13]. Since the winds differ in their rotation direction on opposite sides of the equator, the cyclone cannot cross the equator when it has formed, creating two cyclone belts: one above the equator and one below [13]. Roughly two-thirds of all tropical cyclones form in the northern hemisphere [12]. Besides the difference between the pressures, cyclones are driven by the energy of warm ocean water and the air above. The greater the difference in air and ocean temperature, the greater transfer of energy is given to the cyclone, increasing the potential for a more severe storm [11].

Cyclones' intensity is measured by the Saffir-Simpson Scale and is based on wind speed. The scale is categorized on a scale of one to five, with one being the least intense and five being the most severe. The main use of this scale is to estimate the amount of damage that a cyclone will produce. Table 1 describes the effects of the Saffir-Simpson scale [11, 15]. Additionally, tropical cyclones are known by several names depending on geographic location. The North Atlantic, the Caribbean, the Gulf of Mexico and the eastern North Pacific know cyclones as 'hurricanes'; the western Pacific know them as 'typhoons'; and the Indian Ocean and Australia area know them as 'cyclones' [12, 16]. Between 1980-2000, there was an average of 46 tropical cyclone impacts on countries around the world [13]. This shows that tropical cyclones are a continuing force and one that has a great impact on populations worldwide. Many people are vulnerable to these disasters because of their proximity to coastlines that are at risk.

Additionally, tropical cyclones can be portrayed in what are called hazard maps. A hazard map is a map used for natural disasters to display areas that are or have been affected and vulnerable to specific disasters [17]. The purpose of hazard maps is to inform people in these areas be aware of the potential disasters that could affect their property and well-being. They are tools for prevention to spark preventative or tabletop activities to be performed, and to prepare residents for potential evacuation [17]. Figures 1[1] and 2[1] are examples of different types of hazard maps that depict vulnerable areas to tropical cyclones [18]. Figure 1 shows all tropical cyclone tracks from around the world between the years of 1987 to 2016. This way of representing tropical cyclones is a great way to show where the most intense storms have been in the past 30 years because it is broken down by intensity by wind speed according to the Saffir-Simpson scale (Table 1). It is evident that areas around the Gulf of Mexico and off the western coast of Mexico had the most intense storms during this time period, as seen with category 5 storms shown. Alternatively, Figure 2 depicts the density of all tropical cyclone tracks from around the world during the same time period. Here, it is apparent that southeast Asia and the west coast of Mexico experienced the most storms during this time.

Public Health Consequences of Tropical Cyclones

There are impending public health consequences of tropical cyclones, including mortality, morbidity, and public concern for safety. Damage to homes and businesses can also cause displacement and economic strain [13]. However, mortality and morbidity differ

---

[1] Data for Figures 1 and 2 were obtained from Dr. Dave Eslinger, an oceanographer from the National Oceanic and Atmospheric Administration (NOAA). He provided shapefiles of all tropical cyclone tracks from 1851 to 2017.

among countries of different development. Developing countries, or areas of low-income, that lack an advanced warning system or evacuation routes have higher chances of dying from drowning due to the storm's related storm surge. This accounts for approximately 90% of all cyclone attributable deaths [8, 10]. Developed nations have lower rates of mortality and majority of them occur after the cyclones hits during the recovery phase [8]. Over the past two centuries, cyclones have been responsible for nearly 2 million deaths globally, with an average of about 11,800 deaths occurring every year during the years 1980-2000 [8, 13]. Deaths related to cyclones usually come from three major factors: winds, storm surge and associated secondary disasters [2].

Injuries are the leading cause of morbidity for tropical cyclones, mostly consisting of cuts, blunt force trauma and puncture wounds from debris. Some chronic diseases are also heightened post-cyclone like asthma and emphysema, as well as exposure to hazardous material that could have been displaced. Epidemiologic surveillance has shown that there is usually no increase in communicable diseases [8]. Infectious disease outbreaks post-cyclones are extremely rare, but occasionally arise in developing nations [12]. Long-term effects have been seen in longitudinal cyclone studies, showing that rates of suicide, child abuse, and post-traumatic stress disorders in developing nations increased after a storm affected the area [8].

Cyclones can also produce secondary disasters that can attribute to a morbidity and mortality, including but not limited to flash floods, landslides, mudslides, tornadoes and storm surges. Human-factors can also increase the rates. Poor or nonexistent building codes, structural collapse, increase in coastal land use, poor warning systems or inadequate evacuation notices, can increase mortality and morbidity rates [19]. Prevention and mitigation strategies are the strongest way to limit the amount of morbidity and mortality in areas. Some measures that can be taken include proper building design in cyclone-prone areas, forecasting, early warnings, evacuation routes, land-use planning and response and recovery plans in place [19].

Prediction of Disasters

Many studies have been done to predict the number of storms in a given season or predict the overall impact that a country could expect. However, there have been few studies that look at mortality from these events and attempt to predict the risk of a tropical cyclone season by each specific's country impact and historical record of these storms. According to the United Nations Office for Disaster Risk Reduction (UNISDR), the need for data regarding disaster mitigation and prevention is critical to better response and recovery efforts by agencies [20]. UNISDR also says that disasters are only assessed at the time of the event to try to understand how the impact affected the population at risk. This level of understanding is not valuable for prevention efforts and shows that there is a need to better comprehend historical trends of disasters and how past impacts could be foretelling for present and present events. Creating this model could have assisted with risk assessments made for tropical cyclones and portrayed the potential disaster impact, which could have benefited preparedness and risk reduction initiatives.

METHODS

The data used in this study came from the Centre for Research on the Epidemiology of Disasters (CRED)'s Emergency Events Database (EM-DAT). EM-DAT is an international disaster database that reports on both natural and technological disasters from 1900 to present [21]. The natural disasters classification can also be broken down into 6 sub-groups, followed by 15 disaster types and more than 30 disaster sub-types, while the technological disasters classification can be broken down into 3 sub-groups with 15 disaster types [22]. At least one of the following criteria must be met or known in order for any disaster to be entered into the EM-DAT database: (1) 'ten or more people reported killed'; (2) 'a hundred or more people reported affected'; (3) 'declaration of a state of emergency'; or (4) 'call for international assistance' [21]. The EM-DAT receives their information from multiple sources, including but not limited to the United Nations (UN), governments, non-governmental organizations (NGOs), insurance companies and research institutes. Additionally, these sources can conflict with one another regarding their information. In order to provide the most accurate data in the database, CRED has a ranking system for the agencies that provide information based on their credibility and trustworthiness and how complete the data is, and more often than not, data will not be entered unless at least two sources report it [23].

To obtain the data for this paper, the EM-DAT database was consulted by using the database search option 'Disaster List'. Under 'Search Criteria' the years 1987 to 2016 were entered, followed by selecting all countries under the available country list. Data was collected from the 'Natural disasters' classification with the 'Meteorological' sub-group, and the 'Storm' disaster (sub-type is indicated after). Additionally, 'Total Affected' and 'Total Deaths' were selected. This generated a table that contained every disaster indicated as

'Storm' that was entered into the database from 1987 to 2016 for all countries. The data included the 'Start date', 'End date', 'Country name', 'ISO', 'Location', 'Latitude', 'Longitude', 'Magnitude value', 'Magnitude scale', 'Disaster Type', 'Disaster subtype', 'Disaster name', 'Total deaths', 'Total affected' and 'Disaster No.' (Definitions for the labels are in Appendix I). To ensure the data only included storms that were considered tropical cyclones, the data was filtered to ensure that data was only included if it was indicated as a 'Tropical storm' in the 'Disaster subtype' column.

To uncover countries that had the greatest impact from tropical cyclones, the data was then organized to showcase the countries in order by number of tropical cyclone occurrences. Table 2 shows the top 10 countries with the highest occurrences of tropical cyclones from 1900 to 2016, according to the CRED's EM-DAT database. After investigating Table 2, the data was analyzed to pick three countries that had the most complete data in the CRED EM-DAT database and were in different parts of the world to see how different cyclone basins varied. Australia, Taiwan and the United States were designated as the three countries that were going to be used for this analysis. Although Australia was ranked 12th among the top countries with the most tropical cyclone activity, it was chosen because of the activity of these storms in its basin. Looking at Figures 1 and 2, it is evident that areas around Australia are very active and it was important to have a country that could represent this activity in the analysis. Besides Australia, Taiwan and the Unites States having some of the most complete overall data, they were also chosen based on the credibility of their national or regional meteorological agencies that contain additional reports to fill in the missing pieces.

To devise the prediction models for mortality, the predictors to be used were 'Magnitude value' of the tropical cyclone (categories based on the Saffir-Simpson Scale,

Table 1) and the rate of mortality based on the tropical cyclone occurrence, which would be calculated as the 'Total deaths' divided by the 'Total affected'. However, a great deal of data were still missing from these columns from the EM-DAT database. The original dataset also made it unclear if the storm had made landfall in the designated country or if the country was just feeling outside affects from outer bands. For the purpose of this study, it was imperative to understand the track to see if the storm had a direct impact (landfall) on the country. One storm could have also impacted several countries, both directly and indirectly. Related to understanding the track of the storm is understanding the magnitude value of the storm. It was not indicated at what point of the cyclone's track the magnitude was taken. Tropical cyclones' wind speeds change throughout the day and days of its existence and could have differing speeds when making landfall and when in the open ocean. The magnitude can also differ between landfalls of different countries.

To acquire more credible data, the country or regional meteorological agency sites for each country were consulted to find the individual reports of each storm in order to determine the missing magnitude, number of deaths, or number of affected. Missing data for Australia was found at the website for the Australian Bureau of Meteorology, Taiwan data was found at the website for the Joint Typhoon Warning Center and for the United States, additional missing data were taken from the National Oceanic and Atmospheric Administration (NOAA). These agencies were used to cross-reference with the EM-DAT database because they are some of the main sources that CRED uses for their natural disaster information [23]. Although it is unclear what each of the sources were rated on CRED's 1 to 5 reliability scale. To distinguish between the new data and the EM-DAT data, additional columns in an excel workbook for 'Magnitude', 'Deaths' and 'Affected were created to indicate where the data was derived. Additionally, the individual storm reports were analyzed to see if the magnitude value varied from different countries' landfalls throughout the track to ensure the most accurate wind speed estimation for the model. To compare already available EM-DAT data and data with the other meteorological agencies' data, information regarding magnitude, deaths and affected were also reported to cross-reference with the EM-DAT data for credibility.

A further component that had to be considered for accurate prediction results was the timeframe of historical data collected to include the changing effects of climate change. According to Richard B. Rood, a current climate and space sciences and engineering professor at the University of Michigan, a precise timeframe to look at historical weather data for predicting future weather is 30 years [24]. He continues to say how in the past 30 years all surface temperatures on Earth were above average, with the last recorded surface temperature being at or below average was in February of 1985 [24]. The United States Geological Survey (USGS) claims that global climate change will likely bring more severe weather patterns and storms, including tropical cyclones [25]. To portray these findings in relation to climate change for this study, tropical cyclones that were recorded between 1987 to 2016 were used to portray this 30-year weather pattern for accurate results.

RESULTS

When analyzing the United States data that was given, it was found approximately 57% of tropical cyclone entries did not have values for the 'Magnitude value' from the CRED EM-DAT database. Australia and Taiwan also had missing data for 'Magnitude value' at 35% and 59%, respectively. Furthermore, the United States data were missing 46% of the 'Total affected' information needed to create mortality rates for the model, while Australia was missing about 26% and Taiwan was missing about 64%. Overall, 55% of magnitude entries were missing from the selected dataset and 49% of the 'Total affected' data were missing. Additionally, roughly 32% of the original dataset were missing entries for both columns. Table 3 shows the missing values distribution from our original dataset. For this study, values that were indicated as 0 for 'Magnitude value' and 'Total affected' were indicated as missing because if a tropical cyclone event occurred and was entered into CRED's database, a magnitude of the storm must have occurred and some people would have been affected. It was difficult to determine if there were any missing entries for regarding 'Total deaths' from the database because there were no empty entries for that column, but many indicated as 0.

This study was conducted to create models to predict mortality rates based on historical tropical cyclone data for a country. Unfortunately, the data analyzed was unable to provide enough information to produce reliable results. After cross-referencing the EM-DAT's database with the national or regional meteorological agencies' databases, it was evident that there were many more discrepancies besides the missing data. Many of the 'Total deaths' from our original dataset differed from the individual storm reports that were found, which was surprising since these reports were main sources for our dataset. Furthermore, it was difficult to assess what the magnitude value should be indicated as for each individual country. Many of the older reports (before 2000) only clearly labeled the what the tropical cyclone's highest sustained winds were, but there were difficulties when looking for the wind speed at specific timepoints, like landfalls. Reports that were done after 2000 had more frequent time stamps of the wind speed, making it easier to determine the approximated wind speed at landfall.

Out of the three databases used to find more credible data, NOAA had the most abundant. Most of the reports had easily identifiable magnitudes for specific landfalls and had the deaths written in the same section. This is most likely because NOAA is a United States government agency and could easily obtain the data for the United States. The Joint Typhoon Warning Center, used to find missing data for Taiwan, has their headquarters in Hawaii, which could explain why Taiwan data was not a main focus at times. Nevertheless, NOAA, the Joint Typhoon Warning System and the Australian Bureau of Meteorology did not yield the same results for either 'Magnitude value' or 'Total deaths' on a consistent basis. Again, for 'Magnitude value' this is most likely because the data found in the original dataset most likely had the highest magnitude the cyclone reached (this is based on speculation as there was no definition for how the magnitude value was found on CRED's website), while the newly found data was specific to when the cyclone made landfall. The values for 'Total deaths' were most likely different due to the CRED EM-DAT database also including missing people in the count. The deaths found in the new databases reported on deaths directly linked to the storm and separately indicated indirect deaths, which were not included in this study.

Conversely, none of the new databases had any indication of 'Total affected' to fill in the missing entries from the original dataset. This was the column that was missing the most entries in the entire original dataset and it is obvious that it is because this data is not as readily available as other data collected. It is difficult to obtain a figure that estimates an entire population that was affected by a storm, and by being unable to find it in official storm reports shows that it is sparser than originally anticipated. The 'Total affected' indicator is one of the most important variables for this study because it determines the mortality rate from the event. Without this value, it is impossible to indicate what the mortality rate would be.

DISCUSSION

The EM-DAT database from CRED is one of three global and multi-peril loss databases that is consistently used by other organizations and agencies around the world for their expertise on disaster information [26]. Unfortunately for the purposes of this study, the database did not provide sufficient data to create a reliable prediction model of mortality for different countries from 1987 to 2016. After deciding to look at Australia, Taiwan and the United States, the EM-DAT data was cross-referenced with their main in-country meteorological agency, the Australian Bureau of Meteorology, the Joint Typhoon Warning Center, and the National Oceanic and Atmospheric Administration, respectively, to see how the numbers compared. Many of the data did not match and some data were still missing to a point where prediction was not possible to be conducted. This tells us that there is inconsistent data across databases, which could stem from differing definitions for each indicator collected, such as deaths and those affected, or due to incomplete data collection after a disaster.

As seen in our results, this study was not able to use the indicators given from the CRED EM-DAT database to create a prediction model that could estimate the mortality rate of tropical cyclones for a specific country. This study faced a number of limitations that hindered the furthering of the model that was the end goal. However, these results are not useless and show that there is still a great deal of research and commitment that must be done in order to predict impact and mortality of disasters.

Limitations

An apparent limitation of this study was the inconsistency and missing data from the EM-DAT database. Many of the tropical cyclone events from the original dataset did not have completed data from the database with one or multiple data columns empty or with unreasonable numbers. A number was deemed unreasonable if it appeared too large or too small compared to historical data or other numbers generated from the storm. A consistent trend with missing or unreasonable numbers came from the 'Total affected' column, which was supposed to account for the number of people that were affected by the storm in some way, whether it be from evacuations, property damage, or those requiring immediate assistance due to the event [27]. Data can also be entered incorrectly or interpreted differently when put into a global database. A common interpretation is units being labeled incorrectly, which would express an incorrect impact [28].

Another factor that creates inconsistencies in data is how the country that is experiencing the disaster reports a disaster event. Every country has their own way of reporting loss and other measures when a disaster occurs, with some having high quality data, such as North America, while some countries do not have the means to accurately make reports after a disaster, such as developing countries [28]. Many countries also tend to have abundant reports when the disaster is large or catastrophic, but limited data when a disaster has a slighter impact, leaving some information missing or inaccurate [28]. This can affect how data is interpreted because it might not take into account any information of the smaller disaster. In prediction analysis, this would bias the results to show that these catastrophic impacts are more likely to cause impact and happen more often because there would be no data on the weaker event.

Furthermore, how the EM-DAT database defined a disaster applies thresholds on what is considered a disaster, which can be subjective to other databases for disaster data. Again, at least one of the following criteria must be met or known in order for any disaster to be entered into the EM-DAT database: (1) 'ten or more people reported killed'; (2) 'a hundred or more people reported affected'; (3) 'declaration of a state of emergency'; or (4) 'call for international assistance' [22]. This could limit the database because it neglects some entries that could be considered in other global disaster databases. NatCatSERVICE, a global natural catastrophes database created by Munich Reinsurance Company, has lower standards when it comes to including a disaster into their database [26, 29]. The NatCatSERVICE database will include an event as a disaster in their database if there is any harm to humans or if there is any property damage reported from the disaster and is then categorized into six levels of severity based on the disasters impact [26]. These differences in criteria vastly change the number of events that are considered disasters in each database. NatCatSERVICE adds an average of 800 natural disaster events each year, while EM-DAT adds an average of 300 both natural and technological disaster events each year [26].

The two different criteria pose a problem when it comes to creating prediction models from the data and producing reliable results. First, the number of tropical cyclone entries can affect how the impact of a tropical cyclone would affect mortality in this study. If there are more minor or moderate events taken into account from one database, the results of mortality rate would be different if the database that had fewer events was used. Since the EM-DAT database had less events reported than the NatCatSERVICE database, there is less evidence of these smaller disasters and the actual rate of mortality for these storms is not being accurately portrayed. A study done by Chiung-Ting Chang studied mortality from tropical cyclones in Taiwan from the years 2000 to 2014 [5]. He found that moderate intensity, only described as an intensity level 2 on a scale of 1-3, tropical cyclones cased the majority of deaths in his study timeframe. He attributed these findings to the fact that people are better prepared for more intense storms because they are the most at risk events and the maximum average wind speed in the center of the tropical cyclone. In conclusion, Chang found that moderate storms were reported four times more than intense intensity, described as an intensity level 3 on a scale of 1-3, tropical cyclones but accounted for 16 times more deaths than intense events [5].

With these results from Chang's study, it shows how important it is to include every tropical cyclone event when performing a prediction analysis and not just the worst events or events with the least missing data. This concept of having different criteria among the databases can question the stability of a prediction model. A prediction model that is known as stable is one that can be utilized in different samples from the population being model, while still maintaining the same accuracy [30]. Palmer and O'Connell suggest to perform cross-validation to test the accuracy of a model by taking two random samples from the same population [30]. However, the prediction model that would have been generated from this study could have faced reliability issues and drawn different conclusions because the data obtained did not have a universal criterion.

Another limitation that made predicting the estimated mortality of future cyclones was the lack of a definition from the dataset for the 'Total deaths' column. CRED defines their 'Total deaths' column of data from their database as 'Sum of death and missing' [22]. In *Guidelines on Measuring Losses from Disasters: Human and Economic Impact Indicators* (2015) by the Integrated Research on Disaster Risk (IRDR) Programme, established by the International Council for Science (ICSU), they discuss how impact should be measured and classified by disaster databases [31]. In their manual, they state that the death indicator in these databases should only include the number of deaths associated with the disaster, not the number of missing [31]. Missing persons should not be included in with deaths for disaster data because there is no physical evidence of their death from the disaster. This could make the disaster data unreliable due to incorrect assumptions.

This definition can also be misleading because it is unclear if these deaths include both direct and indirect deaths. A direct death from a disaster is defined as a fatality that is exhibited to be from the disaster, such as drowning from the storm surge generated by a cyclone [32]. However, an indirect death is a death that was not directly exhibited to be from the disaster, but one that would not have happened if the disaster did not occur [32]. An example of an indirect death would be an individual who died in a car accident as he or she was evacuating from the disaster. Additionally, these indirect deaths could be from exacerbated chronic conditions from the stress or devastation from the storm. Rappaport reports that accounting for only direct deaths can affect a storm impact report because it does not 'tell the whole story' [32]. In Doocy et al.'s systematic review on impact from tropical cyclones, they reported that 43% of the deaths that were caused by tropical storms in the United States from 1985 to 2008 were indirect deaths [2]. However, many agencies that create impact reports for cyclones, including the National Hurricane Center, do not account for indirect deaths and only emphasize direct deaths [32]. Kron et al. states that disaster data should be treated with the notion that there are many difficulties when dealing with loss data from disasters and that there could be some underlying characteristics with the data [28]. An example would be how the CRED database will enter a quantitative report when there are no numbers preset. Their guidelines state that if a specific number is not given but 'hundreds' is written from their source, they will automatically put 200 persons, but expect the number to be higher [23].

Another evident limitation for this study was the lack of a denominator to interpret an accurate mortality rate among the affected population of the tropical cyclones used in this study. The denominator that was going to be used was their indicator for an affected population from each storm. However, after analyzing the data, many of the events lacked a number indicating the number of people affected or had a number that was unreliable, such as 0. CRED states that their indicator for 'affected' is used or produced by many different agencies in and that there is a great deal of haziness in the definitions. This creates a vast number of criteria used to estimate those affected, which then makes the numbers incomparable and the numbers more unreliable [22]. Although the CRED database defines those affected as 'people requiring immediate assistance during an emergency situation', it is extremely difficult to come up with an accurate estimation quantitatively. Additionally, CRED did not always use numbers that were generated for number of affected, and instead would look at the number of households that were destroyed from a storm. Their method was to then multiply the number of households by 5 if in a developing county and multiply by 3 if in a developed country to give an estimation of those affected. Also, if there was not a specific number used and 'thousands' was written to indicate the number affected instead, the number 2,000 would be used to indicate the number of affected, which they said was probably underestimated [23]. Although the CRED guidelines state all of this ambiguity with the 'affected' indicator in their database, it makes it extremely difficult to utilize it to create an accurate prediction model for the mortality of a population.

Furthermore, it is important to consider how predictive a prediction model could be, in general, and how predictive it can be for a natural disaster. According to Galit Shmueli, prediction models are used to build upon theories. However, many scientists tend to ignore prediction models [33]. Many believe that prediction is considered 'unscientific' and it would be more beneficial for scientists to just compare current events to past events and see how they have changed over time [33]. This thought process questions the validity of prediction and whether this model would have been useful for predicting mortality for disasters. Although it was mentioned earlier that climate change is causing storms to be more severe, most natural disasters come in an unpredictable circumstance, and some without any advanced warning. Additionally, disasters are more of a random event and many types of disasters have little events to predict with, including tropical cyclones in many countries.

Future Recommendations

For data to be more reliable for prediction analysis, all databases should have consistent criteria that they follow when entering disaster data and pull from similar data sources to ensure the same, accurate information. Having databases that contain different numbers of entries due to conflicting criteria to be considered a disaster can have an impact on the analysis that is done. The same is true for data that is not directly known. It is also important to only include data that has entries of data with known quantitative figures in order to provide the most truthful data. Entries that contain numbers that are unrealistic for the entry or numbers that were estimated by a word indicator such as 'hundreds' or 'thousands' do not provide information that portrays the impact of the disaster.

More variables are needed for a better prediction model as well. A time variable should probably be introduced into the model to indicate when the tropical cyclone formed because storms are more severe when the water is warmer. This could mean that in the off change a cyclone forms when the water is colder, it may not be as severe as one in the middle of summer when the water should be at its hottest. Another variable that should be considered is one for the Fujiwhara effect. Although rare, the Fujiwhara effect is when two cyclones form very close to each, sometimes to the point where the smaller one is absorbed by the larger one to form one intense storm [34]. This effect could have an effect on the intensity of a storm, which could cause a higher mortality rate. A few supplementary variables that look at land coverage on a country may also be beneficial to indicate if they would be prone to secondary disasters like landslides.

REFERENCES

1. Twigg J. *Disaster Risk Reduction*. (2015 edition). London, United Kingdom: Overseas Development Institute.
2. Doocy S., Dick, A, Daniels, A, and Kirsch TD. The human impact of tropical cyclones: A historical review of events 1980-2009 and systematic literature review. PLOS: Current Disasters 2015; 1.
3. Keim ME. Assessing disaster-related health risk: Appraisal for prevention 2017.
4. Jonkman SN and Kelman I. An analysis of the causes and circumstances of flood disaster deaths. Disasters 2005; 29(1): 75-97.
5. Chang C-T. A preliminary study on mortality attributed to tropical cyclones in Taiwan 2000-2014. Environmental Hazards 2016; 15(2): 178-187.
6. United Nations International Strategy for Disaster Research (UNISDR). *2009 UNISDR Terminology on Disaster Risk Reduction*. Geneva, Switzerland: United Nations International Strategy for Disaster Research (UNISDR) 2009.
7. Keim ME. Introduction to Disasters 2017; [PowerPoint slides].
8. Keim ME. Health Impacts of Environmental Disasters. In G. R. Ciottone, P. D. Biddinger, R. G. Darling, S. Fares, M. E. Keim, M. S. Molloy and S. Suner (2nd Edition), *Disaster Medicine* 2016. Philadelphia, PA: Elsevier, Inc.
9. Noji EK. Disasters: Introduction and state of the art. Epidemiologic Reviews 2005; 19:3-8.
10. Zibulewsky J. Defining disaster: The emergency department perspective. BUMC Proceedings 2001; 14:144-149.
11. Keim ME. Tropical Cyclones I 2017; [PowerPoint slides].

12. Keim ME. Cyclones, tsunamis and human health: The key role of preparedness. Oceanography 2006; 19(2):40-49.

13. Shultz J, Russell J, and Espinel Z. Epidemiology of tropical cyclones: The dynamics of disaster, disease and development. Epidemiologic Review 2005; 27:21-35.

14. Tang B and Emanuel K. A ventilation index for tropical cyclones. American Meteorological Society 2012; 1901-1912.

15. National Hurricane Center. Saffir-Simpson Hurricane Wind Scale. *National Oceanic and Atmospheric Administration (NOAA).* Available from www.nhc.noaa.gov. Last accessed April 14, 2018.

16. Malilay J. Tropical cyclones. In E. K. Noji (ed.), *The Public Health Consequences of Disasters* 1997. New York, NY: Oxford University Press.

17. Udono T and Sah AK. Hazard mapping and vulnerability assessment. Regional Workshop on Total Disaster Risk Management 2002.

18. Eslinger D. Data on tropical cyclone tracks, National Oceanic and Atmospheric Administration (NOAA) 2018.

19. Keim ME. Tropical Cyclones II 2017 [PowerPoint slides].

20. United Nations International Strategy for Disaster Research (UNISDRb). Disaster Statistics. Available from www.unisdr.org. Last accessed April 14, 2018.

21. Center for Research on the Epidemiology of Disasters. Frequently asked questions. *EM-DAT: The International Disaster Database*. 2017; Available from: www.emdat.be/. Last accessed March 3, 2018.

22. Center for Research on the Epidemiology of Disasters. Explanatory notes. *EM-DAT: The International Disaster Database*. 2017; Available from: www.emdat.be/. Last accessed March 3, 2018.

23. Center for Research on the Epidemiology of Disasters. Guidelines. *EM-DAT: The International Disaster Database*. 2017; Available from: www.emdat.be/. Last accessed April 4, 2018.

24. Rood RB. Earth has now had 30 straight years of record monthly temperatures. The New Republic 2015.

25. United States Geological Survey (USGS). How can climate change affect natural disasters? Available from www.usgs.gov. Last accessed April 15, 2018.

26. Wirtz A, Kron W, Löw P, and Steuer M. (2012). The need for data: Natural disasters and the challenges of database management. Natural Hazards 2012; 70(1): 135-157.

27. Guhar-Sapir D, Hoyois P, and Below R. Annual disaster statistical review 2015: The numbers and trends. Brussels: Centre for Research on the Epidemiology of Disasters 2016.

28. Kron W, Steuer M, Löw P, and Wirtz A. How to deal properly with a natural catastrophe database—Analysis of flood losses. Natural Hazards and Earth Systems Sciences 2012; 12:535-550.

29. Munich Re. Topics Geo 2016. Able, M. (2017 issue). München, Germany: Kastner & Callway Medien GmbH 2017.

30. Palmer PB and O'Connell DG. Research corner—Regression analysis for prediction: Understanding the process. Cardiopulmonary Physical Therapy Journal 2009; 20(3):23-26.

31. Integrated Research on Disaster Risk (IRDR). Guidelines on measuring losses from disasters: Human and economic impact indicators. (IRDR DATA Publication No. 2). Beijing: Integrated Research on Disaster Risk 2015.

32. Rappaport EN and Blanchard BW. (2016). Fatalities in the United States indirectly associated with Atlantic tropical cyclones. Bulletin of the American Meteorological Society 2016; 97:1139-1148.

33. Shmueli G. To explain or to predict? Statistical Science 2010; 25(3):89-310.

34. National Hurricane Center. Fujiwhara effect. *National Oceanic and Atmospheric Administration (NOAA)*. Available from www.nhc.noaa.gov. Last accessed April 15, 2018.

TABLES

Table 1. Saffir-Simpson scale on tropical cyclone intensity.

Category	Wind Speed (mph) (knots)	Storm Surge* (feet)	Damage	Effects
Tropical Depression	0-38 0-33	0-1	Minimal	No significant damage to buildings or structures
Tropical Storm	39-73 34-63	2-3	Minimal	No significant damage to buildings or structures
1	74-95 64-82	4-5	Minimal	No significant damage to buildings or structures
2	96-100 83-95	6-8	Moderate	Wind speeds can cause damage to windows and roofs
3	111-130 96-112	9-12	Extensive	Serious wind damage and severe coastal flooding
4	131-155 113-136	13-18	Extensive	Damage to well-built structures by wind and major flooding
5	155+ 137+	18+	Catastrophic	Structure failures and severe flooding

***Storm surge use water that is pushed onto land during a cyclone due to water piling up, winds, waves, and sea-level rise.**

Table 2. Top 10 countries with the most tropical cyclone activity from 1900 to 2016, according to the EM-DAT database.

Country	**Number of Occurrences**
Philippines	310
China	150
Japan	133
United States	112
India	105
Mexico	89
Bangladesh	88
Vietnam	84
Taiwan	78
Madagascar	53

***11th Hong Kong (53); 12th Australia (39)**

Table 3. Missing data from our dataset with variables needed to create the model.

Country	Total Entries	Magnitude Value		Total Affected	
		n	%	*n*	%
Australia	23	8	34.8	6	26.1
Taiwan	59	35	59.3	38	64.4
United States	61	35	57.4	28	45.9

FIGURES

Figure 1. The global distribution of tropical cyclone paths from 1987-2016. The highest proportion of tropical cyclones are seen in the western Pacific around the Philippines and Taiwan and the eastern Pacific off the west coast of Mexico. There is also a presence of high tropical cyclone activity on the western Atlantic and the Gulf of Mexico, as well as around Australia and the Indian Ocean.

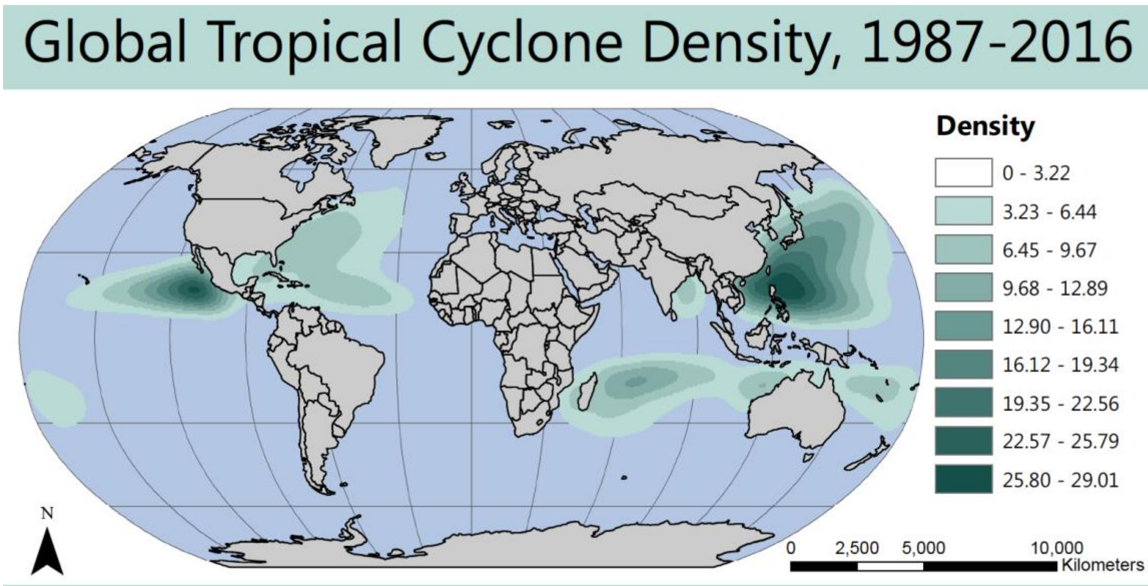
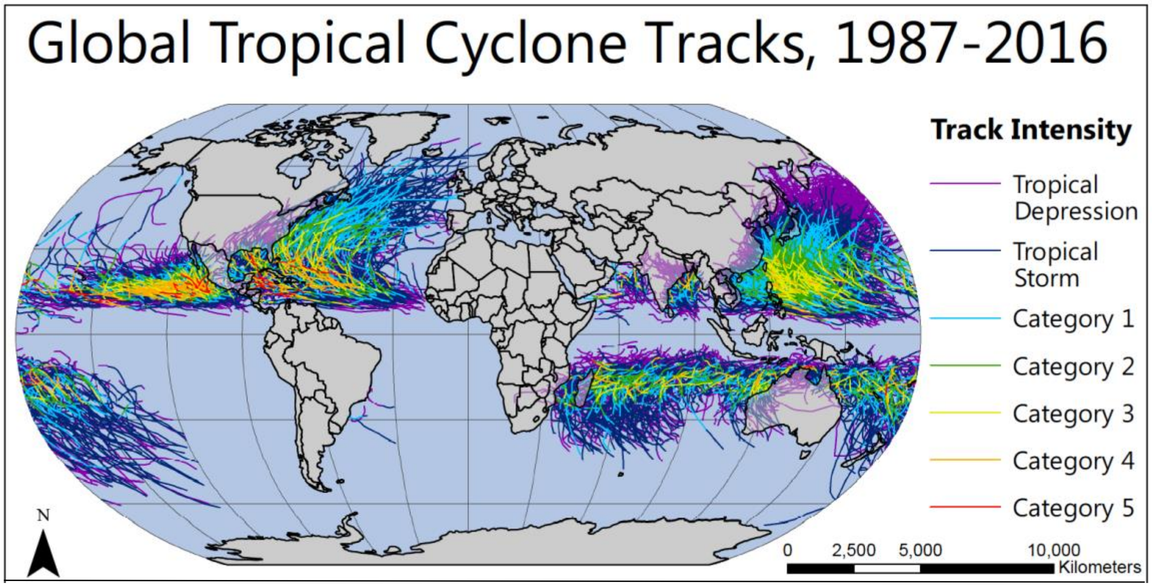


Figure 2. The global tracks of tropical cyclone paths from 1987-2016. The most intense tropical cyclones are seen in the western Atlantic around the Gulf of Mexico and the eastern Pacific off the west coast of Mexico. There is also a presence of high tropical cyclone activity on the western Atlantic and western Pacific off the east coast of Asia, as well as around Australia and the Indian Ocean.



APPENDIX 1

DEFINITION OF LABELS[2]

Start date: The date when the disaster occurred

End date: The date when the disaster ended

Country name: The country in which the disaster has occurred or had impact

ISO: The International Organization for Standardization attributes a 3-letter code to each country. CRED uses the ISA 3166. This field is automatically linked to the country

Location: Geographical location within a specific country

Latitude: North-South coordinates

Longitude: East-West coordinates

Magnitude value: The 'intensity' of a specific disaster

Magnitude scale: kph (kilometers per hour; speed of wind)

Disaster type: Main disaster identified with the event

Disaster subtype: Subdivision related to the disaster type

Disaster name: Any specification related to the disaster which allow its identification

Total deaths: Deaths and people determined as missing

Total affected: This number includes the sum of the injured, affected and homeless

Disaster No.: An 8-digit disaster number that is given to each disaster event

[2] All labels and label definitions came from CRED's database Guidelines on emdat.be [22]