Challenges of Integrating Physical Exposure and Human Impacts Data in Tropical Cyclone Studies

Matthew Hughes and Brooke Anderson

May 25, 2020

Introduction

Tropical cyclones—which encompasses hurricanes as well as tropical storms and tropical depressions— regularly threaten coastal communities across the Eastern and Southern United States. From 2000 to 2019, tropical cyclones cost the United States at least 811 billion dollars in damages(**???** billion dollar disasters). Tropical cyclones in that same time frame resulted in 6,010 human fatalities, averaging 301 deaths per year (**???** billion dollar disasters). Tropical cyclones upset coastal communities and society by damaging property, disrupting local economies, and harming human health. This is why they are so critical to study.

Researchers have observed that in utero exposure to tropical cyclones leads to adverse birth outcomes. (Kinney et al. 2008) observed higher rates of autism in children born to mothers who had higher rates of storm exposure than children born to mothers who were exposed to later intensities. This is not very surprising when one considers that tropical storms are highly stressful events, and stress during a pregnancy is known to have strong impacts on the developing fetus. The scientific literature also reveals evidence of mental health outcomes associated with populations exposed to tropical cyclones. Survivors of tropical storms often report higher levels of depression, anxiety, and PTSD, due to reduced access to important medical and social services, property damages, poor sanitation, and displacement after storms. (Lieberman-Cribbin et al. 2017) found higher levels of PTSD in New York City residents who were exposed to flooding after Hurricane Sandy. Beyond health impacts, both mental and physical, tropical storms create incredible strains on the economies of the Southeastern United States. The average cost of a tropical cyclone event in the US is 21.2 billion per event, CPI-adjusted (**???** billion dollar disasters).

Clearly, tropical cyclones dramatically impact the social, economic, and physical wellbeing of coastal communities. These extreme weather events represent an environmental health threat that is not going to disappear, and given that coastal regions of the Southeastern US are experiencing population growth, it is likely that higher numbers of people will be put at risk in the future. Avoiding these risks is not possible, but building resilience in communities after they experience tropical cyclone events is key to mitigating damages and preparing for future disasters. Creating lasting and resilient communities in areas prone to tropical cyclones requires that researchers understand which populations and locations are at the greatest risk for negative exposures to tropical storms. This requires data that allows researchers to assess where in space and time tropical storms occur, and also where in space and time individuals and populations are experiencing impacts from these storms.

Multidisciplinary teams of researchers are exploring this using different datasets, however a key challenge is integrating data from across disciplines. For example: extensive physical exposure data is often available for tropical cyclones as they near and cross communities in the United States. This data can come both from established monitoring networks, like [NOAA network name?], but may also result from data collection efforts during or after the storm by atmospheric scientists and engineers seeking to characterize a storm. Researchers studying the human impacts of these storms, including epidemiologists, economists, and social scientists are interested in this data as well, but the differences in temporal and spatial resolution makes the data accessible harder to use. Resolving physical exposure and human impact datasets is challenging because the human impact data and physical exposure data often do not have congruent resolutions.

Here we explore cases and implications of integrating data at different temporal and spatial scales, focusing as an example on human impact studies of tropical cyclones in the US. We begin by investigating the reasons that spatial and temporal misaligment exist in the study of tropical cyclones. We then describe the main spatial and temporal scales used, and finally assess some of the consequences that result from integrating physical exposure data with human impacts data.

Spatial and Temporal Misalignment: Origins of Integration Challenges

Questions about the human impacts of tropical cyclones are multidisciplinary, and as such require datsets from different and sometimes seemingly disparate sources. Different disciplines have different methods of collecting data. These differences go beyond the types of software or data management systems used, they often come down to differences the spatial and temporal scales that data points are collected at. These differences and temporal and spatial scales are what we refer to in this paper as spatial and temporal misalignment. In this section we'll describe what drives these differences in temporal and spatial scale, and also inform the reader about the most common spatial and temporal scales used in tropical cyclone research, detailing some of the situations where particular spatial and temporal scales are advantageous.

Many of the drivers for spatial and temporal misalignment in any sort of multidisciplinary research stem from the fact that these disciplines evolved and created their own research methods separately. These separate methods are easily illustrated in the various ways in which data are collected across disciplines. For example, in physical exposure assessments, data is often recorded by sensors in vast monitoring systems that are designed to automatically record a data point at a fixed interval of time. These monitoring systems are often the result of long-standing weather projects such as NOAA and the NWS, and the data is often narrow in temporal and spatial resolution. On the other hand, researchers studying human impacts often collect their data from administrative sources. These sources are often aggregated by geographic region and time, often out of convenience, or a need to preserve the anonymity and privacy of the people whose data is being used. This goes to show that different disciplines have different aims and needs, and this creates spatial and temporal misalignment when multidisciplinary research is conducted.

Differences in spatial and temporal scales are also related to the study question that researchers are asking. If a study is concerned with birth outcomes for example, having weather data on the windspeed every several seconds may not be relevant, because birth outcomes related to storm exposure in utero may operate on a longer time scale. In (S. C. Grabich et al. 2016), the researchers looked at gestational periods and defined pregnancies as exposed to tropical cyclones if they happened before 20 weeks of gestation. If the researchers had been interested in a different question, for example acute injuries due to direct storm exposure, they would have chosen a smaller time scale. There is no correct spatial or temporal scale that works well for all research, it all depends on what is being asked and how that can be ascertained.

Another factor that drives spatial and temporal misalignment in tropical cyclone studies is that researchers studying human impacts often do not collect the physical exposure data themselves. Instead they get this data from large sources of secondary data such as the National Hurricane Center Data Archive from NOAA (National Oceanic and Atmospheric Administration) and the NWS (National Weather Service). Researchers also will use such secondary datasets and sources to compare with primary data sources. For example in (Lieberman-Cribbin et al. 2017), self reported flooding exposure data was compared to FEMA flooding exposure data.

If spatial and temporal misalignment are a result of different disciplines using different methods, asking different questions, and collecting data from different sources, then it is important to understand what those temporal and spatial scales are. There are reasons pertaining to research questions that exposure data may be aggregated at the level of zip code or county as opposed to a specific point location. Likewise, there are reasons that exposures may be analyzed over days, weeks, months or even years. Different scales allow the researchers to make certain inferences and determine how the results of a study can be interpreted.

The remainder of this section will highlight the most common spatial and temporal scales typically used in

tropical cyclone studies. These scales were chosen after conducting a literature review that covered a wide range of human impacts from tropical cyclones. First we will describe spatial scales starting from the smallest resolution of point locations, working up to the level of metropolitan areas and states. Next we will describe temporal scales most commonly used in tropical cyclone studies and again work from smallest to largest resolution.

Spatial Scales

The spatial scale that a researcher uses varies depending on the data available or sampling method used. In human impacts data finer spatial scales will correspond more often to individuals or households, while larger spatial scales will correspond to regions, states, or even countries. Physical exposure data is often at a small point location or a grid, based on where weather monitoring sensors are placed. In the following section we will outline the most common spatial scales used in tropical cyclone studies and include some examples from the literature where they were employed.

Point Location

[BA: Let's think some about the order we want for these sections. We're making several good points / analysis here. First, we're defining what we mean by the resolution ("point location" here). We probably want to start with that. Then we have some examples for studies that have had outcome data at this resolution. Maybe that could go next, to help illustrate the definition we've given. We've got some information on *how* the data at this scale was collected (e.g., geocoding from addresses reported from the study subjects), which I think is really interesting. Finally, we're got some text that talks about how data at this resolution could be integrated with some main formats of exposure data. We might want to end with that (or maybe even, as we work on this draft, that might go into a different section of the paper).]

Point locations are the smallest resolution of spatial data used to assess the exposure to tropical storms and hurricanes, as they represent the specific location of individual, non-aggregated observations on the outcome of interest. In many cases, researchers collect information on the study subject's residential address through some sort of a survey to assess point location (Lieberman-Cribbin et al. 2017), (Jaycox et al. 2010),(Bayleyegn et al. 2006). These surveys are often designed to assess psychological needs of hurricane survivors, as well as medical, financial, and nutritional needs. For example in (Lieberman-Cribbin et al. 2017), New York City residents provided their address in a self reported manner to look at associations between mental health outcomes and flooding data. This residential address served as a point location that could be mapped and was compared to flooding data maps created by FEMA. In other cases, a GPS device is used to record coordinates that mark a specific point location. An example is (Hagy, Lehrter, and Murrell 2006), where specific point locations were used to take water samples were taken to measure parameters of water quality such as salinity, temperature, dissolved oxygen, and turbidity compared before and after Hurricane Ivan in Pensacola Bay, Florida. This is a common practice in ecological research because point locations distributed across a landscape can be used to observe patterns taking geography into account. Point locations are also advantageous when using satellite images in conjuction with analysis of hurricane impact as illustrated in (Bianchette et al. 2009), where Landsat 5 images were used to compare vegetation damage, by looking at specific trees at different elevations to assess the ecological impact of Hurricane Ivan.

The obvious advantage of a point location is that when mapped, it can be overlayed with physical exposure data on a storm or storms to gage a very accurate picture of exposure, taking full advantage of high resolution in the exposure data. Since storm tracks are often spatially represented by the path of the storm's center, having point locations for the exposed units of interest allows researchers to more accurately measure how close each observation was to the storm's central track, and make further conclusions on this. Similarly, point locations can be integrated in a straightforward way with gridded exposure data, as might result from re-analysis datasets or ... [check with James Done about this], as each point location can be assigned the exposure level of the closest gridded measurement.

[Once we give examples, we should talk about what level the physical exposure data was recorded as. Did it line up exactly? Grided data. Some studies avoid the problem by creating a proxy (ex: dist from the storm

track).]

Zip Code/County/Parish

While point locations are very useful, many of the papers cited used larger geographic areas to denote spatial exposure to storms. Zip codes (Bevilacqua et al. 2020),(Lane et al. 2013), are often used to aggregate groups of people living in a given area. Counties are at a higher aggregation level than zip codes (Kinney et al. 2008), (S. C. Grabich et al. 2016), (S. Grabich et al. 2016), (Schwartz et al. 2018), (Harville et al. 2010). Often these levels seem to be used when a specific metropolitan area is being looked at, such as New York City after Hurricane Sandy (Lane et al. 2013), and Houstan after Hurricane Harvey (Schwartz et al. 2018).

Aggregating exposure at the county level is convenient because it utilizes some of the most established methods for assigning exposure status: the storm track trajectory, and FEMA presidential disaster declarations (S. Grabich et al. 2016). The storm track trajectory is typically the path that the tropical cyclone takes, and although the counties immediately crossed can be categorized as exposed, there are methods to calculate distance from the storm center that allow for estimation of exposure at various distances by establishing exposure thresholds.

There are several disadvantages and pitfalls to using this spatial level. For one, not all counties and zip codes(which are called parishes in Louisiana) are the same size or have the same population, so they may not be immediately comparable. Using the county/parish or zip code makes it easier for researchers to misclassify exposure. There are many ways that this can occur in a study on tropical storms; one common example is that counties selected as exposed are those that had the center of the storm pass through their county's physical boundaries. However it is very possible that some individuals lived in a county classified as exposed based on this criteria, but were in a region of the county far enough away from the storm center that they were not severely impacted. These individuals would be classified as exposed when they really were not and it could bias an apparetn association towards the null. Alternatively, individuals who lived in a unexposed county, but were near the border of an exposed county could be incorrectly categorized as being unexposed even if they actually experienced many of the effects of the storm.

State/Metropolitan Region

Many studies used the spatial level of entire states or specific metropolitan areas to gather information on those who were exposed. (Harville et al. 2010) is an interesting paper because it looks at the state level as well as the regional and parish level. In this paper researchers observed birth outcomes in response to Hurricane Katrina in the state of Louisiana as a whole, the New Orleans metropolitan area, and Orleans parish, which is the heart of New Orleans. Looking at these three levels is a way to compare different incident rates and other measures of associations across different spatial scales.

The state or national level is the spatial level of an ecological study and can be useful to compare the emergency preparedness and policies of different states. The potential for the ecological bias is of course present when looking at this spatial scale however, which occurs when the outcomes on the population level (typically an average), do not represent the individual outcomes very well.

Temporal Scales

Thanks to scientitific institutions such as NOAA and the National Weather Service, there are wide networks of sensors and monitoring equipment established across the United States that are capable of recording physical exposure data at a fine enough level as to render it almost continuous. It is possible to know the wind speed, amount of rainfall, and air temperature at very fine temporal scales throughout the duration of a tropical cyclone event. Human impacts data however, is typically not available at such a fine scale, nor is such a scale sometimes even relevant. Whereas physical exposure data may be collected in real time during the storm, many of the human impacts that researchers are interested in may be only known after the storm, and thus estimations may have to be made of what happened in the past. Below are some examples of time scales that are more applicable to a human scale, and why aggregating physical exposure data to these units of time may be necessary.

Day

In the event of a tropical cyclone, there are several situations in which the temporal unit of a day may be used to analyze exposure. Physical exposure data from tropical cyclones will typically be available at this scale anyways, but some studies will look at time series and use daily exposure data from hospitalizations and visits to the emergency room.

One study, [zahrah2013daily] looked at casualty counts per day for counties in the Southeaster United States that were exposed to tropical cyclones.

Week

Week is a very common unit of time used to ascertain exposure, particularly for studies that are concerned with birth outcomes and gestation during hurricane exposure (Kinney et al. 2008), (S. C. Grabich et al. 2016), (S. Grabich et al. 2016). When the week of gestation is known, the timing that the hurricane makes landfall, or has its storm center pass through a county can be matched up to this week of gestation to identify possible "critical periods" of exposure during development.

In the North Atlantic Basin, tropical cyclones typically occur between June and November, and so sometimes month to several month periods are used to aggregate the temporal exposure to them.

Cumulative Measures of Time

It is often the case that specific physical exposures are not considered in real time to ascertain human impacts. Instead, human impacts are assessed after the storm has passed, often noting the number of days or weeks that have passed since the hurricane made landfall. This method is common when assessing damages, recovery efforts, and when human impacts are self reported. When this is the case, it is often useful to take an aggregate exposure an aggregate measure of physical exposure corresponding to this time frame, such as the maximum wind speed or maximum flooding level over the period of time being studied.

Implications of not improving this integration

When researchers study the economic, social, and health impacts that tropical storms and hurricanes have in locales like the Gulf Coast of the United States, it is important to select appropriate spatial and temporal scales to adequately classify exposure. Misalignment in the spatial and temporal scale of exposure data versus outcome data creates challenges when measuring and inferring associations between tropical cyclone exposures and human impacts. This is a problem because it gives an inaccurate picture of how communities and individuals' health are impacted by these storms. Two potential implications are that this mismatch can introduce bias in estimated associations and that it can reduce precision in estimates of those associations. In tropical cyclone studies these biases and reduced precision are usually the result of misclassification or measurement error, categorizing continuous data, and ecological bias.

When researchers are confronted with datasets that are at different temporal or spatial scales, there are typically two options for integrating these datasets. The first option is to aggregate the dataset that is at a finer resolution in order to match it to the other dataset. The other option occurs on spatial scales where you have point levels for all of your datsets but they are spatially misaligned. In this case, you have to utilize methods of matching those point locations, for example residential addresses with the closest weather monitor that measured windspeed or flooding. We will go through each of these methods and point out the ways in which error, bias, and reduced precision can occur.

When Data Have Different Scales

When researchers have physical exposure data at a very fine resolution, perhaps even continuous, and human impacts data at a more aggregate level, it is common and practical to aggregate the physical exposure data. This same practice could also work the other way around, aggregating human impacts data to a physical exposure dataset with a narrower spatial and temporal resolution. In any case, when dealing with aggregated

data of any kind, it is important to realize that information on the individual level is lost. Researchers should be mindful when aggregating data, particularly continuous data, that they are losing information. In this section we will discuss some of the major implications of aggregating data which include ecological bias, misclassification and measurement error, and the process of categorizing continuous data .

Ecological Bias

Physical exposures to tropical cyclones, as well as human impacts, are observed across spatial gradiants, and though specific point locations may exist for a weather monitor recording maximum wind speed or rainfall, that point location data is often aggregated to a larger spatial unit. For example, several weather monitors at specific point locations recording maximum wind speed may be replaced by the highest wind speed in a specific county. Data such as this is known as ecological data, aggregate data, or contextual-level data. Studies that use such data are known as ecological studies (Sedgwick 2014). Ecological bias occurs whenever the aggregate association between an exposure and an outcome does not properly reflect the association on the individual level (Greenland and Morgenstern 1989). There are times when this is not a concern, such as when the aggregate is a count. For example in (Zahran, Tavani, and Weiler 2013), the daily casualty count was reported for individual counties in the Southern United States, using count data from the Spatial Hazard Events and Losses Database. However, when estimates derived from ecological studies are used to infer individual estimates, ecological bias will likely be present. Especially when there is heterogeneity present in an aggregated population, an ecological estimate should not be taken to be representative of individual estimates.

Categorizing Continuous Data

In addition to using larger spatial designations, researchers aggregate the physical exposures themselves, simplifying continuous measurements down to a single exposure metric. While aggregate values often represent the mean of all the values recorded, weather data is typically assessed by the maximum value. Regardless, aggregating physical exposure data requires researchers to categorize continuous data, which involves choosing appropriate thresholds.

Thresholds are often used to assign exposure status to individuals or populations (often using a county as proxy for a population). For example, a county may be classified as exposed or unexposed based on local winds exceeding a threshold (e.g. gale-force winds or higher). S.C. Grabich et al. 2016 classified hurricane exposure in a Florida county using maximum wind speed. Maximum wind speed is a continuous variable, but the study used binary categorizations to divide it into tropical wind speeds, classified as greater than 39 miles per hour, and hurricane wind speeds, classified as greater than 74 miles per hour. Florida counties experiencing maximum wind speeds below 39 miles per hour were considered unexposed. In this example it is noteworthy to examine that all the Florida counties in this paper likely experienced hurricane winds somewhere on this spectrum, but categorizing that continuous data made exposure much simpler and concrete.

The Saffir-Simpson scale is an example of how entire storms are often classified by their maximum wind speed. Forecasters classify hurricanes into categories on the Saffir-Simpson scale based on maximum sustained surface wind speed. This is defined as the peak one minute wind speed at a height of 10 feet over an unobstructed exposure (Taylor et al. 2010). The Saffir-Simpson scale uses five different bins to classify varying levels of wind speed and determine the severity of a storm. The first level, Category 1 is designated for hurricanes and tropical storms with maximum wind speeds of between 64 - 82 knots and is generally considered dangerous to people, livestock, and pets from the hazard of flying and falling debris (Taylor et al. 2010). On the higher end of the scale, Category 5 designates hurricanes with maximum wind speeds above 137 knots and is considered to have catastrophic effect on damage and a high probability of injury or death to people, livestock, and pets even if they are sheltering indoors (Taylor et al. 2010). An important limitation of the Saffir-Simpson scale is that it doesn't account for other hurricane-related impact variables such as storm surges, flooding, and tornadoes (Taylor et al. 2010).

Another scale used to categorize wind speed is the Beaufort scale, created by Admiral Sir Francis Beaufort, used to classify wind speeds both over land and sea. While the Saffir-Simpson scale is only designated for wind speeds that are already at hurricane levels (greater than 64 knots), the Beaufort scale considers the

wind speeds below this. The scale ranges from Force 0 (0-1 knots and calm) to Force 12 (64 to 71 knots and hurricane). Other interesting parts of the scale include Force 3 (4-6 knots) which is a gentle breeze, and Force 8 (34-40 knots) which is considered a gale.

Despite several advantages to dichotomizing continuous variables that we just discussed, the general consensus in epidemiology is not to do it. Statistical power is lost because so much information is lost when categorization occurs (Van Walraven and Hart 2008). This makes sense when you consider that continuous variables allow you to observe nuance in the data and perceive a dose response relationship between the predictor and response variables, should one exist. This effect is masked when researchers categorize data, and even more so when a smaller number of categorical variables are used (for example dichotomization itself at 2). Generally, if you are going to categorize continuous data, it is better to use 3 or more categories rather than just two, because this will capture more of what the data that would otherwise be lost. An example of a paper that used three different bins was (Kinney et al. 2008), which explored the risk of autism after a pregnancy that included exposure to a tropical storm in the state of Louisiana. The study authors classified tropical storm exposure as severe, intermediate, and low exposure, and these exposure classifications were determined based on whether a mother lived in a Louisiana parish that had both of the exposure factors of interest: storm intensity and storm vulnerability. Storm vulnerability in this case was based on another dichotomy: whether or not the storm center passed through the parish of interest. Storm vulnerability was a measure of how vulnerable the inhabitants of the parish were to the effects of a storm (higher socioeconomic neighborhoods and parishes have more resources to withstand and recover from a tropical storm for example).

Another obvious problem with categorizing continuous data is that the cutoff points are often arbitrary. In the case of dichotomization, the median is often used, but there is typically no reason to assume that the median is a reasonable cutoff point. Because different samples will have different medians, this automatically makes many categorical bins difficult to compare across studies (Altman and Royston 2006). Further, choosing optimal cutoff points that give the smallest p-values can lead to spurious results (Altman and Royston 2006).

Not surprisingly, dichotomizing continuous variables can bias results. A study by Selvin showed that the odds ratios can be significantly different depending on the chosen cutoff that is implemented in a study (Van Walraven and Hart 2008). Categorical variables can also put otherwise similar observations into separate bins if they are close but on opposite sides of the cutoff (Altman and Royston 2006). Choosing a median as a cutoff is intended to delineate bins, but if the bins are a "high" and "low" group, two individual observations that may only be a fraction different, but on either sides of the mean, will be classified as high and low respectively, and give the false impression that they are significantly different.

Although using a single exposure value can simplify analysis and interpretation, particularly over an extended temporal scale, there are some obvious drawbacks to relying on one single aggregate value. For example, the Saffir Simpson categories typically correspond only to the geographic point location where the maximum wind speed was observed (Taylor et al. 2010). Hurricane Wilma in 2005 for example, was a Category 3 hurricane when it made landfall on the southwest coast of Florida, but it created Category 1 and Category 2 conditions for the more populous Miami-Dade, Broward, and Palm Beach counties when it finally reached them (Taylor et al. 2010).

Misclassification and Measurement Error in Aggregating Data

When aggregating data, another concern that arises is misclassification or measurerement error. Misclassification error occurs when exposure and outcome variables are measured in categories and the wrong category is assigned to a particular case/observation - for example when a case that is exposed is incorrectly categorized as unexposed. Failure to classify exposure accurately(for example, classifying certain observations as exposed to a storm when they really were not, or vice-versa), allows misclassification bias to move the results of the study further from the true parameter . Measurement error occurs when the variables being measured are continuous, such as the amount of precipitation or the wind speed that was measured during a tropical cyclone.

Environmental epidemiology studies are often prone to misclassification error because the methods of assessing exposure are not always congruent with the way that researchers conduct human impact studies. It is easy to map the path of a tropical cyclone's center, and categorize every county it passes through as an exposed

county. However, this information by itself would not give the researcher any information about population centers that the storm passed through or near to. A town within an exposed county may or may not have been close to the storm's path. Conversely,a town in an unexposed county could be located very close to the border of an exposed county, and even be closer to the storm's track than a different town within that exposed county. An example of a study that could be prone to this kind of bias is (Kinney et al. 2008) where Louisiana parishes were considered vulnerable to hurricane exposure based on whether or not the storm center passed through that parish. It is possible that the cases considered exposed based on living in these parishes were not in fact exposed since the storm may have passed through only a certain part of the parish. Never the less, all cases in a parish are considered exposed or unexposed in the aggregate.

A potential solution to this problem of misclassification of populations is through the use of dasymmetric mapping. This is a method that creates heat maps, using different colors to illustrate differences in population density among other things. Overlaying storm tracks on dasymmetric maps is a method that could be employed to record differences in

When Data Have the Same Scale but are at Different Locations

Sometimes, researchers may have access to data that is down to the point source, both for physical exposures and also for human impacts. Very likely however, these point sources will not be the same. Here the issue is not of integrating different resolution levels, but rather of matching different point locations. Weather monitoring sensors may be set up regularly in a grid formation in a geographic region, but the human impacts point locations could be tied to a residential address. In this case, the residential addresses would have to be united with the closest weather monitoring sensor.

Misclassification for Same Scale Different Locations

Misclassification is again another potential source of bias in this situation. A current weather monitor or sensor may give a certain reading for a maximum wind speed, but it will be the closest weather monitor for multiple different residential addresses that all experience different maximum wind speeds.

Misclassification error / measurement error. One pathway for problems is through misclassification / measurement error bias. Misclassification error occurs when exposure and outcome variables are measured in categories and the wrong category is assigned to a particular case/observation - for example when a case that is exposed is incorrectly categorized as unexposed. Failure to classify exposure accurately(for example, classifying certain observations as exposed to a storm when they really were not, or vice-versa), allows misclassification bias to move the results of the study further from the true parameter . Measurement error occurs when the variables being measured are continuous, such as the amount of precipitation or the wind speed that was measured during a tropical cyclone.

Environmental epidemiology studies are often prone to misclassification error because the methods of assessing exposure are not always congruent with the way that researchers conduct human impact studies. It is easy to map the path of a tropical cyclone's center, and categorize every county it passes through as an exposed county. However, this information by itself would not give the researcher any information about population centers that the storm passed through or near to. A town within an exposed county may or may not have been close to the storm's path. Conversely,a town in an unexposed county could be located very close to the border of an exposed county, and even be closer to the storm's track than a different town within that exposed county. Where physical exposure data is often collected at point locations, human impact data is often at the level of zip code, county, metropolitan area, or state. It is easy to here how spatially, physical exposure data and human impact data are collected at different resolutions that increase the risk for misclassification error.

Another source of misclassification error in tropical cyclone impact studies is self reported data. Self reported data is used to assess human impacts that are often not known or apparent until after the tropical cyclone event. A great example of this is in (Lieberman-Cribbin et al. 2017) where study subjects were asked to report their own flooding exposure and their mental health symptoms of depression, anxiety, and PTSD. It is reasonable to believe that self perceived exposure to hurricane related flooding would not be independent from

perceived negative mental health symptoms and thus potentially contribute to differential misclassification error in this situation.

Dichotomizing continuous exposure measurements. Sometimes, researchers use an agreed upon threshold to split a continuous metric into a binary classification (exposed or unexposed). For example, a county may be classified as exposed or unexposed based on local winds exceeding a threshold (e.g. gale-force winds or higher). S.C. Grabich et al. 2016 classified hurricane exposure in a Florida county using maximum wind speed. Maximum wind speed is a continuous variable, but the study used binary categorizations to divide it into tropical wind speeds, classified as greater than 39 miles per hour, and hurricane wind speeds, classified as greater than 74 miles per hour. Florida counties experiencing maximum wind speeds below 39 miles per hour were considered unexposed.

Researchers typically dichotomize or categorize continuous variables in several situations for several reasons. They do this typically because it simplifies the data and allows for easier analysis and interpretation (Naggara et al. 2011). Additionally, it is very common in clinical settings to categorize continuous variables, for example hypertensive or not hypertensive, overweight or not overweight, dead or alive, etc. (Van Walraven and Hart 2008).

Despite several advantages to dichotomizing continuous variables that we just discussed, the general consensus in epidemiology is not to do it. Statistical power is lost because so much information is lost when categorization occurs (Van Walraven and Hart 2008). This makes sense when you consider that continuous variables allow you to observe nuance in the data and perceive a dose response relationship between the predictor and response variables, should one exist. This effect is masked when researchers categorize data, and even more so when a smaller number of categorical variables are used (for example dichotomization itself at 2). Generally, if you are going to categorize continuous data, it is better to use 3 or more categories rather than just two, because this will capture more of what the data that would otherwise be lost. An example of a paper that used three different bins was (Kinney et al. 2008), which explored the risk of autism after a pregnancy that included exposure to a tropical storm in the state of Louisiana. The study authors classified tropical storm exposure as severe, intermediate, and low exposure, and these exposure classifications were determined based on whether a mother lived in a Louisiana parish that had both of the exposure factors of interest: storm intensity and storm vulnerability. Storm vulnerability in this case was based on another dichotomy: whether or not the storm center passed through the parish of interest. Storm vulnerability was a measure of how vulnerable the inhabitants of the parish were to the effects of a storm (higher socioeconomic neighborhoods and parishes have more resources to withstand and recover from a tropical storm for example).

Another obvious problem with categorizing continuous data is that the cutoff points are often arbitrary. In the case of dichotomization, the median is often used, but there is typically no reason to assume that the median is a reasonable cutoff point. Because different samples will have different medians, this automatically makes many categorical bins difficult to compare across studies (Altman and Royston 2006). Further, choosing optimal cutoff points that give the smallest p-values can lead to spurious results (Altman and Royston 2006).

Not surprisingly, dichotomizing continuous variables can bias results. A study by Selvin showed that the odds ratios can be significantly different depending on the chosen cutoff that is implemented in a study (Van Walraven and Hart 2008). Categorical variables can also put otherwise similar observations into separate bins if they are close but on opposite sides of the cutoff (Altman and Royston 2006). Choosing a median as a cutoff is intended to delineate bins, but if the bins are a "high" and "low" group, two individual observations that may only be a fraction different, but on either sides of the mean, will be classified as high and low respectively, and give the false impression that they are significantly different.

While dichotomizing continuous variables is something that can be done for either the exposure or the outcome of interest in a study, for our purposes we are primarily interested in continuous *exposures*. This means that we are primarily interested in the effects of dichotomizing variables such as wind speed, rainfall, temperature, distance from storm center, and distance from coastline, among other factors. Many epidemiology studies will dichotomize continuous outcome variables such as blood pressure, body weight (BMI), and length of pregnancy in order to gage medical concern and priorities, but because we are concerend with creating a data framework that makes storm exposure data accessible for epidemiologists, exposure scientits, economists, and other scientists to use, we have a priority to look at exposure variables.

Scales for Categorizing Wind Speeds There are several methods in existence for categorizing wind speed, one of the most frequently used variables for estimating exposure to hurricanes and tropical storms. The first is the Saffir-Simpson scale, which uses five different bins to classify varying levels of wind speed and determine the severity of a storm. The first level, Category 1 is designated for hurricanes and tropical storms with maximum wind speeds of between 64 - 82 knots and is generally considered dangerous to people, livestock, and pets from the hazard of flying and falling debris (Taylor et al. 2010). On the higher end of the scale, Category 5 designates hurricanes with maximum wind speeds above 137 knots and is considered to have catastrophic effect on damage and a high probability of injury or death to people, livestock, and pets even if they are sheltering indoors (Taylor et al. 2010).

Forecasters classify hurricanes into categories on the Saffir-Simpson scale based on maximum sustained surface wind speed. This is defined as the peak one minute wind speed at a height of 10 feet over an unobstructed exposure (Taylor et al. 2010). An important limitation of the Saffir-Simpson scale is that it doesn't account for other hurricane-related impact variables such as storm surges, flooding, and tornadoes (Taylor et al. 2010).

Another scale used to categorize wind speed is the Beaufort scale, created by Admiral Sir Francis Beaufort, used to classify wind speeds both over land and sea. While the Saffir-Simpson scale is only designated for wind speeds that are already at hurricane levels (greater than 64 knots), the Beaufort scale considers the wind speeds below this. The scale ranges from Force 0 (0-1 knots and calm) to Force 12 (64 to 71 knots and hurricane). Other interesting parts of the scale include Force 3 (4-6 knots) which is a gentle breeze, and Force 8 (34-40 knots) which is considered a gale.

Categorizing wind speeds presents researchers with some of the same problems mentioned above that happen when dealing with continuous data, but both scales are based off associations between winds at certain speeds and observed damage and health impacts to communities exposed to these wind speeds.

Aggregate Hurricane Exposure Metrics

Another method of assessing damage and impact of tropical storms and hurricanes is through a single aggegrate exposure metric. While aggregate values often represent the mean of all the values recorded, weather data is typically assessed by the maximum value. This could be something like the maximum wind speed reached in a particular county or parish, or the total monetary cost in damage due to flooding in a metropolitan statistical area. The Saffir-Simpson scale is an example of how entire storms are often classified by their maximum wind speed.

Although using a single exposure value can simplify analysis and interpretation, particularly over an extended temporal scale, there are some obvious drawbacks to relying on one single aggregate value. For example, the Saffir Simpson categories typically correspond only to the geographic point location where the maximum wind speed was observed (Taylor et al. 2010). Hurricane Wilma in 2005 for example, was a Category 3 hurricane when it made landfall on the southwest coast of Florida, but it created Category 1 and Category 2 conditions for the more populous Miami-Dade, Broward, and Palm Beach counties when it finally reached them (Taylor et al. 2010).

Single exposure metrics are often used after a storm event has happened. They are very common in assessing ecological damage after a large hurricane.

Ecological Bias/The Ecological Fallacy Because studying tropical storm and hurricane exposures requires us to look at different spatial scales, we run the risk of encountering the ecological bias when looking at larger spatial aggregations. Ecological bias occurs whenever the aggregate association between an exposure and an outcome does not properly reflect the association on the individual level (Greenland and Morgenstern 1989). Ecological studies themselves don't look at individuals, but rather at an aggregate value, usually within a defined geographic region. Looking at national levels of obesity, cancer, or life expectancy, and comparing countries with respect to these outcomes and some exposure is an example of what ecological studies aim to achieve.

An example of a study that could be prone to this kind of bias is (Kinney et al. 2008) where Louisiana parishes were considered vulnerable to hurricane exposure based on whether or not the storm center passed through that parish. It is possible that the cases considered exposed based on living in these parishes were

not in fact exposed since the storm may have passed through only a certain part of the parish. Never the less, all cases in a parish are considered exposed or unexposed in the aggregate.

[BA: I'm adding some additional text/notes we can work into Claim 2 as appropriate. I drafted these while working on another manuscript but they were more detailed than we needed there, so we can work them in here.]

Measurement error can be either random or systematic. Systematic error can often be corrected with adjustment if the direction and typical size of the error is understood. Random error cannot in the same way. Either type of measurement error can be either differential (associated with the probability of the outcome) or non-differential (independent of the distribution / probability of the outcome). In simpler models, this characteristic might help in predicting whether the resulting bias is likely toward the null; however, more complex models (e.g., statistical models with adjustment for potential confounders) are trickier to diagnose in terms of the likely implications of differential versus non-differential measurement error [would need a ref for this].

When a single value of exposure is assigned across an aggregated level (e.g., a single exposure measurement for a county or ZIP code), it assumes constant exposure across that area. However, this will typically not be the case—hazards like storm-associated wind, rain, and flooding can vary in intensity across these spatial areas. For some hazards, this variation can be notable. Storm surge, for example, will typically be limited to coastal areas of a county or ZIP code. Other hazards, like storm-associated wind and rain, are more likely to be more homogeneous across space, and so have less within-county/ZIP code variation. The rainfields for tropical cyclones are very large, and while there are rainbands within the storm that might have particularly high rates of precipitation, these progress over the course of the storm, and it is unlikely that a county will have one area that experienced very extreme precipitation while another experienced very little [BA: We could see if we could find a good ref. or two on this point]. Similarly, while topographic features and other variability can create variation in the sustained and gust windspeeds experienced in an area from a storm, it is unlikely that one part of a county would experience high-impact winds from a storm while other parts of the county experienced mild wind [BA: We could look for a ref for this, too].

If you use a single exposure estimate for everyone in an area, there is the chance that some people within that area will be misclassified (if exposure is measured as exposed/unexposed) or have exposure measured with error (if a continuous metric of exposure is being used), unless the exposure is perfectly homogeneous across the area. This exposure misclassification or measurement error can lower the power of the study to detect a clear association between exposure to a storm hazard and a certain societal impact, as this smoothing drops information inherent in the within-county variation in exposure levels. It can also bias estimates of the association between exposure and outcome in the same way exposure misclassification through any other mechanism would.

When a proxy exposure estimate (e.g., county-level average exposure level) is used for a group of individuals in the study, it can result in a type of exposure measurement error called Berkson error. In this case, the true exposure of each individual is randomly distributed around the proxy or mean exposure level assigned to him or her. In other words, the group as a whole is assigned a common exposure level, based on the average exposure across that group, when in fact the individuals' true exposure levels are randomly distributed around this common assigned exposure level. [BA: I think this type of error might be a risk when aggregating exposure data, but we should look into it a bit more to make sure I'm right.]

It can be important to think about the scale at which the process happens. For something very local (e.g., aggregating to a very small neighborhood scale), much less information will be lost compared to aggregated to a large scale (e.g., state). If an exposure tends to be fairly homogenous across the spatial scale used for aggregation, then these concerns are lessened (Wakefield and Haneuse 2008).

Data that are aggregated across a spatial area—for example, the total number of deaths in a geographic area in a certain time period—is known as ecological data, aggregate data, or contextual-level data. Studies that

use such data are known as ecological studies (Sedgwick 2014).

Aggregated or ecological data can be used to infer a contextual effect, for example. Sometimes, however, aggregated data are used to infer individual-level associations. While the first type of inference seeks to answer questions like how the county-wide rate of an outcome of interest changes when the county is exposed to a storm hazard, the second seeks to determine how a person's individual risk of an outcome changes if he or she is personally exposed to the hazard. The second type of inference can be prone to bias that results from cross-level inference—the data used to model the association is at the contextual level (e.g., county-level) while the inference is for the individual association between exposure and outcome. When an individual-level association is estimated from ecological data, the estimate can be very biased from the true association, event to the point of reversing the effect estimate—estimating a protective effect, for example, when the true effect is detrimental (Wakefield and Haneuse 2008). This type of bias is called ecological or cross-level bias (Greenland and Robins 1994; Idrovo 2011), and the misconception that associations estimated from data at the ecological/aggregated level provide an unbiased estimate of individual-level associations between exposure and risk of the outcome is called the ecological fallacy (Wakefield and Shaddick 2006; Portnov, Dubnov, and Barchana 2007).

Ecological bias can result both from individual-level exposure measurement error inherent in assigning a common exposure estimate to everyone in an area, while the exposure varies in intensity across that area. It can also result from confounding, even if the confounders are controlled at the ecological level. When data are aggregated across a spatial area, information is lost about how all relevant factors—exposure level, outcome risk, confounders, and even potential effect modifiers—vary within that spatial area. Just as aggregation smooths over within-area variation in exposure levels, it also smooths over within-area variation in levels of potential confounders. Depending on the patterns of this within-area variation, a result could be that ecological-level control of the confounders does not, in fact, control for their role at the individual level, and so the association inferred at the ecologic level continues to be confounded by them when inferred to the individual level. In other words, a factor could still confound the inference of an individual-level association, even if it is controlled at a population level in an ecological model. For example, a study of the association between risk of pre-term birth and tropical cyclone exposure could control for county-level smoking when modeling county-level storm exposure and county-level rates of pre-term births. Even with this control, an observed association could result from differences in individual smoking status, if there is within-county variation in smoking and if this has a different pattern across people in the county than variation in exposure from the county-wide exposure estimate.

If individual-level inference is the aim, and population-level data is available, there are some methods for using it while still aiming to avoid ecological bias. Indeed, it can be helpful to use population-level data, as it is often available for a large population, improving the power and precision of the study (Wakefield and Haneuse 2008; Wakefield and Shaddick 2006). Further, the level of exposure might vary a lot more over the population captured with population-level data compared to the variation that captured in a smaller sample of individual-level data (Wakefield and Haneuse 2008). This can contribute both to statistical power and improve external validity (as the study data will cover more of the range of exposure that might ever be expected). There are ways, for example, to supplement population-level data with samples of individual-level data through two-level, semi-ecologic study designs (Wakefield and Haneuse 2008). Other study designs can also be used to leverage ecological data while minimizing risk from ecological bias. For example, potential confounders like age distribution and smoking rates vary much less within a county over time than comparing between counties. Time series-style study designs, which compare a county to itself over time, therefore allow for very similar covariate distributions between exposure and non-exposure. This can help since the mechanism for ecological bias depends on the joint distribution between individual exposure, outcome, and covariates, if any are included in the model. Other studies add to this design by stabilizing for temporal confounding through the addition of counties that were never exposed, allowing for a differences-in-differences style approach to calibrate for seasonal or longer-term trends that might otherwise create confounding. For example, many health outcomes have a strong seasonal trend, with peak rates in the winter and lows in the summer. Since the hurricane season stretches from summer into fall, a study design that compares the rate of a health outcome in an exposed county to the rate two weeks before the exposure might be biased away from the null, since baseline rates of the health outcome will typically be moving up over most of the hurricane

season.

Ecological bias can also complicate estimation of effect modification, which otherwise could help in identifying vulnerabilities and susceptibilities among certain subpopulations (Wakefield and Haneuse 2008).

For disasters, there are added nuances. First, in some cases, the ecologic-level effect (contextual effect) will be directly of interest. For example, public health planners in a city may be more interested in knowing how a storm hazard exposure is likely to change city-wide rates of certain outcomes than in how it would change individual-level risk. In this case, it is appropriate to use of ecologic-level data, and resulting estimates will not be prone to ecological bias) (Idrovo 2011; Greenland and Robins 1994), although when inferring contextual effects from ecologic data without considering individual-level factors, there is a chance for the *sociologistic fallacy* (Idrovo 2011). (BA: In this last sentence, the cited article is a letter to the editor, and it in turn is referencing some other papers that more fully define these ideas. If we keep this, we should go back and read more deeply into those cited papers from the Idrovo 2011 reference).

Second, for a disaster, the relevant exposure might be not just at the individual level (e.g., winds or flooding at the individual's residence), but also throughout a broader area surrounding the individual. Disasters bring physical hazards that can harm people directly, but also through indirect pathways. The causal pathways for tropical cyclones to affect human health and cause other societal impacts therefore differ from those for a dangerous substance, like air pollutants, in which the substance itself must enter the body to cause harm. While some health risk comes directly from the storm (e.g., deaths and injuries from trees falling on homes or drowning from flooding), there are many more pathways that are indirect. These include pathways that go through the way that the storm's damage affects community infrastructure and access to medical care. For example, a tropical cyclone can bring high winds that cause power outages, and as a result those affected could be exposed to more outdoor hazards (outdoor air pollution, heat), struggle to safely store perishable food and medications, and lose means to power medical equipment. While extreme winds at a person's residence would increase their risk of a power outage, outages could also be caused by damage to the grid in another part of the community. In some cases, then, the level of exposure in a person's community may be as important in opening a pathway of risk as exposure at the person's immediate location.

Finally, if the disaster has a large health impact, the health outcome of one person in the community could affect the risk of the outcome (or other adverse outcomes) for others. This situation is often only the case for infectious diseases, where one person with the disease can spread it to others. However, if the community-wide impact is large enough, it can affect access to and effectiveness of medical care for everyone in the community. if hospitals in the community are over capacity or have to evacuate, this could increase health risk for people in a fairly large "catchment" area for that hospital. This effect has been seen recently with Covid 19—attempts to "flatten the curve" aim to avoid moving into a state where a community's health system becomes overwhelmed and can no longer deliver a typical level of care to those in the community. This effect could happen with either infectious or non-infectious diseases. Also, there may be confounders that are relevant at the contextual, rather than individual level, as well as modifiers. For example, whether the county is coastal could be a contextual-level confounder and effect modifier. This will influence whether the county is exposed to that storm or not, since storms usually weaken rapidly when the center is over land. In terms of confounding pathways, coastal communities might tend to have lower levels of air pollution, because sea breezes clear the pollution regularly. They might also be wealthier on average, since property on or near the beach is desirable. Finally, they might be better prepared for or more hardened against tropical cyclones at the community-wide level (e.g., through hardier power infrastructure, more rigorous building codes, higher likelihood of evacuating in advance of a threatening storm) compared to nearby inland counties.

When inferring individual-level associations from individual-level data, without considering an additional role of ecological-level factors, this is known as the *psychologistic* or *individualistic fallacy* (Idrovo 2011). (BA: In this last sentence, the cited article is a letter to the editor, and it in turn is referencing some other papers that more fully define these ideas. If we keep this, we should go back and read more deeply into those cited papers from the Idrovo 2011 reference).

Discussion

???

Terms

These are terms we're using right now that we might want to iterate on, in conjunction with our colleagues on the project, to make sure we have terms that are precise and consistent across the document:

- **physical exposure data**: By this, we mean things that are measured about the storm like wind speed, rainfall, measures of flooding, and other things that might be considered more in the realm of what an atmospheric scientist or engineer might measure about the storm. We're contrasting this with data that for human impacts studies on outcomes among humans (e.g., pregnancy outcomes, economic outcomes like unemployment)
- **resolution**: We're using this right now to talk about spatial and temporal levels of aggregation. Sometimes, we're using "scales" instead, I think.

-**misalignment**

-**data integration**

- **maximum sustained wind speed**

- **storm track**

- **precipitation**

- **hazard**

- **tropical cyclone**

- **Atlantic Basin**

- **interpolation**

- **aggregation**

- ******

Annual Reviews of Microbiology Annual Reviews of Statistics Annualreviews.org < Good for learning about stuff for interdisciplinary work.

(S. C. Grabich et al. 2016) was another paper that looked at birth outcomes after tropical storms. The researchers in this case found a positive association between exposure to a hurricane and the risk of a pre-term birth.

(Bevilacqua et al. 2020) also found higher levels of PTSD, as well as probable depression and anxiety among residents with a higher Hurricane Exposure Score in Houston, Texas. Displaced Puerto Ricans living in Florida after Hurricane Maria also exhibited higher rates of depression, anxiety, and PTSD(Scaramutti et al. 2019). These mental health outcomes were compared to Puerto Ricans living on the island, and the individuals who had migrated reported higher frequencies of mental health problems than those who had not. Displacement after a tropical storms is a common human impact that leads to other mental health effects, as well as economic, social, and environmental effects.

Tropical cyclones often disrupt local economies of coastal communities. For example, in Florida, hurricanes lead to demand shocks in the economy with a positive net effect on earnings and negative net effect on employment - counties directly hit by hurricanes experienced up to 4.35% increases in earnings and 4.76% decreases in employment (Belasen and Polachek 2008). This confusing paradox makes sense when one considers that while hurricanes may wipe out local businesses, the post-hurricane recovery period boosts certain businesses and sectors. New Orleans in the aftermath of Hurricane Katrina is evidence of this. While the city initially experienced the negative effects of a shut down economy, several years after the storm revealed

that victims of Hurricane Katrina in New Orleans experienced increased income relative to cities not affected by the storm, perhaps due to a strengthed labor market in post-Katrina New Orleans (Deryugina, Kawano, and Levitt 2018). High costs of rebuilding infrastructure, providing resources to displaced populations, medical bills, and loss of businesses after a tropical storm all drive these economic burdens.

Non-differential misclassification refers to misclassification of either the exposure or the outcome, that is unrelated to the other (Aschengrau and Seage 2013). The effect of misclassifying exposures will often, though not always, bias the results of outcome towards the null (Armstrong 1998). In effect, this will weaken or obscure any associations that are present that the researcher may hope to observe in the data (Armstrong 1998).

Differential misclassification error occurs when the misclassification of the outcome is related to the misclassification of the exposure or vice versa (Aschengrau and Seage 2013). While non-differential misclassification often (though not always) has the effect of moving the observed association or parameter towards the null, differential misclassification can move the observation in either direction.

Spatial and temporal misalignment is a problem that researchers run into when integrating data from human impact studies with physical exposure data. For example, physical exposure data on windspeed may have a very fine resolution, possibly down to seconds or minutes, while data on birth outcomes may be at a temporal scale of weeks or even months.

References

Altman, Douglas G, and Patrick Royston. 2006. "The Cost of Dichotomising Continuous Variables." *Bmj* 332 (7549). British Medical Journal Publishing Group: 1080.

Aschengrau, Ann, and George R Seage. 2013. *Essentials of Epidemiology in Public Health.* Jones & Bartlett Publishers.

Bayleyegn, Tesfaye, Amy Wolkin, Kathleen Oberst, Stacy Young, Carlos Sanchez, Annette Phelps, Joann Schulte, Carol Rubin, and Dahna Batts. 2006. "Rapid Assessment of the Needs and Health Status in Santa Rosa and Escambia Counties, Florida, After Hurricane Ivan, September 2004." *Disaster Management & Response* 4 (1). Elsevier: 12–18.

Belasen, Ariel R, and Solomon W Polachek. 2008. "How Hurricanes Affect Wages and Employment in Local Labor Markets." *American Economic Review* 98 (2): 49–53.

Bevilacqua, Kristin, Rehana Rasul, Samantha Schneider, Maria Guzman, Vishnu Nepal, Deborah Banerjee, Joann Schulte, and Rebecca M Schwartz. 2020. "Understanding Associations Between Hurricane Harvey Exposure and Mental Health Symptoms Among Greater Houston-Area Residents." *Disaster Medicine and Public Health Preparedness.* Cambridge University Press, 1–8.

Bianchette, TA, K-B Liu, NS-N Lam, and LM Kiage. 2009. "Ecological Impacts of Hurricane Ivan on the Gulf Coast of Alabama: A Remote Sensing Study." *Journal of Coastal Research.* JSTOR, 1622–6.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." *American Economic Journal: Applied Economics* 10 (2): 202–33.

Grabich, SC, J Horney, C Konrad, and DT Lobdell. 2016. "Measuring the Storm: Methods of Quantifying Hurricane Exposure with Pregnancy Outcomes." *Natural Hazards Review* 17 (1). American Society of Civil Engineers: 06015002.

Grabich, Shannon C, Whitney R Robinson, Stephanie M Engel, Charles E Konrad, David B Richardson, and Jennifer A Horney. 2016. "Hurricane Charley Exposure and Hazard of Preterm Delivery, Florida 2004." *Maternal and Child Health Journal* 20 (12). Springer: 2474–82.

Greenland, Sander, and Hal Morgenstern. 1989. "Ecological Bias, Confounding, and Effect Modification."

International Journal of Epidemiology 18 (1). Oxford University Press: 269–74.

Greenland, Sander, and James Robins. 1994. "Invited Commentary: Ecologic Studies—biases, Misconceptions, and Counterexamples." *American Journal of Epidemiology* 139 (8). Oxford University Press: 747–60.

Hagy, James D, John C Lehrter, and Michael C Murrell. 2006. "Effects of Hurricane Ivan on Water Quality in Pensacola Bay, Florida." *Estuaries and Coasts* 29 (6). Springer: 919–25.

Harville, Emily W, Tri Tran, Xu Xiong, and Pierre Buekens. 2010. "Population Changes, Racial/Ethnic Disparities, and Birth Outcomes in Louisiana After Hurricane Katrina." *Disaster Medicine and Public Health Preparedness* 4 (S1). Cambridge University Press: S39–S45.

Idrovo, Alvaro J. 2011. "Three Criteria for Ecological Fallacy." *Environmental Health Perspectives* 119 (8). National Institute of Environmental Health Sciences: a332–a332.

Jaycox, Lisa H, Judith A Cohen, Anthony P Mannarino, Douglas W Walker, Audra K Langley, Kate L Gegenheimer, Molly Scott, and Matthias Schonlau. 2010. "Children's Mental Health Care Following Hurricane Katrina: A Field Trial of Trauma-Focused Psychotherapies." *Journal of Traumatic Stress: Official Publication of the International Society for Traumatic Stress Studies* 23 (2). Wiley Online Library: 223–31.

Kinney, Dennis K, Andrea M Miller, David J Crowley, Emerald Huang, and Erika Gerber. 2008. "Autism Prevalence Following Prenatal Exposure to Hurricanes and Tropical Storms in Louisiana." *Journal of Autism and Developmental Disorders* 38 (3). Springer: 481–88.

Lane, Kathryn, Kizzy Charles-Guzman, Katherine Wheeler, Zaynah Abid, Nathan Graber, and Thomas Matte. 2013. "Health Effects of Coastal Storms and Flooding in Urban Areas: A Review and Vulnerability Assessment." *Journal of Environmental and Public Health* 2013. Hindawi.

Lieberman-Cribbin, Wil, Bian Liu, Samantha Schneider, Rebecca Schwartz, and Emanuela Taioli. 2017. "Self-Reported and Fema Flood Exposure Assessment After Hurricane Sandy: Association with Mental Health Outcomes." *PLoS One* 12 (1). Public Library of Science: e0170965.

Naggara, O, J Raymond, F Guilbert, D Roy, A Weill, and Douglas G Altman. 2011. "Analysis by Categorizing or Dichotomizing Continuous Variables Is Inadvisable: An Example from the Natural History of Unruptured Aneurysms." *American Journal of Neuroradiology* 32 (3). Am Soc Neuroradiology: 437–40.

Portnov, Boris A, Jonathan Dubnov, and Micha Barchana. 2007. "On Ecological Fallacy, Assessment Errors Stemming from Misguided Variable Selection, and the Effect of Aggregation on the Outcome of Epidemiological Study." *Journal of Exposure Science & Environmental Epidemiology* 17 (1). Nature Publishing Group: 106–21.

Scaramutti, Carolina, Christopher P Salas-Wright, Saskia R Vos, and Seth J Schwartz. 2019. "The Mental Health Impact of Hurricane Maria on Puerto Ricans in Puerto Rico and Florida." *Disaster Medicine and Public Health Preparedness* 13 (1). Cambridge University Press: 24–27.

Schwartz, Rebecca M, Stephanie Tuminello, Samantha M Kerath, Janelle Rios, Wil Lieberman-Cribbin, and Emanuela Taioli. 2018. "Preliminary Assessment of Hurricane Harvey Exposures and Mental Health Impact." *International Journal of Environmental Research and Public Health* 15 (5). Multidisciplinary Digital Publishing Institute: 974.

Sedgwick, Philip. 2014. "Ecological Studies: Advantages and Disadvantages." *Bmj* 348. British Medical Journal Publishing Group.

Taylor, Harvey Thurm, Bill Ward, Mark Willis, and Walt Zaleski. 2010. "The Saffir-Simpson Hurricane Wind Scale." *Atmospheric Administration: Washington, DC, USA*.

Van Walraven, Carl, and Robert G Hart. 2008. "Leave'em Alone-Why Continuous Variables Should Be Analyzed as Such." *Neuroepidemiology* 30 (3). S. Karger AG: 138.

Wakefield, Jon, and Sebastien J-PA Haneuse. 2008. "Overcoming Ecologic Bias Using the Two-Phase Study Design." *American Journal of Epidemiology* 167 (8). Oxford University Press: 908–16.

Wakefield, Jon, and Gavin Shaddick. 2006. "Health-Exposure Modeling and the Ecological Fallacy." *Bio-*

statistics 7 (3). Oxford University Press: 438–55.

Zahran, Sammy, Daniele Tavani, and Stephan Weiler. 2013. "Daily Variation in Natural Disaster Casualties: Information Flows, Safety, and Opportunity Costs in Tornado Versus Hurricane Strikes." *Risk Analysis* 33 (7). Wiley Online Library: 1265–80.