Tornado-Level Estimates of Socioeconomic and Demographic Variables

Tyler Fricker[1]

Abstract: Tornadoes create a threat to human life. Knowing the conditions that make people vulnerable to this threat is vitally important. Yet, socioeconomic and demographic data are not consistently available at the tornado level, making it hard to obtain this knowledge. In response to this limitation, here a method to estimate socioeconomic and demographic variables in a consistent manner at the tornado level for historical events is implemented and assessed. The dasymetric method uses data from the 1990 and 2000 Censuses, as well as the 2010 American Community Survey together with tornado reports over the period 1995–2016. Results show that a typical casualty-producing tornado affects 34 people with an interquartile range between 4 and 198 people. Results also show that the Detroit tornado of July 2, 1997, with its 90 known injuries, likely affected nearly 101,752 people. Comparisons between estimates using the actual path and a simplified modeled path show strong correspondence (percentage errors averaging less than 10%) and estimates compare favorably (correlations exceeding 0.90) with known demographic numbers from a sample of tornadoes, indicating the procedure provides useful information for statistical studies of tornado vulnerability. **DOI: 10.1061/(ASCE)NH.1527-6996.0000379.**

Author keywords: Tornado; Casualties; Demographics.

Introduction

The United States experiences more tornadoes—rapidly rotating columns of air—than anywhere else on Earth (Grazulis 1990). Tornadoes affecting urban areas have the potential to cause hundreds or thousands of casualties—injuries and deaths. Data from the Storm Prediction Center (SPC) of the National Oceanic and Atmospheric Administration show that the April 27, 2011, Tuscaloosa-Birmingham, Alabama, tornado resulted in 1,564 casualties, including 64 deaths. Less than a month later, the May 22, 2011, Joplin, Missouri, tornado resulted in 1,308 casualties, including 158 deaths. In 2011 alone, tornadoes were responsible for 553 deaths in the United States (NCEI 2012). More recently, the December 26, 2015, Garland-Rowlett, Texas, tornado resulted in 478 casualties including 10 deaths.

Fast winds, short warning times, and quality of built environments are the leading proximal causes of casualties (Greenough et al. 2001). In addition, research shows that beyond tornado strength and the number of people in harm's way (Fricker et al. 2017a), per-tornado casualty counts depend on housing stock (permanent or mobile) and age and income of the occupants (Bohonos and Hogan 1999; Greenough et al. 2001; Simmons and Sutter 2005, 2008, 2009). Other socioeconomic and demographic determinants include poverty and education (Simmons and Sutter 2014; Lim et al. 2017), but the significance and relative importance of these other factors remain in question, partly owing to the application of different estimation approaches.

Using county-level socioeconomic and demographic data and a straight line as a model for the so-called tornado footprint, Simmons and Sutter (2005, 2008, 2009) evaluate risk factors for casualties and find a positive correlation with income and percentage of mobile homes after controlling for the damage rating. Lim et al. (2017) find similar correlations but note a positive relationship between deaths and percentage of households headed by women when other variables like income, poverty rate, educational attainment, and mobile homes are omitted. Simmons and Sutter (2014) use the same approach to predict per-tornado fatalities during 2011. Fricker et al. (2017a) use a more detailed model for the tornado footprint and produce tornado-level estimates of energy dissipation and population with a dasymetric approach on grid-level data. Masoomi and van de Lindt (2018) use a similar footprint model to produce tornado-level estimates of population and housing units from Census block-level data. With the considerably more granular data they improve on the predictive skill of Simmons and Sutter (2014)—on the same set of violent tornadoes in 2011—using maximum damage rating, path length, and number of people in the tornado path as fixed effects.

The aforementioned studies demonstrate the potential of statistical models for understanding tornado vulnerability. However, a drawback to realizing this potential is the inconsistency in the data, which can lead to contradictory results even when the model is the same, as well as the different tornado models and levels of aggregation used by researchers. In response, the main purpose of the present work is to provide a consistent set of socioeconomic and demographic numbers at the tornado level in order to benchmark the performance of models used in analyzing and predicting tornado casualties. The approach to obtain these numbers is to use a suitable level of aggregated socioeconomic and demographic data and to include potential relevant variables like race, age, and income. The result is a data set of per-tornado estimates that will be used to analyze and predict casualties with greater accuracy and consistency. Importantly, the numerical estimates highlighted in this paper can be reproduced exactly using the available code developed from the R project for statistical computing (R Core Team 2016).

[1]Dept. of Geography, Florida State Univ., 113 Collegiate Loop, Tallahassee, FL 32306. ORCID: https://orcid.org/0000-0002-8186-7399. Email: tfricker@fsu.edu

Note. This manuscript was submitted on May 28, 2018; approved on November 15, 2019; published online on March 17, 2020. Discussion period open until August 17, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Natural Hazards Review***, © ASCE, ISSN 1527-6988.

Data

Historical Tornado Reports

Official tornado reports are obtained from the US SPC. The SPC database has information related to the spatial location and dimensions of damage paths. The database is compiled from National Weather Service (NWS) *Storm Data* and includes all known tornadoes dating back to 1950. Specifically, each tornado record includes information on initiation point (latitude and longitude), date, length and width of the damage path, and maximum damage rating from 0 to 5 [in Fujita (F) (prior to February 2007) or Enhanced Fujita (EF) (thereafter) damage scale]. It also includes the number of deaths and injuries. Tornado reports in the database are compiled initially by the NWS offices and reviewed by the National Centers for Environmental Information (formerly the National Climate Data Center) (Verbout et al. 2006) before being entered into the database. The database is available in a shapefile format, with each tornado represented geometrically as a straight-line track between the start (initiation point) and end locations. Here all tornadoes in the database from the historical period 1995–2016 are considered. The start year coincides with the period of records in which maximum path width was adopted by the NWS. The end year is the last year available to the authors at the time of analysis.

Tornado reports include information on the number of deaths and injuries. When available, they also include information on the age and sex of each fatality. For example, the *Storm Events* database lists the fatality details from the May 22, 2011, tornado in Joplin, Missouri, showing the age and sex of each death. But these event-fatality details are not available for every tornado. Details from the May 29, 1995, Great Barrington, Massachusetts, tornado that produced 27 casualties, with 3 deaths, for example, include no information on either the age or sex of the victims. Moreover, missing from all tornado reports is information on the race or socioeconomic status of the victims. For example, although it is known that 54% of the deaths from the 2011 Joplin, Missouri, tornado were women, no information on the race or income of these casualties exists.

Census Tract Socioeconomic and Demographic Data

Socioeconomic and demographic data are obtained from the US Census Bureau and the American Community Survey (ACS). The associated boundary shapefiles are obtained from the topologically integrated geographic encoding and referencing (TIGER) database. The ACS is a nationwide survey that collects and produces information on demographic, social, economic, and housing characteristics every year. The survey includes information at the state, county, tract, block group, and block level. Here census tract data are used so as to be large enough to include estimates of most socioeconomic and demographic variables and small enough to give more detailed estimates than can be given for other geographies. This study includes the 1990 and 2000 Censuses, as well as the 2010 ACS 5-year estimates. The 2010 Census is not considered because it is a short-form-only census that asks for information on just name, sex, age, date of birth, race, ethnicity, relationship, and housing tenure.

Differences in the quality of data exist between the decennial census—here 1990 and 2000—and the ACS. Though the long form remained largely consistent between the 2000 Decennial Census and the ACS, the data collection strategy changed quite drastically. For example, the 2000 Census sampled approximately 19 million households to collect long-form data, while the ACS contacts a sample of households each month, amounting to approximately 3.54 million households each year (Folch et al. 2016). As a result of this reduction in sample size, the margins of error associated with ACS estimates at most geographies increased, making it important to recognize the challenges of using ACS estimates, in part, to their potentially high levels of uncertainty (Macdonald 2006; Salvo and Lobo 2006; Bazuin and Fraser 2013; Spielman et al. 2014; Folch et al. 2016).

Tornado Model and Dasymetric Method

A model for each tornado path is made using a buffer on the straight-line track in accordance with the recorded path width. Over the conterminous United States during the period 1995–2016, there were 26,863 tornadoes. Of these, 2,208 were casualty-producing, and of these 2,201 were casualty-producing tornadoes that occurred over a geography with socioeconomic and demographic data available (Fig. 1). Estimates of social correlates within the path are computed for each tornado. This is done using a dasymetric procedure similar to that used in Fricker et al. (2017b). The procedure requires two sets of volumetric areal data. The first set is the spatial path of the tornado. The second set is socioeconomic and demographic data and its areal representation as census tracts.

The dasymetric calculations are similar to those made in Fricker et al. (2017b) but differ in perspective. Fricker et al. (2017b) use tornado-level information (e.g., the number of injuries) to estimate where (e.g., what jurisdiction) those injuries are most likely to have occurred. Here jurisdiction-level information (e.g., census tracts) is used to estimate how many people by socioeconomic or demographic group were in the path of the tornado. More specifically, the central premise is that a reliable estimate of tornado-level socioeconomic and demographic data can be made with spatial apportionment of the census information. Using a ratio of the fraction of the tornado path that occurs within a census tract and the total area of the census tract, weighted estimates of variables (e.g., household median income) can be made for each fraction of the tornado path. When added together, the result is an estimate of the variable (e.g., household median income) for the entire tornado path (Fig. 2). Final values are determined using linear interpolation based on the year of occurrence. Tornadoes on or after 2010 are assigned estimates based on the year 2010.

With the goal of better understanding tornado casualties, socioeconomic and demographic variables that can influence the rate of tornado casualties are further examined. Variables chosen are based on previous research (Simmons and Sutter 2005, 2008, 2009, 2011; Ashley 2007; Paul and Stimers 2012; Paul et al. 2014; Lim et al. 2017; Masoomi and van de Lindt 2018) and include total population, population density, male population, female population, white population, black population, household median income, and number of mobile homes. Of additional interest is population by age groups (under 18, 18–44, 45–64, and over 65). Population data (total population, male population, female population, white population, black population, and population by age) are in number of people. Population density is in people per square kilometer. Household median income data are in 2015 dollars, converted using the Consumer Price Index (CPI-U), and mobile home data are in number of mobile homes.

As previously mentioned, it is important to recognize the challenges in using ACS estimates. Here, the challenge of creating reliable estimates is rooted in the margins of error (MOE) attached to the 2010 ACS 5-year estimates (Table 1). For the seven variables of interest with direct MOE (total population, number of males, number of females, white population, black population, household

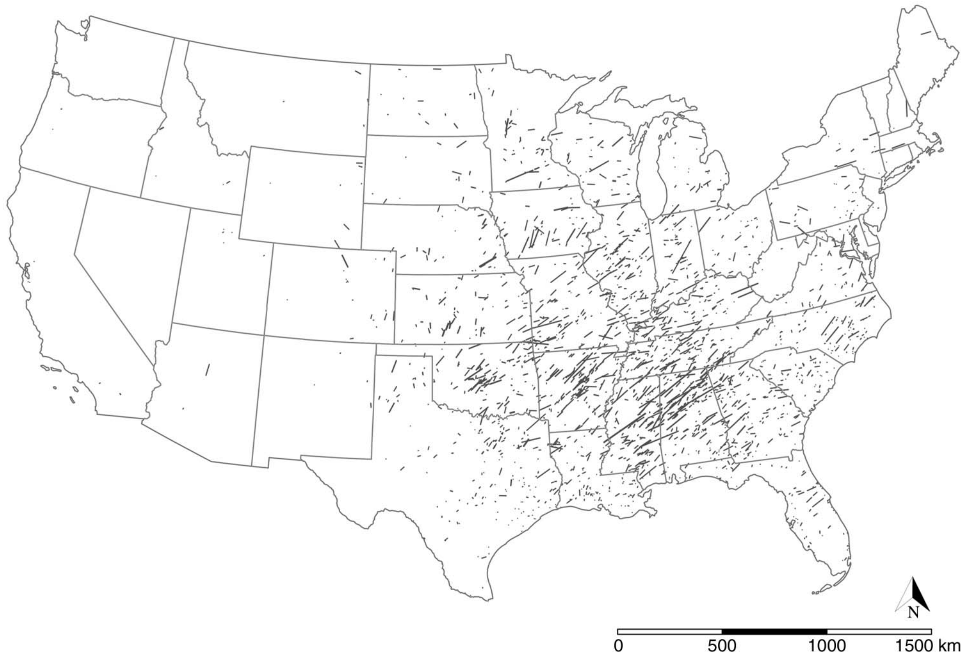

Fig. 1. Casualty-producing tornadoes with socioeconomic and demographic data available. Paths of tornadoes with at least one casualty over the period 1995–2016 are shown as (thin) rectangles.

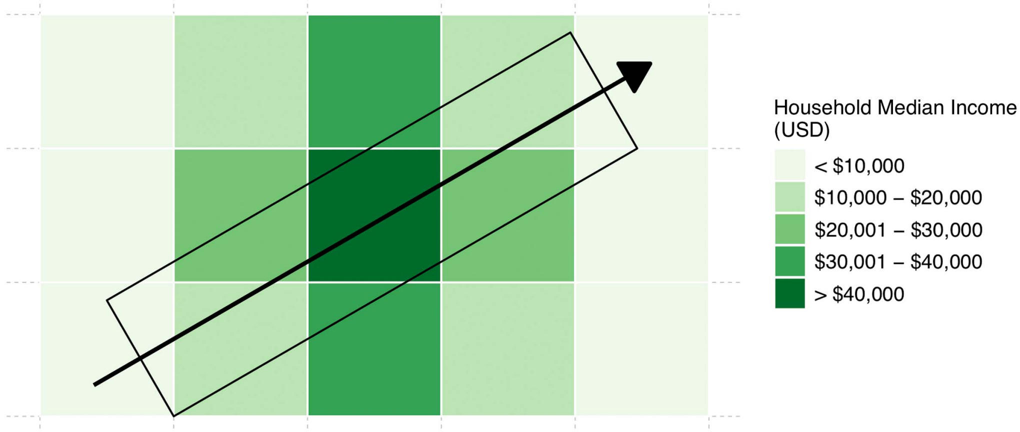

Fig. 2. Idealized models of tornado path and household median income. The straight-line tornado track (arrow) and damage path (rectangle) are shown over household median income represented by the shaded square cells. The ratio of the fraction of the tornado path that falls within a census tract (square cells) and the total area of the census tract are determined by a percent overlay. The weighted estimates of each segment of the tornado path are added to come up with an estimate of household median income for the entire tornado path.

Table 1. Summary of margins of error associated with 2010 ACS 5-year estimates

Variable	Mean	25th percentile	Median	75th percentile
Total population	9.2	6.5	8.3	10.5
Number of males	12	8.9	11	14
Number of females	12	8.5	11	13
White population	13	7.7	11	15
Black population	46	28	44	57
Household median income	15	11	14	18
Number of mobile homes	39	24	36	56

Note: Margins of error are reported as percent error (%). Margins of error for the number of mobile homes is calculated with tracts containing an estimate of zero removed.

median income, and number of mobile homes), the lowest associated MOE for all tracts is found in population counts (excluding black population) and household median income. Conversely, the largest associated MOE is found in the number of mobile homes, which is likely the result of a small number problem. In fact, 35% of all tracts in the 2010 ACS 5-year estimates have less than 50 mobile homes. For the five variables of interest with no direct MOE (population density and population by age groups), a new MOE was calculated using the Census Bureau's directions for derived estimates (US Census Bureau 2008). The associated MOE range from an average low of 2% (people 18–44) to an average high of 9% (population density), with a median low of 2% (people 18–44) to a median high of 8% (population density), and an interquartile range between 2% and 11%.

Table 2. Top 10 tornadoes ranked by total population in their path

Location	Date	Injuries	Deaths	Casualties	Population
Detroit	July 2, 1997	90	0	90	101,752
St. Louis	May 31, 2013	8	0	8	36,840
Pittsburgh	June 2, 1998	50	0	50	27,818
Tuscaloosa-Birmingham, AL	April 27, 2011	1,500	64	1,564	25,793
Springfield, MA	June 1, 2011	200	3	203	20,448
Minneapolis	May 22, 2011	48	1	49	19,174
Bridge Creek-Moore, OK	May 3, 1999	583	36	619	19,157
St. Louis	April 22, 2011	5	0	5	18,594
Hackleburg-Phil Campbell, AL	April 27, 2011	145	72	217	17,141
Wichita, KS	April 14, 2012	38	0	38	16,612

Note: The number of injuries and deaths are given in the SPC tornado database.

Results

Socioeconomic/Demographic Estimates Per Tornado

The procedure results in estimates of 12 variables that can be analyzed independently or in combination with other attributes in the SPC database. For the set of 2,201 tornadoes, the median total population is 33.7 people with an interquartile range between 3.71 and 198 people. The median population density is 20.1 people per square kilometer with an interquartile range between 7.83 and 65.6 people per square kilometer. The method estimates that as many as 101,752 people—given tract-level information of residential population and not actual population numbers—were in the path of the July 2, 1997, Detroit tornado that resulted in 90 injuries. Of the top 10 tornadoes ranked by total population (Table 2), only 1 (the 1997 Detroit tornado) is estimated to have impacted over 100,000 people. The next closest tornado (the 2013 St. Louis tornado) is estimated to have impacted 36,840 people—given tract-level information of residential population and not actual population numbers.

For the same set of tornadoes, the median number of males is 16.9 with an interquartile range between 1.9 and 98. Similarly, the median number of females is 16.9 with an interquartile range between 1.9 and 100. The median white population is 26.8 people with an interquartile range between 2.82 and 155 people, and the median black population is 0.85 people with an interquartile range between 0.04 and 13.2 people. On average, casualty-producing tornadoes have impacted nearly three times as many white people (344 people) as black people (130 people).

Additionally, the median household income is $46,988 with an interquartile range between $39,559 and $56,366, and the median number of mobile homes is 1.68 with an interquartile range between 0.18 and 9.5. It is estimated that as many as 821 mobile homes were in the path of the April 27, 2011, Hackleburg–Phil Campbell, Alabama, tornado that resulted in 145 injuries and 72 deaths. Of the top 10 tornadoes ranked by the number of mobile homes (Table 3), only 2 (1999 Bridge Creek–Moore, Oklahoma, and 2012 Wichita, Kansas, tornadoes) occurred in states outside of the Southeast. Five of the top 10 tornadoes ranked by the number of mobile homes occurred in the state of Alabama alone.

Comparisons between Estimates Using Actual and Modeled Paths

The reliability of the per-tornado estimates of socioeconomic and demographic variables hinges on the assumption that a rectangular—or buffered straight-line—path is a reasonable approximation to the actual path. The start and end locations plus the path width define a rectangular path model, but tornadoes never exactly fit this model. Fig. 3 shows modeled and actual paths for 20 casualty-producing tornadoes. The actual paths are downloaded from the NWS Damage Assessment Toolkit (DAT), which is a geographic information system (GIS)–based framework for collecting, storing, and analyzing damage survey data, using the EF scale for damage classification (Fricker et al. 2014). The choice of tornadoes for comparison is a combination of random selection and availability conditional on the most extreme events—with the highest number of casualties. As can be seen, for some tornadoes the model provides an excellent overlay to the actual path, but not for all. Important are the differences in socioeconomic and demographic estimates for modeled and actual paths, which can be seen in Table 4.

Of the 20 casualty-producing tornadoes with both modeled and actual paths, the largest overestimation for total population is the 2011 Tuscaloosa-Birmingham, Alabama, tornado. The procedure

Table 3. Top 10 tornadoes ranked by number of mobile homes in their path

Location	Date	Injuries	Deaths	Casualties	Mobile homes
Hackleburg-Phil Campbell, AL	April 27, 2011	145	72	217	821
Wichita, KS	April 14, 2012	38	0	38	635
Shoal Creek Valley-Ohatchee, AL	April 27, 2011	85	22	107	616
Vilonia, AR	April 25, 2011	16	4	20	598
Cordova, AL	April 27, 2011	54	13	67	561
Bridge Creek-Moore, OK	May 3, 1999	583	36	619	519
Tuscaloosa-Birmingham, AL	April 27, 2011	1,500	64	1,564	494
Auburn, AL	November 16, 2011	4	0	4	451
Tallulah-Yazoo City-Durant, LA	April 24, 2010	146	10	156	409
Little Rock, AR	March 1, 1997	40	10	50	389

Note: The number of injuries and deaths are given in the SPC tornado database.

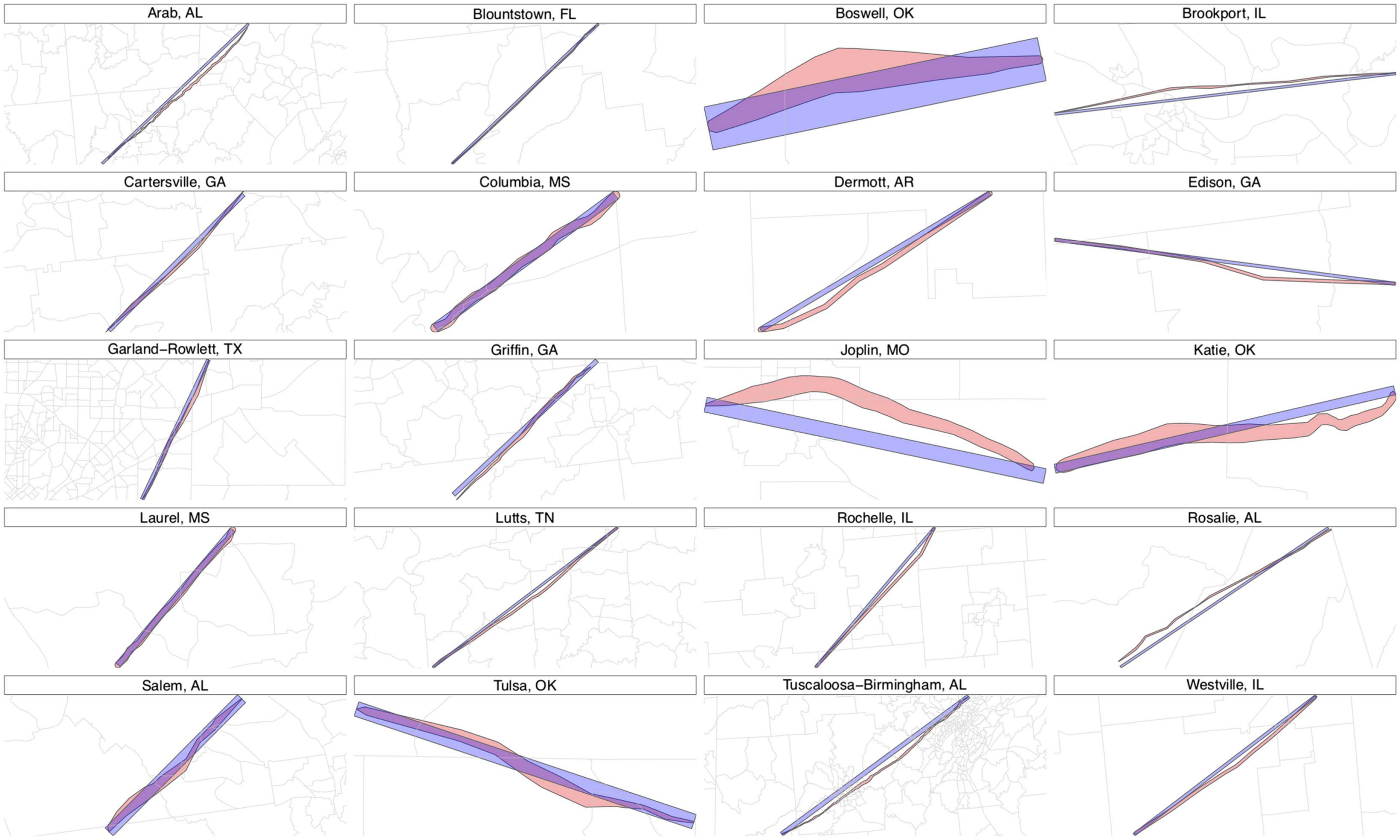

Fig. 3. Actual and modeled paths for 20 casualty-producing tornadoes. The actual tornado paths are curved and the modeled tornado paths are rectangular. Census tract boundaries are displayed below the tornado paths.

Table 4. Differences in demographic numbers estimated from actual and modeled tornado paths

	Total population		Population density		Median household income		Mobile homes	
Location	Actual path	Model path	Actual path	Model path	Actual path	Model path	Actual path	Model path
Tuscaloosa-Birmingham, AL	18,915 (±1,740)	25,793 (±2,373)	78 (±7)	106 (±10)	$48,211 (±$7,232)	$50,361 (±$7,554)	343 (±134)	494 (±192)
Joplin, MO	14,062 (±1,294)	2,688 (±242)	352 (±32)	67 (±6)	$43,612 (±6,542)	$48,647 (±$7,297)	143 (±56)	122 (±48)
Garland-Rowlett, TX	5,051 (±465)	4,837 (±445)	609 (±55)	584 (±53)	$82,456 (±$12,369)	$82,514 (±12,377)	48 (±19)	40 (±15)
Arab, AL	2,014 (±185)	2,668 (±245)	42 (±4)	56 (±5)	$44,493 (±$6,674)	$44,306 (±6,646)	128 (±50)	158 (±61)
Salem, AL	1,033 (±95)	942 (±87)	65 (±6)	59 (±5)	$52,204 (±$7,831)	$51,497 (±$7,724)	125 (±49)	136 (±53)
Columbia, MS	961 (±88)	756 (±69)	76 (±7)	59 (±5)	$37,028 (±5,555)	$36,964 (±5,554)	34 (±13)	26 (±10)
Griffin, GA	844 (±78)	1,184 (±109)	38 (±3)	54 (±5)	$51,408 (±$7,711)	$51,030 (±$7,654)	78 (±31)	101(±40)
Rochelle, IL	834 (±77)	648 (±60)	34 (±3)	27 (±2)	$67,139 (±10,071)	$66,928 (±10,039)	24 (±9)	19 (±8)
Lutts, TN	783 (±72)	618 (±57)	18 (±2)	14 (±1)	$39,800 (±$5,970)	$39,585 (±5,938)	84 (±33)	67 (±26)
Tulsa, OK	665 (±61)	1,037 (±95)	65 (±6)	101 (±9)	$57,431 (±$8,614)	$53,914 (±$8,087)	51 (±20)	52 (±20)
Cartersville, GA	591 (±54)	765 (±70)	25 (±2)	33 (±3)	$64,495 (±$9,674)	$63,063 (±$9,459)	35 (±14)	50 (±20)
Brookport, IL	508 (±47)	924 (±85)	21 (±2)	38 (±3)	$50,811 (±$7,622)	$50,850 (±$7,628)	47 (±18)	67 (±26)
Westville, IL	385 (±35)	301 (±28)	38 (±4)	30 (±3)	$55,898 (±8,385)	$56,319 (±$8,448)	27 (±11)	22 (±9)
Laurel, MS	211 (±19)	159 (±15)	89 (±8)	67 (±6)	$50,749 (±7,613)	$50,850 ($7,628)	17 (±6)	13 (±5)
Boswell, OK	154 (±14)	208 (±19)	3 (±0.3)	4 (±0.4)	$33,729 (±5,059)	$33,582 (±5,091)	16 (±6)	22 (±9)
Katie, OK	93 (±9)	49 (±4)	23 (±2)	12 (±1)	$39,750 (±$5,963)	$39,582 (±$5,937)	5 (±2)	2 (±1)
Blountstown, FL	89 (±8)	83 (±14)	8 (±1)	14 (±1)	$36,387 (±5,458)	$36,236 (±5,435)	12 (±5)	20 (±7)
Rosalie, AL	84 (±8)	98 (±9)	26 (±2)	31 (±3)	$38,799 (±5,820)	$38,827 (±$5,824)	8 (±3)	9 (±4)
Dermott, AR	54 (±5)	63 (±6)	17 (±2)	20 (±2)	$34,082 (±$5,112)	$31,898 (±$4,785)	3 (±1)	3 (±1)
Edison, GA	53 (±5)	42 (±4)	12 (±1)	9 (±1)	$27,400 (±4,110)	$27,426 (±$4,114)	6 (±3)	5 (±2)

Note: Associated margins of error are shown in parentheses.

estimates that as many as 25,793 (±2,373) people were in the tornado's path given a modeled path, which is 6,878 more people than the estimated 18,915 (±1,740) people in the tornado's path given the actual path. The largest underestimation for total population is the 2011 Joplin, Missouri, tornado. The procedure estimates that as many as 2,688 (±242) people were in the tornado's path given a modeled path, which is 11,374 fewer than the estimated 14,062 (±1,294) people in the tornado's path given the actual path.

Table 5. Relationship between modeled path estimates of socioeconomic and demographic variables and actual path estimates of socioeconomic and demographic variables

Variable	r	RMSE	Percentage error (%)
Total population	0.84	775	3.7
Population density	0.89	57	2.8
Number of males	0.85	277	1.3
Number of females	0.83	498	4.8
White population	0.67	650	6
Black population	0.99	68	121
Household median income	0.99	575	0.01
Number of mobile homes	0.98	44	2.8

Note: r = Pearson correlation between modeled path estimates and actual path estimates; RMSE = root-mean-square error between modeled path estimates and actual path estimates; and percentage error = percentage difference between RMSE and actual path estimates.

These large over- and underestimates, however, appear to be outliers, as the remaining 18 tornadoes with both modeled and actual paths do not differ by more than 700 people.

Further, modeled path estimates of socioeconomic and demographic values are compared to actual path estimates of socioeconomic and demographic variables for the set of 20 casualty-producing tornadoes. The Pearson correlation between modeled and actual path estimates is above 0.67 for all variables and above 0.83 when the white population variable is removed (Table 5). The root-mean-square error (RMSE) between modeled and actual path estimates for total population is 775 people, and the RMSE for population density, number of males, number of females, white population, and black population is 57 people per square kilometer, 277 people, 498 people, 650 people, and 68 people, respectively. The RMSE between modeled and actual path estimates for household median income is \$575, and the RMSE for the number of mobile homes is 44.

The percentage error is the RMSE divided by the actual path estimates. The average percentage error for all but one (black population) of the variables is below 10% (Table 5). The high percentage error of the estimated black population is attributed to the low numbers of black people affected by the 20 tornadoes chosen. When a subset of the 20 tornadoes consisting of only tornadoes with an estimated black population of at least 100 is considered, the percentage error diminishes to 0.05%. As such, the combination of a very high correlation between modeled and actual path estimates and low percentage errors indicates the methodology is quite useful.

To account for potential sampling errors associated with the ACS, COVs are calculated for each census tract intersected by actual tornado paths (Table 6). A COV measures the relative amount of sampling error that is associated with a sample estimate. A small COV indicates that the standard error is small relative to the estimate, while a large COV indicates that the standard error is large relative to the estimate. Small COVs are more reliable than large COVs. For the set of seven variables of interest with direct MOE (total population, number of males, number of females, white population, black population, household median income, and number of mobile homes), the lowest associated COV for tracts is found in population and median household income, while the largest associated COV is found in black population and the number of mobile homes.

There are a few different thresholds for reliability when comparing and evaluating ACS data. High and medium reliability exists at a COV below 30%, while low reliability exists at a COV above 30%. Of the 141 census tracts impacted by an actual path, 54 (38%) have a COV above 30% for black population, while 42 (30%) have a COV above 30% for the number of mobile homes. When only census tracts with associated COVs below 30% are used, the average COV drops to 17% in black populations and 18% in the number of mobile homes—both of which are reliable uncertainty measurements.

Table 6. Summary of COVs associated with 2010 ACS 5-year estimates

Variable	Mean	25th percentile	Median	75th percentile
Total population	6.3	4.5	5.6	7.0
Number of males	8.1	5.9	7.1	8.9
Number of females	7.8	5.7	6.9	8.5
White population	10	5.3	6.6	9.7
Black population	40	15	28	55
Household median income	11	6.7	9.5	12
Number of mobile homes	35	16	22	37

Note: COVs are reported as percentage error (%).

To evaluate whether or not differences in estimates can be found by comparing modeled and actual paths, only those tornadoes impacting tracts with COVs below 30% are considered. Of the initial 20 tornadoes, only 5 impacted at least 1 census tract with a COV above 30% for any variable. For the remaining 15 tornadoes, RMSE and percentage error between the modeled and actual path estimates decrease from previous calculations. Thus, the combination of a very high correlation between modeled and actual path estimates and low percentage errors continues to exist when only those tornadoes overlaying reliable data are used, again indicating the methodology is quite useful.

Correlations between Actual and Estimated Deaths by Groups

The approach to obtaining reliable estimates is validated by comparing results with demographic statistics of fatalities available in the *Storm Events* database (https://www.ncdc.noaa.gov/stormevents/). The database records all tornado segments from 1950 through 2016. The tornado segments are divided by county, and each segment includes both an episode narrative and event narrative. Event fatality data are available within each event and include the type of death—direct or indirect—along with the age, sex, and location—if known—of the victim. No information about injuries is available.

To link the tornado in the SPC database with the associated tornado segments in the *Storm Events* database, information available in both data sources—both state of occurrence and number of deaths—is used. For example, the 2011 Joplin, Missouri, tornado caused 158 direct deaths. To find this tornado in the *Storm Events* database, search for "Missouri" in the State/Area drop-down menu. Next, choose "22 May 2011" as the beginning and end date and "tornado" as the event type. After sorting by Death/Injury, the *Storm Events* database shows an event occurring in Jasper County, Missouri, that caused 158 deaths. Choosing the hyperlinked location opens the *Storm Events* database, where information on the number of deaths and the ages and sex of the deceased exists.

Estimated deaths by age and sex from the methodology are compared to observed deaths available in the *Storm Events* database for two dozen tornadoes. Estimated deaths by age and sex are found by multiplying the ratio of the age and sex populations relative to total population by the number of recorded deaths. Tornadoes were chosen as a representative sample to create a wide range of possible fatality estimates. Of these tornadoes, the average number of deaths

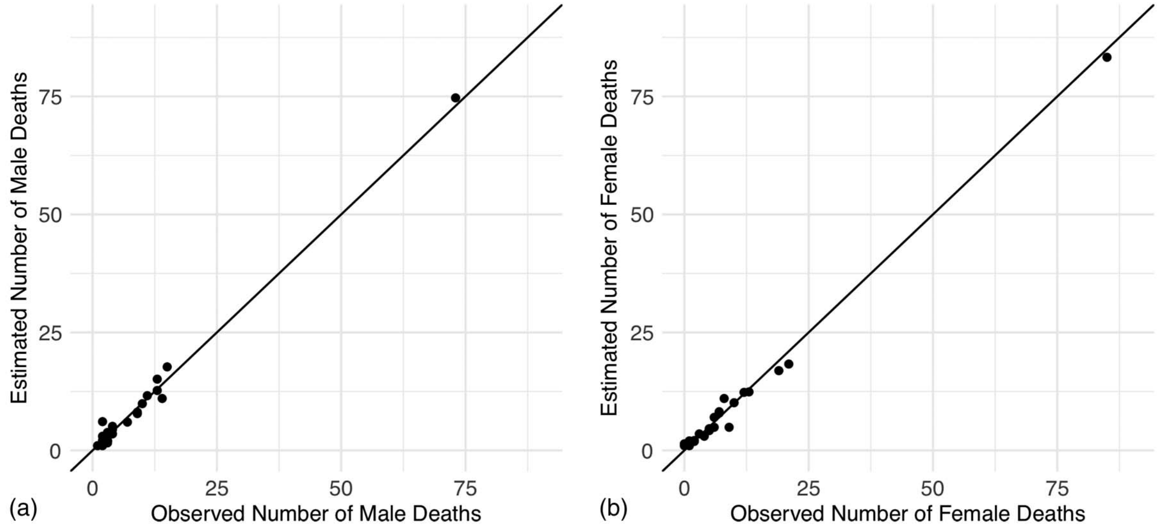

Fig. 4. (a) Observed and estimated number of male deaths; and (b) observed and estimated number of female deaths.

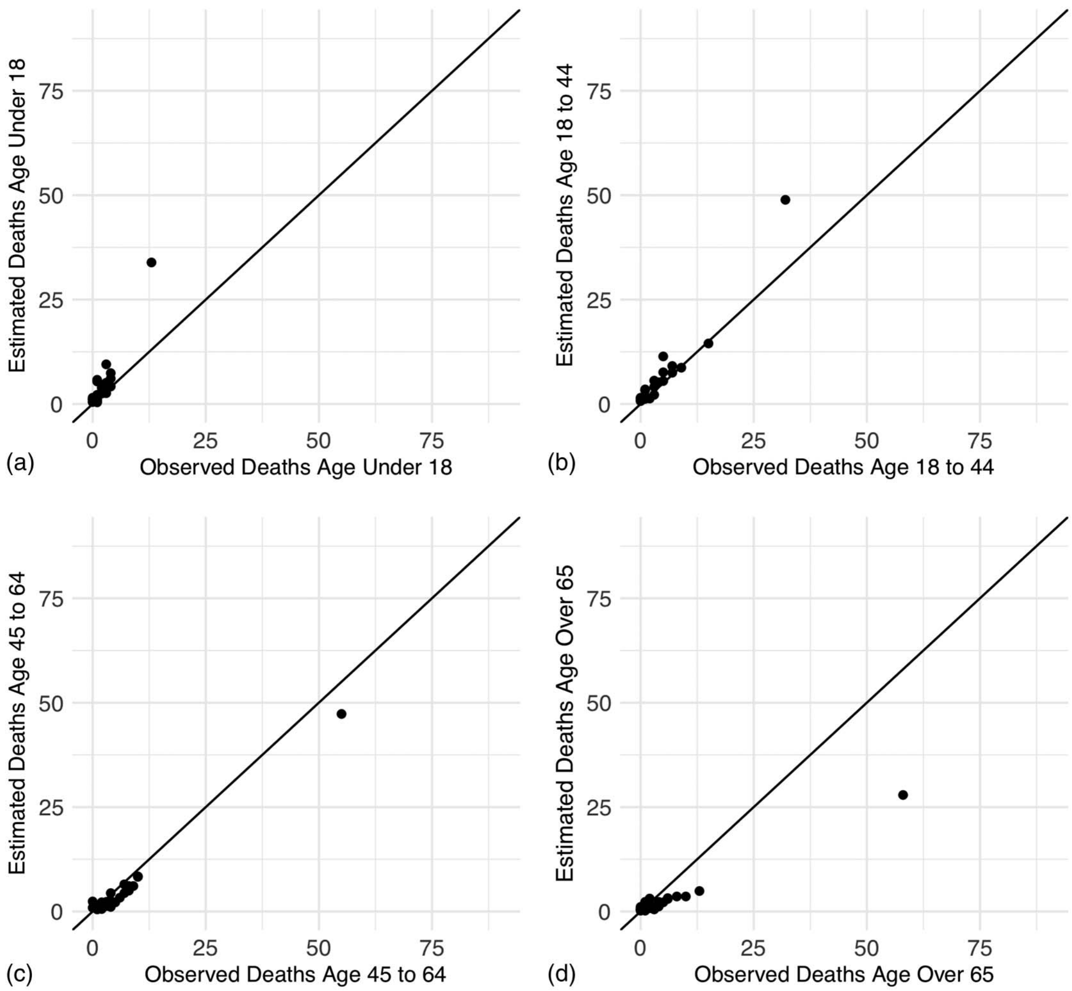

Fig. 5. (a) Observed and estimated number of deaths among people under age 18; (b) observed and estimated number of deaths in 18–44 age group; (c) observed and estimated number of deaths in 45–64 age group; and (d) observed and estimated number of deaths among people over 65.

is 18.3, with a minimum of 2 and a maximum of 158. The Pearson correlation between observed and estimated male deaths is 0.99 ($p < 0.001$), and the correlation between observed and estimated female deaths is 0.99 ($p < 0.001$), both indicating a strong relationship (Fig. 4). The Pearson correlation between observed and estimated deaths for people under the age of 18 is 0.93 ($p < 0.001$), and the correlation between observed and estimated deaths for people 18 to 44, people 45 to 64, and people over 65

(a)

(b)

Fig. 6. Top ten tornadoes by (a) estimated white casualties; and (b) estimated black casualties.

is 0.97 ($p < 0.001$), 0.99 ($p < 0.001$), and 0.99 ($p < 0.001$), respectively, again indicating a strong relationship (Fig. 5).

The very high correlation between the estimated and observed demographics of fatalities again indicates that the methodology is quite useful. Obviously in cases where observed demographics are available, they should be used rather than the estimates. However, observed demographics on tornado casualties are limited to deaths and for sex and age groups only. Importantly, the methodology provides estimates for *any* social or demographic variable of interest and appears to do so in a reliable way. Adding additional verification data will likely change the correlation with an anticipated decrease given the large influence of the Joplin tornado. In fact when that verification point is removed, the correlation drops to 0.72 or above ($p < 0.001$) for age and 0.95 or above ($p < 0.001$) for sex, both of which remain statistically significant.

Utility of the Estimates

Having accurate estimates of socioeconomic and demographic variables at the tornado level makes it possible to infer aggregate demographics. For instance, using the ratio of the white population

Nat. Hazards Rev., 2020, 21(3): 04020018

Fig. 7. Top ten tornadoes by (a) estimated young casualties; and (b) estimated older adult casualties.

relative to the total population, the number of white casualties per tornado can be estimated. Similarly, using the ratio of the black population relative to total population, the number of black casualties per tornado can be inferred. The median number of white casualties for the set of 2,201 tornadoes is 2 (±0.26) people with an interquartile range of between 1 (±0.13) and 6 (±0.78). In comparison, the median number of black casualties for the same set of tornadoes is 0.12 (±0.06) people with an interquartile range between 0.01 (±0.005) and 0.63 (±0.30). The average number of white casualties is nine (±1.2) and the average number of black casualties is slightly less than two [1.9 (±0.87)]. Of the top 10 tornadoes ranked by white casualties (Fig. 6), 3 occurred in the state of Oklahoma, 2 in the state of Alabama, and 1 each in Arkansas, Kentucky, Missouri, Georgia, and Texas. Of the top 10 tornadoes ranked by black casualties (Fig. 6), only 1 (the 1997 Detroit tornado) occurred in a state outside of the Southeast.

Using the ratio of the number of young people (under 18 years old) relative to the total population, the number of young casualties per tornado can be estimated. Similarly, using the ratio of the number of older adults (over 65 years old) relative to the total population, the number of older adult casualties per tornado can be estimated. The median number of young casualties for the set of

Nat. Hazards Rev., 2020, 21(3): 04020018

Table 7. Relationship between estimated socioeconomic and demographic variables and number of deaths and number of injuries

Variable	r_d	r_i
Total population	0.19	0.31
Population density	−0.01	0.01
Number of males	0.20	0.32
Number of females	0.19	0.31
White population	0.32	0.43
Black population	0.05	0.13
Household median income	0.00	0.04
Number of mobile homes	0.43	0.42

Note: r_d = Pearson correlation between estimated variable and number of deaths; and r_i = Pearson correlation between estimated variable and number of injuries.

2,201 tornadoes is 0.64 (±0.09) people with an interquartile range between 0.28 (±0.04) and 1.78 (±0.27). In comparison, the median number of older adult casualties is 0.37 (±0.03) with an interquartile range between 0.17 (±0.01) and 0.99 (±0.08). The average number of young casualties is about three [2.9(±0.04)] and the average number of older adult casualties is less than two [1.6(±0.13)]. Of the top 10 tornadoes ranked by young and older adult casualties (Fig. 7), only 1 (the May 30, 1998 Spencer, South Dakota, tornado) occurred in a state outside the southern Great Plains or Southeast.

With tornado-level aggregated socioeconomic and demographic information, the next step toward a better understanding of casualties is to examine how these factors relate to death and injury counts. For example, the Pearson correlation between estimated socioeconomic and demographic numbers and the counts of deaths and injuries is seen in Table 7. The coefficients range between −0.01 and 0.43 for deaths and between 0.01 and 0.42 for injuries. Moderate correlations are noted between deaths and population, males, females, whites, blacks, and mobile homes. Bivariate relationships have limited utility in this setting where there are many interacting factors, but they provide clues on what factors might be important. For example, the relatively high correlation with mobile homes is consistent with previous research (Ashley 2007; Simmons and Sutter 2005, 2008, 2009; Sutter and Simmons 2009; Lim et al. 2017), as is the relatively high correlation with race (Donner 2007).

How climate change will influence tornado activity and, in turn, tornado casualties remains an open and challenging question. Statistically, given a tornado that produces at least one casualty, the casualty rate depends on the number of people in harm's way and on the power of the winds inside the vortex. Using a regression model, Fricker et al. (2017a) find that casualties increase by 33% (±3%) with a doubling of the tornado energy and that casualties increase by 21% (±3%) with a doubling of the number of people affected, on average. Including an interaction term in the regression model provides a better description of casualties given population and energy (Elsner et al. 2018), but these findings are only the beginning because socioeconomic and demographic variables likely impact the casualty rate. Having estimates of these variables at the individual tornado level provides the information needed to build a model that can predict casualty rates from changes in socioeconomic and demographic variables controlling for population and energy.

Summary

Research shows that beyond tornado strength and the number of people in the tornado's path, per-tornado casualty counts depend on a number of socioeconomic and demographic factors. The significance and relative importance of these factors remain in question owing to inconsistencies in the approaches used to estimate them. In response, this paper provides validated estimates of socioeconomic and demographic numbers at the tornado level. The numbers are validated using known fatalities and actual paths. The strong correlation between estimated and observed fatalities, exceeding 0.93 for four distinct age groups and 0.99 for sex, provides a high level of confidence in the estimates.

The socioeconomic and demographic estimates made are influenced by many factors, including the assumption of a fixed, straight-line representation of the damage path and census data that are based on the location of residence. Greater spatial precision on the damage path like those available in the NWS DAT would undoubtedly improve the estimates. Moreover, it is reasonable to assume that a residential population more closely approximates the number of people in a tornado's path when the tornado strikes at night—and the vast majority of people are at home—than when it strikes during the day. Thus it is likely that time of day will be an important variable in any model that explains casualty rates.

Estimates of socioeconomic and demographic variables can provide new insights into the profiles of the populations affected by tornadoes. Additionally, they can be used to create a spatial understanding of certain place vulnerabilities. Furthermore, these estimates can be used to build risk models for tornado casualties that, along with exposure and energy, include information about the people that are exposed while controlling for factors like poverty, age, ability, and household status. Importantly, building type, building codes, and how building regulations are unequally implemented and enforced across communities will likely be key model variables. Such models will help articulate why some areas are more vulnerable than others and will help to address questions about tornado casualties related to changing socioeconomic and demographic variables. The models can also be used to project future vulnerability by integrating available Census Bureau demographic projections. And while the science behind a potential link between tornadoes and climate change is still in its infancy (Elsner et al. 2015, 2019), any links that do arise can be incorporated through a modification of the energy term.

Finally, it is envisioned that this approach to augmenting an existing per-event database with casualty demographics could be used to spark other, similar approaches for databases on flood and hurricane casualties. The hope is that this work will spur greater efforts toward keeping track of socioeconomic and demographic information on casualties arising from tornadoes, which may provide insight into vulnerabilities that can drive mitigation strategies and policy changes.

Acknowledgments

The author thanks the two anonymous reviewers for their recommendations and feedback on an earlier draft. The work was partially supported by crowd source funding to the author through experiment.com. The code used to produce the tables and graphs, as well as the data set of tornado-level estimates of socioeconomic and demographic variables, is available at https://github.com/tfricker/Tornado-level-socioeconomic-and-demographic-estimates.

References

Ashley, W. 2007. "Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005." *Weather Forecasting* 22 (6): 1214–1228. https://doi.org/10.1175/2007WAF2007004.1.

Bazuin, J. T., and J. C. Fraser. 2013. "How the ACS gets it wrong: The story of the American Community Survey and a small, inner city neighborhood." *Appl. Geogr.* 45 (Dec): 292–302. https://doi.org/10.1016/j.apgeog.2013.08.013.

Bohonos, J. J., and D. E. Hogan. 1999. "The medical impact of tornadoes in North America." *J. Emergency Med.* 17 (1): 67–73. https://doi.org/10.1016/S0736-4679(98)00125-5.

Donner, W. R. 2007. "The political ecology of disaster: An analysis of factors influencing U.S. tornado fatalities and injuries, 1998–2000." *Demography* 44 (3): 669–685. https://doi.org/10.1353/dem.2007.0024.

Elsner, J. B., S. C. Elsner, and T. H. Jagger. 2015. "The increasing efficiency of tornado days in the United States." *Clim. Dyn.* 45 (3–4): 651–659. https://doi.org/10.1007/s00382-014-2277-3.

Elsner, J. B., T. Fricker, and W. D. Berry. 2018. "A model for U.S. tornado casualties involving interaction between damage path estimates of population density and energy dissipation." *J. Appl. Meteorol. Climatol.* 57 (9): 2035–2046. https://doi.org/10.1175/JAMC-D-18-0106.1.

Elsner, J. B., T. Fricker, and Z. Schroder. 2019. "Increasingly powerful tornadoes in the United States." *Geophys. Res. Lett.* 46 (1): 392–398. https://doi.org/10.1029/2018GL080819.

Folch, D. C., D. Arribas-Bel, J. Koschinsky, and S. E. Spielman. 2016. "Spatial variation in the quality of American Community Survey estimates." *Demography* 53 (5): 1535–1554. https://doi.org/10.1007/s13524-016-0499-1.

Fricker, T., J. B. Elsner, P. Camp, and T. H. Jagger. 2014. "Empirical estimates of kinetic energy from some recent U.S. tornadoes." *Geophys. Res. Lett.* 41 (12): 4340–4346. https://doi.org/10.1002/2014GL060441.

Fricker, T., J. B. Elsner, and T. H. Jagger. 2017a. "Population and energy elasticity of tornado casualties." *Geophys. Res. Lett.* 44 (8): 3941–3949. https://doi.org/10.1002/2017GL073093.

Fricker, T., J. B. Elsner, V. Mesev, and T. H. Jagger. 2017b. "A dasymetric method to spatially apportion tornado casualty counts." *Geomatics Nat. Hazard. Risk* 8 (2): 1768–1782. https://doi.org/10.1080/19475705.2017.1386724.

Grazulis, T. P. 1990. *Significant tornadoes, 1880–1989: Discussion and analysis*. St. Johnsbury, VT: Environmental Films.

Greenough, G., M. McGeehin, S. M. Bernard, J. Trtanj, J. Riad, and D. Engelbert. 2001. "The potential impacts of climate variability and change on health impacts of extreme weather events in the United States." *Environ. Health Perspect.* 109 (May): 191–198. https://doi.org/10.1289/ehp.109-1240666.

Lim, J., S. Loveridge, R. Shupp, and M. Skidmore. 2017. "Double danger in the double wide: Dimensions of poverty, housing quality and tornado impacts." *Reg. Sci. Urban Econ.* 65 (Jul): 1–15. https://doi.org/10.1016/j.regsciurbeco.2017.04.003.

Macdonald, H. 2006. "The American community survey: Warmer (more current), but fuzzier (less precise) than the decennial census." *J. Am. Plann. Assoc.* 72 (4): 491–503. https://doi.org/10.1080/01944360608976768.

Masoomi, H., and J. W. van de Lindt. 2018. "Fatality and injury prediction model for tornadoes." *Nat. Hazard. Rev.* 19 (3): 04018009. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000295.

NCEI (National Centers for Environmental Information). 2012. *State of the climate: Tornadoes-Annual 2011*. Asheville, NC: NCEI.

Paul, B. K., and M. Stimers. 2012. "Exploring probable reasons for record fatalities: The case of 2011 Joplin, Missouri, tornado." *Nat. Hazard.* 64 (2): 1511–1526. https://doi.org/10.1007/s11069-012-0313-3.

Paul, B. K., M. Stimers, and M. Caldas. 2014. "Predictors of compliance with tornado warnings issued in Joplin, Missouri, in 2011." *Disasters* 39 (1): 108–124. https://doi.org/10.1111/disa.12087.

R Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Salvo, J. J., and A. P. Lobo. 2006. "Moving from a decennial census to a continuous measurement survey: Factors affecting nonresponse at the neighborhood level." *Popul. Res. Policy Rev.* 25 (3): 225–241. https://doi.org/10.1007/s11113-006-0008-0.

Simmons, K. M., and D. Sutter. 2005. "WSR-88d radar, tornado warnings, and tornado casualties." *Weather Forecasting* 20 (3): 301–310. https://doi.org/10.1175/WAF857.1.

Simmons, K. M., and D. Sutter. 2008. "Tornado warnings, lead times, and tornado casualties: An empirical investigation." *Weather Forecasting* 23 (2): 246–258. https://doi.org/10.1175/2007WAF2006027.1.

Simmons, K. M., and D. Sutter. 2009. "False alarms, tornado warnings, and tornado casualties." *Weather Clim. Soc.* 1 (1): 38–53. https://doi.org/10.1175/2009WCAS1005.1.

Simmons, K. M., and D. Sutter. 2011. *Economic and societal impacts of tornadoes*. Boston: American Meteorological Society.

Simmons, K. M., and D. Sutter. 2014. "Fatality prediction for the 2011 tornado season based on historical extreme weather data." *Nat. Hazard. Rev.* 15 (3): 04014005. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000144.

Spielman, S. E., D. Folch, and N. Nagle. 2014. "Patterns and causes of uncertainty in the American community survey." *Appl. Geogr.* 46 (Jan): 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002.

Sutter, D., and K. M. Simmons. 2009. "Tornado fatalities and mobile homes in the united states." *Nat. Hazard.* 53 (1): 125–137. https://doi.org/10.1007/s11069-009-9416-x.

US Census Bureau. 2008. *A compass for understanding and using American Community Survey data: What general data users need to know*. Suitland, MD: US Census Bureau.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz. 2006. "Evolution of the U.S. tornado database: 1954-2003." *Weather Forecasting* 21 (1): 86–93. https://doi.org/10.1175/WAF910.1.