HHS Public Access
Author manuscript
J Food Prot. Author manuscript; available in PMC 2024 December 06.

Published in final edited form as:
J Food Prot. 2024 December ; 87(12): 100379. doi:10.1016/j.jfp.2024.100379.

Disparities in salmonellosis incidence for US counties with different social determinants of health profiles are also mediated by extreme weather: a counterfactual analysis of Laboratory Enteric Disease Surveillance (LEDS) data from 1997–2019

Daniel L. Weller[1], **Reese Tierney**[1], **Sarah Verlander**[1,2], **Beau B. Bruce**[1], **Erica Billig Rose**[1]

[1]Division of Foodborne, Waterborne, and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, CDC

[2]Oak Ridge Institute for Science and Education, Oak Ridge, TN, United States

Abstract

Understanding disparities in salmonellosis burden is critical for developing effective, equitable prevention programs. Past efforts to characterize disparities were limited in scope and by the analytical methods available when the study was conducted. We aim to address this gap by identifying disparities in salmonellosis incidence between counties with different determinant of health (DOH) profiles. Using national U.S. Laboratory-based Enteric Disease Surveillance (LEDS) data for 1997–2019, age-adjusted county-level salmonellosis incidence/100,000 persons was calculated and linked to publicly available DOH data. We used hurdle counterfactual random forests (CFRF) to quantify, for each DOH, the risk that (i) ≥1 versus no cases were reported by a county, and (ii) when ≥1 case was reported, whether a high (≥16 cases/100,000 persons) or low incidence (≥1<4 cases/100,000 persons) was reported. Risk in both models was significantly associated with demographic DOH, suggesting a disparity between counties with different demographic profiles. Risk was also significantly associated with food, healthcare, physical, and socioeconomic environment. The risk was generally greater for counties with more negative food resources, and for under-resourced counties (e.g., fewer healthcare and social services, fewer grocery stores). Risk was also significantly higher if any extreme weather event occurred. The study also found that underreporting and underascertainment appeared to result in underestimation of salmonellosis incidence in economically marginalized and under-resourced communities. Overall, our analyses indicated that, regardless of other county

*****Correspondence:** Daniel Weller, dweller@cdc.gov, 1600 Clifton Rd, Atlanta, GA.
Author Contributions: DW and ER conceived the project. DW and BB developed the analysis plan, which DW implemented with support from RT. SV compiled the surveillance data. DW compiled the metadata. DW led data cleaning efforts with support from RT and SV. All authors contributed to manuscript development.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Disclaimer: The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control (CDC).

characteristics, extreme weather was associated with increased salmonellosis incidence, and that certain communities were differentially disadvantaged toward a higher incidence. This information can facilitate the development of community-specific prevention efforts.

Keywords

Salmonellosis; Health Equity; Health Disparities; Counterfactual Random Forest; Extreme Weather; Social Determinants of Health; Laboratory Enteric Diseases Surveillance System

Introduction

Salmonella is a leading cause of foodborne illness, hospitalization, and death in the U.S. and globally (Havelaar et al., 2012; Havelaar et al., 2015; Lake et al., 2010; Monteiro Pires et al., 2019; Park et al., 2015; Shah et al., 2024). Indeed, Scallan et al. (2015) determined that *Salmonella* was the leading foodborne cause of disability adjusted life years, hospitalization and death in the U.S. While the estimated mean cost of salmonellosis hospitalization in the U.S. ranged from $12,995 to $29,690 per patient (Cummings et al., 2015; Dhaliwal et al., 2021), the US Department of Agriculture reported that *Salmonella* infections generated $4,142,179,161 in medical costs in the U.S. in 2018 alone (Hoffmann, 2014; Hoffmann et al., 2012). *Salmonella* infections and outbreaks also generate costs for public health agencies and industry. For example, a 1988 outbreak of *Salmonella* linked to eggs cost U.K. public health agencies 26.2 million in USD (Knowles et al., 2007), while a 2008 *Salmonella* outbreak linked to muskmelons cost the fresh produce industry $50 million in USD (Ribera et al., 2012). Reducing salmonellosis burden is, therefore, a public health and economic priority, and the US government set a goal of reducing salmonellosis incidence to 11.5 cases/100,000 persons by 2030 (U.S. Department of Health and Human Services, 2023). While studies have shown that infectious disease burden often varies between communities and populations, recent incidence estimates suggest a lack of progress toward these disease reduction goals (Collins et al., 2022; Shah et al., 2024). The sentinel surveillance system that tracks progress toward Healthy People 2030 foodborne disease goals reported no change in *Salmonella* incidence in its ten-state catchment area in 2023 (16.6 cases/100,000 persons) compared to the baseline period used to measure progress (2016–2018; 16.9 cases/100,00 persons)(Shah et al., 2024). To meet the 2030 goals, it is important to identify communities that are differentially disadvantaged toward an increased salmonellosis burden so we can more effectively tailor public health interventions and guidance.

Determinants of health (DOH) are the underlying socioeconomic or environmental conditions in which people are born, live, and work, including the physical (e.g., land use, weather, extreme weather), healthcare (e.g., access to and quality of care), food (e.g., access to positive and negative food resources) and socioeconomic (e.g., employment rate, educational attainment) environment (see Tables 1 & S1 for additional examples). Multiple studies have identified disparities in foodborne and enteric disease incidence, clinical outcomes, treatment, and disease reporting for individuals living in communities with different DOH profiles (Chang et al., 2009; Gourishankar, 2021; Hadler et al., 2019; Ray et al., 2022; Shaw et al., 2016). Chang et al. (2009) quantified the association between county-level salmonellosis incidence and 26 demographic, socioeconomic, and healthcare environment characteristics in the United States during 1993–2002, and determined that incidence was positively associated with urbanicity, the percent of residents living below the poverty level, and the percent of residents who identified as Black or African American (Black) or Hispanic or Latino (Hispanic). However, Chang et al. (2009) did not consider DOH related to the food or physical (e.g., land use, weather) environments. Chang et al.'s (2009) conclusions were also limited by the data and statistical methods available when their study was conducted; for example, they acknowledged their analyses did not have sufficient confounding control. Shaw et al. (2016) assessed the association between DOH and salmonellosis incidence at the zip code-level and found evidence of state-specific associations (e.g., broiler operation presence was associated with significantly higher salmonellosis rates in three of seven states included in their analyses). While these findings highlight the importance of understanding the association between DOH and salmonellosis incidence, most US studies analyzed data from ≤10 states, and considered a limited number of DOH (Henly et al., 2017; Ray et al., 2024; Younus et al., 2007). Over the last two decades, there have been multiple advancements in disease detection and reporting, as well as the type and diversity of DOH data and analytic methods available (e.g., the development of culture-independent methods) (Collins et al., 2022; Delahoy et al., 2023; Healy et al., 2023; Ray et al., 2022). A national analysis of the determinants of salmonellosis that considers a broad range of DOH and utilizes more recent data and an analytical approach with strong confounding control is therefore needed. This analysis aims to meet these needs by identifying potential disparities in salmonellosis between counties with different DOH profiles and characterizing communities that are differentially disadvantaged toward higher salmonellosis incidence. This analysis adapts a novel analytical approach, counterfactual random forest (CFRF), which has not been used to characterize disparities in infectious disease incidence. Therefore, this study provides a blueprint for how CFRF can be used in future analyses.

Methods

Data:

The CDC Laboratory-based Enteric Disease Surveillance (LEDS) system conducts passive, national, isolate-based salmonellosis surveillance. Cases are reported to LEDS if persons experiencing illness seek care, their doctor orders a test that can detect *Salmonella,* and clinical diagnostic laboratories submit *Salmonella*-positive specimens to the state or local public health lab. LEDS collects patient demographic, serotype, and specimen source data, including county of residence. Using finalized LEDS data from 1997 through 2019, we calculated age-adjusted salmonellosis incidence (per 100,000 persons) for each US county, which was linked to county-level DOH data. Non-US residents and cases that lacked county-of-residence data were excluded.

DOH data were obtained from publicly available sources, such as the Agency for Healthcare Research and Quality (AHRQ; Table S1). The original set of DOH (henceforth the "feature set"; see Table S1) was reduced as previously described (Ray et al., 2024). For each DOH, missing data were singly imputed for each factor using impute.rfsrc from the randomForestSRC package in R with 1,000 trees and 5 iterations, and otherwise default parameters (Ishwaran & Kogalur, 2023). Three rounds of feature reduction were then implemented to remove irrelevant variables that were initially retained to improve imputation (e.g., motor vehicle death rate), remove conceptually redundant or nested variables, and ensure that Spearman's correlation between all retained DOH was <0.65. If two DOHs were correlated or conceptually redundant, the DOH with (i) greater temporal and spatial resolution, (ii) less missing data in the unimputed dataset, (iii) lowest correlation with all other DOHs in the feature set, (iv) a more plausible biological or social association with risk of contracting salmonellosis and having their illness reported to LEDS, or (v) relatively easy interpretation for use in guiding follow-on studies and public health efforts was preferentially retained. DOH were also selected to ensure that metrics representing demographic and socioeconomic characteristics, and food, healthcare, and physical environment were retained (see Table S4 for final feature set). Following feature reduction, all continuous factors, including age-adjusted incidence, were converted to five-year rolling averages using a Cauchy distribution to reduce the impact of single-year events (e.g., outbreaks), differences in temporal resolution between DOH, and the influence of small population counties on the incidence distribution [see Table S4; (Wainer, 2019)].

While conceptually redundant variables (e.g., unemployment and employment rate) represent similar phenomena, correlated variables were not interchangeable. Because only one of each set of correlated DOH was included in the CFRF, our findings must be interpreted cautiously. Temporary shelters/1,000 persons was correlated with all other social services considered (e.g., emergency relief services/1,000 persons, community housing/1,000 residents), and, thus, served as a proxy for social services/1,000 persons. Obesity rate was correlated with other comorbidity rates (e.g., diabetes rate, mental distress rate), and served as an indicator of comorbidity burden. Because the percentage of households without adequate kitchen facilities and plumbing were correlated, the percentage of households without adequate kitchen facilities served as an indicator of housing quality and the availability of in-home sanitation/hygiene facilities. While not correlated with other DOH, flu vaccination and diabetes monitoring rates were considered indicators of healthcare quality.

Counterfactual Random Forest (CFRF):

CFRF (Dasgupta et al., 2014; Malley et al., 2012) was implemented as previously described (Jackson et al., 2021; Ray et al., 2024) to identify DOH associated with annual age-adjusted salmonellosis incidence for each US county during 1997–2019, while adjusting for all other DOH in the model. Random forest analysis has several advantages over other analytical approaches, like regression: (i) more robust confounding control, (ii) scalability (i.e., can handle a large number of features), (iii) does not require explicit specification of model structure, (iv) fewer assumptions, (v) can handle non-parametric data, and (vi) automatically considers all possible interactions in the data (Jackson et al., 2021; Kuhn & Johnson, 2016; Malley et al., 2012). However, the interpretability of traditional random forest algorithms is limited (Jackson et al., 2021; Kuhn & Johnson, 2016; Malley et al., 2012; Weller et al., 2021).

CFRF overcomes this limitation by generating output that is directly comparable to regression (Jackson et al., 2021; Weller et al., 2021). Specifically, CFRF generates odds or risk ratios, which are interpreted as the average change in risk (or odds) when a feature changes by one unit and all other features are held constant (Dasgupta et al., 2014).

In the present study, separate random forest models were developed using those counties with each value of the DOH under consideration. Estimated probabilities for each outcome were generated for each county annually between 1997 through 2019, including counties that reported a different DOH value than the one represented by the given random forest (i.e., counterfactuals). Tables scaled to the exposure margins constructed using the sum of the predicted probabilities tables. Adjusted risk ratios were then calculated using standard methods.

In addition to the DOH in Table S4, year and separate dummy variables for each state and the District of Columbia were included as covariates. Incidence was categorized because CFRF cannot handle continuous outcomes. While random forest can be implemented using continuous covariates, continuous covariates need to be categorized when calculating risk ratios; to facilitate interpretation continuous covariates were categorized prior to running any part of the CFRF analyses. Future studies should consider categorizing after running the random forests but before calculating risk ratios. In the present study, if the distribution of a DOH was strongly skewed (e.g., right-censored, left-censored), categorization was implemented to minimize the effect of skew. For example, chicken density (chickens/100 km2) was strongly right-skewed. Many counties had <10 chickens/100 km2, and most counties with ≥10 chickens/100 km2 had between 10 and 5000 chickens/100 km2. To categorize chicken density, a separate category for counties <1 chicken/100 km2 was created, and thresholds of 5, 10, 50, 100, 500, 1000, and 5000 were applied to counties with ≥1 chicken/100 km2. All other DOH were categorized into ten approximately equal-sized groups using the cut2 function in the Hmisc package (Harrell & Dupont, 2024). The reference level for environmental and demographic DOH was the first group (e.g., <1 chicken/100 km2). The second group was the reference for all other DOH and incidence to account for increased variability associated with small populations (Wainer, 2019)

Because multiple counties did not report any cases each year, separate CFRF were implemented where the outcome was (i) if no cases (reference) or ≥1 case was reported, and (ii) county-level incidence for counties reporting ≥1 case. Like continuous DOH, incidence was converted to a 10-level factor using 1, 4, 7, 10, 13, 16, 19, 22, and 25 cases/100,000 persons as the cut-offs. Counties with ≥16 cases/100,000 persons were considered “high” incidence and counties with <7 cases/100,000 persons were considered “low” incidence for interpretation. The reference was ≥1 but < 4 cases/100,000 persons. These values were selected after considering the mean and five-number summary for smoothed age-adjusted incidence for counties reporting ≥1 case.

The following sensitivity analyses were performed to assess how changes in LEDS reporting over time, and variability in the number of cases with county of residence data affected results.

i. Using data from 2004–2019;

ii. Four separate analyses excluding data from states in years during 1997–2019, where 70%, 50%, 25%, or 5% of cases lacked county of residence data;

iii. Four separate analyses excluding data from states, where 70%, 50%, 40%, or 30% of cases during 1997–2019 lacked county of residence data.

Results

During 1997–2019, 89% of the 922,836 salmonellosis cases reported to LEDS could be linked to county of residence. The percentage of cases missing county of residence varied considerably by state (min=0.3%; max=73%; median=9%; mean=16%; standard deviation [SD]=18%).

The final dataset comprised 71,052 rows, and after smoothing, 10% of rows represented years where the given county reported no cases. The number of counties reporting no cases decreased steadily from 1997 through 2019. When only counties reporting ≥1 case were considered, mean incidence increased consistently from 1997 (mean=9.3; SD=19.4) through 2019 (mean=19.4; SD=15.8). Incidence also varied spatially and appeared to be higher in certain regions (e.g., the Delmarva, Southeast, Mississippi Delta) and lower in others (e.g., New England, Rocky Mountains, Pacific Northwest; Fig. 1).

Counterfactual Random Forest (CFRF).

Results from all sensitivity analyses were generally consistent with the results from the model where no state or year exclusions were applied. Thus, only findings from the latter model are presented here.

Year.—The CFRF identified a strong temporal trend (Fig 3–4, S1; Table S1). Risk that a county reported ≥1 case was >4 times higher in 2015–2016 (95% Confidence Interval [CI]=3.5–4.4) and 2017–2018 (95% CI=3.5–4.4) than 1997–1998. A county's risk of being in a high-incidence group (≥16.0 cases/100,000 persons) increased significantly over time (e.g., risk that a county reported ≥19 and <22 cases/100,000 persons was 7.0 [95% CI=6.5–9.2] times higher in 2019 than 1997–1998; Table S1).

Environmental Context: Agricultural Intensity/Land Cover.—Risk of a county reporting ≥1 case was higher for counties with more wetlands (as percent of area), a higher cattle, chicken or swine (i.e., livestock) density, and more pollutant discharge permits/1,000 km of stream (a surface water quality indicator; Fig 3, S2). Risk of a county being in a high-incidence group if it reported ≥1 case also increased as livestock density and wetland cover increased (Fig 4; Table S5–S6).

Environmental Context: Weather.—Risk of a county reporting ≥1 case was generally lower for counties in years with higher annual precipitation and annual maximum temperatures (Fig 2, S3; Table S5). Risk of a county being in a high incidence group generally increased as annual precipitation increased but decreased as annual maximum temperature increased (Fig 4, S3).

Irrespective of the type of extreme weather (i.e., extreme heat, extreme rain, flooding, hurricanes/tropical storm, or storm) event, the likelihood of a county reporting ≥1 case and being in a high incidence group was significantly higher for counties in years when an extreme weather event occurred, regardless of event type (Fig 2–4; Table 1–2 & S5–S6). For example, risk of reporting ≥1 case and of being in a high incidence group was 2.98 (95% CI=2.4–3.6) and 2.05 (95% CI=1.8–2.4) times greater, respectively, if a hurricane occurred.

Socioeconomic.—Risk that a county reported ≥1 case was significantly higher for counties with higher high school graduation and unemployment rates, per capita income, and income inequality (Fig 3, S4–S5; Table S5). Risk was lower for counties where more residents worked in agriculture, forestry, fishing, hunting, or mining. Risk was 1.34 (95% CI=1.04–1.72) times higher for counties where the largest employer was food manufacturing and processing (Table 1).

The likelihood of a county being in a high incidence group was negatively associated with the percent of residents working in agriculture, forestry, fishing, hunting or mining but positively with high school graduation and unemployment rates, per capita income, the percentage of households without adequate kitchen facilities, and income inequality (Fig 4, S2, S4–S5; Table S6). Risk was higher where the largest employer was food manufacturing and processing (Table 2).

Healthcare Environment: Access.—Risk that a county reported ≥1 case was significantly lower for counties with more diagnostic laboratories, hospitals, and social services (Fig 3, S6–S7; Table 1& S4). Counties with more hospitals, social services, diagnostic labs, and physician offices generally had a lower risk of being in a high-incidence group (Fig 4, S6–S7; Table S6). Medically underserved counties, as defined by AHRQ (see Table S1), were less likely to report ≥1 case (0.91; 95% CI=0.88–0.94) but were, if they reported ≥1 case, more likely to be in a high incidence group (Tables 1–2).

Healthcare Environment: Affordability.—Risk of a county reporting ≥1 case was significantly higher for counties with higher healthcare costs per capita and where more residents were unable to seek medical care due to the cost (Fig 3, S6–S7; Table S5). Risk of a county reporting ≥1 case was lower for counties where a higher percentage of adult residents lacked health insurance. The likelihood of a county being in the highest incidence category increased as per capita healthcare costs and the percentage of residents who were unable to seek care increased (Fig 4, S6–S7; Table S6).

Healthcare Environment: Quality of Healthcare.—Risk that a county reported ≥1 case was significantly higher for counties where more residents reported being obese, and with higher flu vaccination and diabetes monitoring rates (Fig 3, S8; Table S5). The likelihood of a county being in a high-incidence group generally increased as the obesity, flu vaccination, and diabetes monitoring rates increased (Fig 4, S8; Table S6).

Food Environment.—Counties with more fast-food restaurants (FFR) and negative food resources (NFR) had a significantly higher risk of reporting at least one case (Fig 3, S9). Conversely, risk was notably lower for counties with more convenience stores, full-service restaurants (FSR), and grocery stores, and where more residents lived in food deserts (Fig 3, S5, S9–S10). For instance, risk increased steadily from 0.9 (95% CI=0.8–0.9) when there were 0.1–0.7 NFR/1,000 persons to 1.4 (95% CI=1.3–1.5) when there were 1.2–1.3 NFR/1,000 persons; risk remained high for counties with more than 1.2 NFR/1,000 persons (Table S5).

Risk that a county reported a high salmonellosis incidence was positively associated with convenience stores, FFR, FSR, and NFR/1,000 persons and per capita FFR expenditures (Fig 4, S9–S10). Risk was negatively associated with grocery stores/1,000 persons increased (Fig S10; Table S6).

Demographics.—Counties where a higher proportion of residents identified as Black or African American (Black), Hispanic or Latino (Hispanic), or Native Hawaiian or other Pacific Islander (NHPI) had a significantly greater risk of reporting ≥1 case (Fig 3, S11; Table 1, S4–S5). Risk of reporting ≥1 case increased steadily from 0.6 (95% CI=0.5–0.6) when 0.0%–0.5% of residents identified as Hispanic to 1.3 (95% CI=1.3–1.4) when 4.0%–8.0% of residents identified as Hispanic. Counties including American Indian or Alaska Native (AIAN) tribal lands or NHPI homelands had a 1.44 (95% CI=1.32–1.57) times higher risk than those that did not. Risk that ≥1 case was reported increased consistently with population density and was marginally higher for more urban counties (Fig S1; Tables 1, S4).

Risk of a county reporting a high incidence increased substantially as the percentage of residents who identified as Black or Hispanic increased (Fig 4, S11; Table 2, S5). Risk was significantly lower for counties where more residents identified as AIAN. Risk was significantly higher for less urban counties, and counties that included AIAN tribal lands or NHPI homelands.

Discussion:

We address key gaps in our current understanding of the determinants of salmonellosis in the United States and identify characteristics of communities that may be differentially disadvantaged toward a higher incidence. This analysis used CFRF to characterize associations between salmonellosis incidence and individual DOH while controlling for all other DOH in the CFRF and all possible interactions between DOH. CFRF results can be interpreted as the change in the risk that a county moved from a low to high incidence group if a single DOH, such as unemployment rate, changed while controlling for all other DOH in the model.

Temporal trends may be due to changes in detection methods.

The risk ratios associated with year were among the largest observed. Due to improvements in LEDS reporting and advancements in illness diagnostics and outbreak detection over the last two decades, the strong association between incidence and year is not unexpected (Healy et al., 2023; Li et al., 2020). Increased use of culture-independent syndromic panel tests (CIDTs) since 2015 has been repeatedly linked to increases in reported enteric illness incidence (Collins et al., 2022; Delahoy et al., 2023; Healy et al., 2023; Ray et al., 2024;

Ray et al., 2022). Despite substantial increases in the incidence of illnesses reported to the CDC, (Collins et al., 2022; Delahoy et al., 2023), Healy et al. (2023) determined that, in the absence of CIDTs, observed salmonellosis incidence would have decreased. Preliminary 2023 data from the FoodNet catchment, suggest a 97% increase in the percent of *Salmonella* cases diagnosed by CIDT between 2023 and 2016–2018, the baseline FoodNet uses for monitoring trends in enteric illnesses (Centers for Disease Control and Prevention, 2023).

Environmental context appears to predispose some counties toward higher salmonellosis incidence.—More agricultural counties, counties with more wetland cover, and counties with more impaired surface waterways were differentially disadvantaged toward higher salmonellosis incidence. Multiple studies have found similar positive associations between bacterial disease incidence and water quality violations, wetland cover, and agricultural activity/intensity (Henly et al., 2017; Holcomb et al., 2022; Huang et al., 2017; Shaw et al., 2016). Such findings are also consistent with our understanding of *Salmonella* ecology and epidemiology — *Salmonella* can survive and persist in aquatic and wetland environments, and livestock can serve as non-human hosts for *Salmonella* (Hanson et al., 2016; Mentaberre et al., 2013; Weller, Belias, et al., 2020; Weller, Brassill, et al., 2020). These environments and the animals in them may also serve as reservoirs and/or vectors for *Salmonella* and contribute to *Salmonella* transmission to humans (Gorski et al., 2013; Huang et al., 2017; Marin et al., 2013; Mermin et al., 2004). Multiple salmonellosis outbreaks have been linked back to contact with surface water, wild and domestic animals, and animal products [e.g., outbreaks among food workers; (Hedican et al., 2010; Marus et al., 2019; Mermin et al., 2004)]. We also found that extreme weather, regardless of extreme weather event type, predisposed a county toward a greater salmonellosis incidence. After accounting for a county's demographic, socioeconomic, and physical (e.g., average weather, land use) environment, the occurrence of any extreme weather event was significantly associated with an increased risk of said county reporting ≥1 case and reporting a high incidence. These findings are well-supported by studies conducted in the United States and other countries on the impact of extreme weather on enteric illness incidence [e.g., (Jiang et al., 2015; Lal et al., 2013; Lee et al., 2019; Levy et al., 2016; Morgado et al., 2021; Quist et al., 2022)]. Jiang et al. (2015) observed a 4.1% and a 5.6% increase in salmonellosis risk for each 1 unit increase in extreme temperature and precipitation, respectively, and determined that the impact was disproportionately greater in coastal compared to non-coastal Maryland, USA communities. Lee et al. (2019) found that extreme rain events were associated with a 5% increase in salmonellosis risk in Georgia, USA; however, the effect of these events was mediated by rainfall patterns antecedent to the extreme event and was substantially higher for illnesses caused by *Salmonella* serotypes commonly isolated from wildlife and environmental sources. Overall, our findings suggest that environmental context predisposes a county toward a higher salmonellosis incidence, regardless of other DOH characteristics, and that extreme weather event, regardless of event type, was associated with increased salmonellosis incidence. Because climate change will differentially impact marginalized and under-resourced communities (Ruth & Ibarrarán, 2009; Thomas et al., 2019), additional research is needed to determine how increases in extreme weather will impact communities that are already at greater enteric illness risk due to environmental or socioeconomic context and what prevention measures may reduce the differential impact of extreme weather on disease burden in these communities.

Economically marginalized and under-resourced communities appear to have a higher salmonellosis incidence.—Medically underserved counties, and counties with higher per capita healthcare costs, greater income inequality, higher unemployment rates, more homes without adequate sanitation/hygiene facilities, and more persons who could not afford medical care were more likely to report a higher incidence as were counties with higher healthcare quality metrics, per capita income, and graduation rates. Past studies noted that more affluent persons were more likely to engage in behaviors associated with potential *Salmonella* exposure, such as international travel, owning a pet, and eating outside the home (Varga et al., 2013; Younus et al., 2007), and a study on crowd-sourced foodborne illness reporting concluded that indicators of affluence were positively associated with higher reporting rates while indicators of lower socioeconomic status were associated with reduced reporting (Henly et al., 2017). The authors attributed this to residents of affluent communities being more likely and better able to report an illness rather than lower rates of illness in less affluent communities (Henly et al., 2017). This conclusion is supported by studies conducted in countries with universal healthcare where lower socioeconomic status was associated with increased salmonellosis risk (Borgnolo et al., 1996; Varga et al., 2013); universal healthcare may reduce underreporting so findings from these studies, thus, better reflect the true impact of socioeconomic status on salmonellosis risk. Our analysis found that medically underserved counties and counties with reduced access to healthcare were significantly less likely to report any salmonellosis cases, which may be due to a truly lower disease incidence in these counties or, more plausibly, to reduced rates of care-seeking. Overall, it is likely that the association between affluence indicators and salmonellosis is an artifact of improved reporting and patient ascertainment, while the association with indicators of lower socioeconomic status represents a true disparity. There is a need for additional research on enteric disease disparities using approaches that can delineate the relative impact of socioeconomic factors on under-ascertainment/underreporting and true disease burden. One way to do this may be to focus on disparities in severe illness incidence (e.g., hospitalization, death) where under-ascertainment/underreporting is unlikely (Rose et al., 2020).

Communities with fewer positive and more negative food resources were differentially disadvantaged toward higher salmonellosis incidence.—Counties with more negative food resources, FFR, FFR sales, and fewer grocery stores per capita were more likely to both report ≥1 case and be in a high incidence group if they reported ≥1 case. Communities with limited access to affordable, healthy food (i.e., positive food resources) are considered by the USDA to be food deserts. A recent study found a substantial disparity in the quality and safety of food available to Houston, TX neighborhoods classified as low-income food deserts compared to high-income neighborhoods (Sirsat et al., 2021). Similar conclusions were reached by two Philadelphia studies which noted a higher incidence of unsafe food storage and handling practices in low-income, food deserts (Koro et al., 2010; Signs et al., 2011). These and other studies (Brown et al., 2016; Koro et al., 2010; Olive, 2020; Signs et al., 2011; Weisbecker, 2010) highlighted that non-traditional (e.g., FFR, convenience stores, dollar stores) and independently-owned food sources were more common in low-income food deserts, and often experienced more and different barriers to food safety than conventional food sources (e.g., limited resources, poor infrastructure, small/untrained staff, nontraditional supply). Individuals living in rural food deserts may travel substantial distances to purchase food, increasing the potential for temperature abuse between store and home (Griffing et al., 2018). In this context, our findings suggest that counties with reduced access to positive and increased reliance on negative resources are disproportionately disadvantaged toward higher salmonellosis incidence.

This study also found that restaurants per capita and related metrics were associated with higher reported salmonellosis incidence. Similarly, a study that examined online business reviews by consumers found that one of the strongest predictors of reported county-level foodborne illness incidence was restaurants per capita (Henly et al., 2017). These findings are unsurprising given the number of outbreaks linked to restaurants [e.g., (Gould et al., 2013; Hedberg et al., 1991; Nettleton et al., 2021)]. Sixty-six percent (300/457) of outbreaks reported by FoodNet sites during 2006–2007 were restaurant-associated (Gould et al., 2013). Therefore, our findings suggest that food preparer outreach and interventions (e.g., trainings) tailored to and developed in collaboration with the restaurant industry and workers may be appropriate.

Demographic characteristics were strongly associated with salmonellosis incidence, suggesting a disparity for counties with different demographic profiles.—Multiple studies identified disparities in disease burden between communities and individuals with different demographic profiles [e.g., (Chang et al., 2009; Shaw et al., 2016)]. The present study found that counties where a higher proportion of residents identified as Black, Hispanic or NHPI, or included AIAN tribal lands or an NHPI homeland were more likely to report ≥1 salmonellosis cases and a high incidence if they reported ≥1 case. However, counties with more residents who identified as AIAN were more likely to report a low incidence. This was true whether we included AIAN tribal land or NHPI homeland presence in the model or not, suggesting that different processes are driving salmonellosis transmission, care seeking, and/or disease reporting for AIAN populations living on versus off tribal lands. While only 13% of AIAN persons live on tribal lands, a significantly higher percentage of AIAN adults living on tribal lands reported having a usual place where they received healthcare compared to AIAN adults living off tribal lands(Ng et al., 2023). AIAN persons living on tribal lands were also significantly more likely to receive preventative care than AIAN person living off tribal lands, while AIAN persons living outside tribal lands were more likely to report not seeking care due to the cost (Ng et al., 2023). This may suggest that AIAN persons living on tribal lands have better access to healthcare or that under-ascertainment and underreporting are higher for counties with a large AIAN population living outside tribal lands. This knowledge gap highlights the need for research and public health prevention efforts tailored to the needs of specific communities. These findings further emphasize the need for enteric disease disparity studies that use approaches that can differentiate the relative impact of socioeconomic factors, such as access to healthcare, on under-ascertainment/underreporting and true (as opposed to reported) disease burden.

The demographic disparities identified here are not attributable to biological factors and are most likely caused by structural inequities too complex to be captured statistically and difficult to target with prevention efforts. The repeated identification of race or ethnicity-based disparities for foodborne and enteric diseases [e.g., (Arshad et al., 2007; Chang et al., 2009; Gourishankar, 2021; Quinlan, 2013; Shaw et al., 2016)] raises the question of which structural inequities are driving these disparities and how public health professionals can best develop interventions to address them. Follow-on studies at finer spatial scales are needed to move beyond discussing race and ethnicity-based disparities, and to identify, understand, and address inequities. Despite this need, our findings suggest that culturally competent outreach and intervention efforts tailored to specific communities are needed for salmonellosis prevention.

Limitations:

This analysis was conducted at the county level, and US counties are heterogeneous. Follow-on studies conducted at finer spatial scales that better reflect the lived experiences of ill persons are therefore needed to substantiate our findings.

To ensure sufficient confounding control, a variety of DOH were included in our model. While this is a strength of our approach, it complicates interpretation due to correlation between retained and excluded DOH (e.g., between the percentage of residents that identify as Black or African American [Black] and poverty rate) and inclusion of conceptually linked variables (e.g., presence of an AIAN or NHPI homeland, and percentage of residents that identify as AIAN or as NHPI). Also, certain variables function as proxies for broader phenomena (e.g., temporary shelters/1,000 persons functions as a proxy for social service availability).

Like all passive surveillance systems, LEDS relies on voluntary illness reporting, and may lead to selection bias, underreporting, and under ascertainment (e.g., ill persons not seeking care, under-reporting caused by failure to diagnose, differences in reporting practices between jurisdictions) (Losos, 1996; Scallan et al., 2011). Our findings provide insights into disparities among reported cases, but may not reflect disparities for populations for whom salmonellosis is under-detected. Follow-on studies that focus on outcomes robust to underreporting and under ascertainment (e.g., severe illness, death) are needed. Despite these limitations, our analysis provides a foundation on which future analyses and public health efforts and interventions to address disparities in salmonellosis incidence can build.

Conclusion:

This study is novel in its use of counterfactual random forest (CFRF), which is a pseudo-causal method that allows us to characterize associations between salmonellosis incidence and individual DOH while controlling for all other features included in the model as well as all possible interactions between these DOHs. This analysis provides a conceptual framework for how CFRF can be applied to quantify health disparities in infectious diseases with strong confounding control.

By identifying disparities in salmonellosis incidence for counties with different DOH profiles, this analysis suggests that known county characteristics and unmeasured structural inequities predispose certain counties toward a high salmonellosis incidence. Communities with specific environmental (e.g., extreme weather), socioeconomic (e.g., under-resourced), or demographic characteristics were disproportionately disadvantaged toward an increased incidence even after controlling for all other DOHs in the model. This analysis provides key insights into salmonellosis epidemiology, including the importance of understanding how environmental and social disparities interact, and how extreme weather may magnify existing disparities. Such understanding is critical for developing equitable, effective, and culturally-competent intervention and prevention efforts; such efforts need to be co-created with communities and proactively tailored to their needs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We would like to thank Sara Luna (deceased) and Eleanor Click for their support during ideation, and Sean Browning and Bruce Gutelius for their assistance with surveillance data preparation. We would like to thank Molly Steele and Renee Calanan for their technical expertise during manuscript development.

Data Availability Statement:

The surveillance data used here are available upon request to phlissupport@cdc.gov. All other data are publicly available and can be downloaded from the portals listed in Table S1.

References

Arshad M, Wilkins M, Downes F, Rahbar M, Erskine R, Boulton M, & Saeed A (2007). A Registry-Based Study on The Association between Human Salmonellosis And Routinely Collected Parameters in Michigan, 1995–2001. Foodborne Pathogens and Disease, 4, 16–25. 10.1089/fpd.2006.48 [PubMed: 17378704]

Borgnolo G, Barbone F, Scornavacca G, Franco D, Vinci A, & Iuculano F (1996). A case-control study of Salmonella gastrointestinal infection in Italian children. Acta Paediatr, 85(7), 804–808. 10.1111/j.1651-2227.1996.tb14155.x [PubMed: 8819545]

Brown W, Ryser E, Gorman L, Steinmaus S, & Vorst K (2016). Temperatures experienced by fresh-cut leafy greens during retail storage and display.

Centers for Disease Control and Prevention. (2023). FoodNet Fast: Pathogen Surveillance Tool. https://www.cdc.gov/foodnet/foodnet-fast.html

Chang M, Groseclose SL, Zaidi AA, & Braden CR (2009). An ecological analysis of sociodemographic factors associated with the incidence of salmonellosis, shigellosis, and E. coli O157:H7 infections in US counties. Epidemiol Infect, 137(6), 810–820. 10.1017/S0950268808001477 [PubMed: 18947443]

Collins J, Shah HJ, Weller DL, Ray LC, Smith K, McGuire S, Trevejo RT, Jervis RH, Vugia DJ, Rissman T, Garman KN, Lathrop S, LaClair B, Boyle MM, Harris S, Kufel JZ, Tauxe RV, Bruce BB, Rose EB, … Payne DC (2022). Preliminary incidence and trends of infections caused by pathogens transmitted commonly through food - Foodborne Diseases Active Surveillance Network, 10 U.s. sites, 2016–2021. MMWR Morb. Mortal. Wkly. Rep, 71(40), 1260–1264. 10.15585/mmwr.mm7140a2 [PubMed: 36201372]

Cummings PL, Kuo T, Javanbakht M, Shafir S, Wang M, & Sorvillo F (2015). Salmonellosis Hospitalizations in the United States: Associated Chronic Conditions, Costs, and Hospital Outcomes, 2011, Trends 2000–2011. Foodborne Pathogens and Disease, 13(1), 40–48. 10.1089/fpd.2015.1969 [PubMed: 26545047]

Dasgupta A, Szymczak S, Moore JH, Bailey-Wilson JE, & Malley JD (2014). Risk estimation using probability machines. BioData Mining, 7(1), 2. 10.1186/1756-0381-7-2 [PubMed: 24581306]

Delahoy MJ, Shah HJ, Weller DL, Ray LC, Smith K, McGuire S, Trevejo RT, Scallan Walter E, Wymore K, Rissman T, McMillian M, Lathrop S, LaClair B, Boyle MM, Harris S, Zablotsky-Kufel J, Houck K, Devine CJ, Lau CE, … Payne, D. C. (2023). Preliminary Incidence and Trends of Infections Caused by Pathogens Transmitted Commonly Through Food - Foodborne Diseases Active Surveillance Network, 10 U.S. Sites, 2022. MMWR Morb Mortal Wkly Rep, 72(26), 701–706. 10.15585/mmwr.mm7226a1 [PubMed: 37384552]

Dhaliwal S, Hoffmann S, White A, Ahn J-W, McQueen RB, & Scallan Walter E (2021). Cost of Hospitalizations for Leading Foodborne Pathogens in the United States: Identification by International Classification of Disease Coding and Variation by Pathogen. Foodborne Pathogens and Disease, 18(11), 812–821. 10.1089/fpd.2021.0028 [PubMed: 34591654]

Gorski L, Jay-Russell MT, Liang AS, Walker S, Bengson Y, Govoni J, & Mandrell RE (2013). Diversity of Pulsed-Field Gel Electrophoresis Pulsotypes, Serovars, and Antibiotic Resistance Among Salmonella Isolates from Wild Amphibians and Reptiles in the California Central Coast. Foodborne Pathogens and Disease, 10(6), 540–548. 10.1089/fpd.2012.1372 [PubMed: 23577627]

Gould LH, Rosenblum I, Nicholas D, Phan Q, & Jones TF (2013). Contributing Factors in Restaurant-Associated Foodborne Disease Outbreaks, FoodNet Sites, 2006 and 2007†. Journal of Food Protection, 76(11), 1824–1828. 10.4315/0362-028X.JFP-13-037 [PubMed: 24215683]

Gourishankar A (2021). Geospatial analysis of salmonellosis and its association with socioeconomic status in Texas. Fam Med Community Health, 9(4). 10.1136/fmch-2021-001214

Griffing DA, Kriese-Anderson LA, Mullenix M, Wang L, & Bratcher CL (2018). A Study of Consumer Handling Behaviors during Transport from Retail to Residence Utilizing an Electronic Questionnaire. Meat and Muscle Biology, 2(1). 10.22175/mmb2017.12.0058

Hadler JL, Clogher P, Libby T, Wilson E, Oosmanally N, Ryan P, Magnuson L, Lathrop S, Mcguire S, Cieslak P, Fankhauser M, Ray L, Geissler A, & Hurd S (2019). Relationship Between Census Tract–Level Poverty and Domestically Acquired Salmonella Incidence: Analysis of Foodborne Diseases Active Surveillance Network Data, 2010–2016. The Journal of Infectious Diseases, 222(8), 1405–1412. 10.1093/infdis/jiz605

Hanson DL, Loneragan GH, Brown TR, Nisbet DJ, Hume ME, & Edrington TS (2016). Evidence supporting vertical transmission of Salmonella in dairy cattle. Epidemiol Infect, 144(5), 962–967. 10.1017/s0950268815002241 [PubMed: 26419321]

Harrell FEJ, & Dupont C (2024). Hmisc: Harrell Miscellaneous. https://cran.r-project.org/web/packages/Hmisc/index.html

Havelaar AH, Haagsma JA, Mangen M-JJ, Kemmeren JM, Verhoef LPB, Vijgen SMC, Wilson M, Friesema IHM, Kortbeek LM, van Duynhoven YTHP, & van Pelt W (2012). Disease burden of foodborne pathogens in the Netherlands, 2009. International Journal of Food Microbiology, 156(3), 231–238. 10.1016/j.ijfoodmicro.2012.03.029 [PubMed: 22541392]

Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, de Silva NR, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleesschauwer B, & on behalf of World Health Organization Foodborne Disease Burden Epidemiology Reference, G. (2015). World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLOS Medicine, 12(12), e1001923. 10.1371/journal.pmed.1001923 [PubMed: 26633896]

Healy JM, Ray L, Tack DM, Eikmeier D, Tobin-D'Angelo M, Wilson E, Hurd S, Lathrop S, McGuire SM, & Bruce BB (2023). Modelling counterfactual incidence during the transition towards culture-independent diagnostic testing. Int J Epidemiol. 10.1093/ije/dyad133

Hedberg CW, White KE, Johnson JA, Edmonson LM, Soler JT, Korlath JA, Theurer LS, MacDonald KL, & Osterholm MT (1991). An Outbreak of Salmonella enteritidis Infection at a Fast-Food

Restaurant: Implications for Foodhandler-Associated Transmission. The Journal of Infectious Diseases, 164(6), 1135–1140. 10.1093/infdis/164.6.1135 [PubMed: 1955712]

Hedican E, Miller B, Ziemer B, LeMaster P, Jawahir S, Leano F, & Smith K (2010). Salmonellosis Outbreak Due to Chicken Contact Leading to a Foodborne Outbreak Associated with Infected Delicatessen Workers. Foodborne Pathogens and Disease, 7(8), 995–997. 10.1089/fpd.2009.0495 [PubMed: 20470192]

Henly S, Tuli G, Kluberg SA, Hawkins JB, Nguyen QC, Anema A, Maharana A, Brownstein JS, & Nsoesie EO (2017). Disparities in digital reporting of illness: A demographic and socioeconomic assessment. Preventive Medicine, 101, 18–22. 10.1016/j.ypmed.2017.05.009 [PubMed: 28528170]

Hoffmann S (2014). Cost Estimates of Foodborne Illnesses. United States Food and Drug Administration. https://www.ers.usda.gov/data-products/cost-estimates-of-foodborne-illnesses.aspx

Hoffmann S, Batz MB, & Morris JG (2012). Annual Cost of Illness and Quality-Adjusted Life Year Losses in the United States Due to 14 Foodborne Pathogens††The views expressed herein are those of the authors and do not necessarily reflect the views of the Economic Research Service or the U.S. Department of Agriculture. Journal of Food Protection, 75(7), 1292–1302. 10.4315/0362-028X.JFP-11-417 [PubMed: 22980013]

Holcomb DA, Quist AJL, & Engel LS (2022). Exposure to industrial hog and poultry operations and urinary tract infections in North Carolina, USA. Science of The Total Environment, 853, 158749. 10.1016/j.scitotenv.2022.158749 [PubMed: 36108846]

Huang JY, Patrick ME, Manners J, Sapkota AR, Scherzinger KJ, Tobin-D'Angelo M, Henao OL, Cole DJ, & Vieira AR (2017). Association between wetland presence and incidence of Salmonella enterica serotype Javiana infections in selected US sites, 2005–2011. Epidemiol Infect, 145(14), 2991–2997. 10.1017/S0950268817001790 [PubMed: 28803563]

Ishwaran H, & Kogalur UB (2023). randomForestSRC: Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). https://cran.r-project.org/web/packages/randomForestSRC/index.html

Jackson BR, Gold JAW, Natarajan P, Rossow J, Neblett Fanfair R, da Silva J, Wong KK, Browning SD, Bamrah Morris S, Rogers-Brown J, Hernandez-Romieu AC, Szablewski CM, Oosmanally N, Tobin-D'Angelo M, Drenzek C, Murphy DJ, Hollberg J, Blum JM, Jansen R, … Bruce BB (2021). Predictors at Admission of Mechanical Ventilation and Death in an Observational Cohort of Adults Hospitalized With Coronavirus Disease 2019. Clin Infect Dis, 73(11), e4141–e4151. 10.1093/cid/ciaa1459 [PubMed: 32971532]

Jiang C, Shaw KS, Upperman CR, Blythe D, Mitchell C, Murtugudde R, Sapkota AR, & Sapkota A (2015). Climate change, extreme events and increased risk of salmonellosis in Maryland, USA: Evidence for coastal vulnerability. Environment International, 83, 58–62. 10.1016/j.envint.2015.06.006 [PubMed: 26093493]

Knowles T, Moody R, & McEachern MG (2007). European food scares and their impact on EU food policy. British Food Journal, 109(1), 43–67. 10.1108/00070700710718507

Koro ME, Anandan S, & Quinlan JJ (2010). Microbial Quality of Food Available to Populations of Differing Socioeconomic Status. American Journal of Preventive Medicine, 38(5), 478–481. 10.1016/j.amepre.2010.01.017 [PubMed: 20409496]

Kuhn M, & Johnson K (2016). Applied Predictive Modeling. Springer Nature.

Lake RJ, Cressey PJ, Campbell DM, & Oakley E (2010). Risk Ranking for Foodborne Microbial Hazards in New Zealand: Burden of Disease Estimates. Risk Analysis, 30(5), 743–752. 10.1111/j.1539-6924.2009.01269.x [PubMed: 19645753]

Lal A, Ikeda T, French N, Baker MG, & Hales S (2013). Climate Variability, Weather and Enteric Disease Incidence in New Zealand: Time Series Analysis. PLoS One, 8(12), e83484. 10.1371/journal.pone.0083484 [PubMed: 24376707]

Lee D, Chang HH, Sarnat SE, & Levy K (2019). Precipitation and Salmonellosis Incidence in Georgia, USA: Interactions between Extreme Rainfall Events and Antecedent Rainfall Conditions. Environ Health Perspect, 127(9), 97005. 10.1289/EHP4621 [PubMed: 31536392]

Levy K, Woster AP, Goldstein RS, & Carlton EJ (2016). Untangling the Impacts of Climate Change on Waterborne Diseases: a Systematic Review of Relationships between Diarrheal Diseases and

Temperature, Rainfall, Flooding, and Drought. Environmental Science & Technology, 50(10), 4905–4922. 10.1021/acs.est.5b06186 [PubMed: 27058059]

Li X, Singh N, Beshearse E, Blanton JL, DeMent J, & Havelaar AH (2020). Spatial Epidemiology of Salmonellosis in Florida, 2009–2018. Front Public Health, 8, 603005. 10.3389/fpubh.2020.603005 [PubMed: 33681114]

Losos JZ (1996). Routine and sentinel surveillance methods. EMHJ - Eastern Mediterranean Health Journal, 2(1), 46–50.

Malley JD, Kruppa J, Dasgupta A, Malley KG, & Ziegler A (2012). Probability Machines. Consistent Probability Estimation Using Nonparametric Learning Machines, 51(01), 74–81. 10.3414/ME00-01-0052

Marin C, Ingresa-Capaccioni S, González-Bodi S, Marco-Jiménez F, & Vega S (2013). Free-Living Turtles Are a Reservoir for Salmonella but Not for Campylobacter. PLoS One, 8(8), e72350. 10.1371/journal.pone.0072350 [PubMed: 23951312]

Marus JR, Magee MJ, Manikonda K, & Nichols MC (2019). Outbreaks of Salmonella enterica infections linked to animal contact: Demographic and outbreak characteristics and comparison to foodborne outbreaks—United States, 2009–2014. Zoonoses and Public Health, 66(4), 370–376. 10.1111/zph.12569 [PubMed: 30821071]

Mentaberre G, Porrero MC, Navarro-Gonzalez N, Serrano E, Domínguez L, & Lavín S (2013). Cattle Drive Salmonella Infection in the Wildlife–Livestock Interface. Zoonoses and Public Health, 60(7), 510–518. 10.1111/zph.12028 [PubMed: 23253262]

Mermin J, Hutwagner L, Vugia D, Shallow S, Daily P, Bender J, Koehler J, Marcus R, Angulo FJ, & for the Emerging Infections Program FoodNet Working, G. (2004). Reptiles, Amphibians, and Human Salmonella Infection: A Population-Based, Case-Control Study. Clinical Infectious Diseases, 38(Supplement_3), S253–S261. 10.1086/381594 [PubMed: 15095197]

Monteiro Pires S, Jakobsen LS, Ellis-Iversen J, Pessoa J, & Ethelberg S (2019). Burden of Disease Estimates of Seven Pathogens Commonly Transmitted Through Foods in Denmark, 2017. Foodborne Pathogens and Disease, 17(5), 322–339. 10.1089/fpd.2019.2705 [PubMed: 31755845]

Morgado ME, Jiang C, Zambrana J, Upperman CR, Mitchell C, Boyle M, Sapkota AR, & Sapkota A (2021). Climate change, extreme events, and increased risk of salmonellosis: foodborne diseases active surveillance network (FoodNet), 2004–2014. Environmental Health, 20(1), 105. 10.1186/s12940-021-00787-y [PubMed: 34537076]

Nettleton WD, Reimink B, Arends KD, Potter D, Henderson JJ, Dietrich S, & Franks M (2021). Protracted, Intermittent Outbreak of Salmonella Mbandaka Linked to a Restaurant - Michigan, 2008–2019. MMWR Morb Mortal Wkly Rep, 70(33), 1109–1113. 10.15585/mmwr.mm7033a1 [PubMed: 34411074]

Ng AE, Adjaye-Gbewonyo D, & Vahratian A (2023). Health Conditions and Health Care Use Among American Indian and Alaska Native Adults by Tribal Land Residential Status : United States, 2019–2021 [Report]. https://stacks.cdc.gov/view/cdc/125982 (National Health Statistics Reports; no. 185)

Olive R (2020). Final report for GNC18-267 SARE Grant Management System (Farm to Fridge: Assessing Need and Availability of Underutilized Refrigeration in Rural Grocery Stores for Use by Fruit and Vegetable Farmers, Issue. https://projects.sare.org/project-reports/gnc18-267/

Park MS, Kim YS, Lee SH, Kim SH, Park KH, & Bahk GJ (2015). Estimating the Burden of Foodborne Disease, South Korea, 2008–2012. Foodborne Pathogens and Disease, 12(3), 207–213. 10.1089/fpd.2014.1858 [PubMed: 25622301]

Quinlan JJ (2013). Foodborne illness incidence rates and food safety risks for populations of low socioeconomic status and minority race/ethnicity: a review of the literature. Int J Environ Res Public Health, 10(8), 3634–3652. 10.3390/ijerph10083634 [PubMed: 23955239]

Quist AJL, Fliss MD, Wade TJ, Delamater PL, Richardson DB, & Engel LS (2022). Hurricane flooding and acute gastrointestinal illness in North Carolina. Science of The Total Environment, 809, 151108. 10.1016/j.scitotenv.2021.151108 [PubMed: 34688737]

Ray L, Weller D, Rounds J, Trevejo RT, Wilson E, Burzlaff K, Garman KN, Lathrop S, Rissman T, Wymore K, Wozny S, Wilson S, Watkins Francois L, Bruce BB, & Payne DC (2024). Syndromic gastrointestinal panel diagnostic tests have changed our understanding of the epidemiology of Yersinia enterocolitica — Foodborne Diseases Active Surveillance Network, 2010–2021.

Ray LC, Griffin PM, Wymore K, Wilson E, Hurd S, LaClair B, Wozny S, Eikmeier D, Nicholson C, Burzlaff K, Hatch J, Fankhauser M, Kubota K, Huang JY, Geissler A, Payne DC, & Tack DM (2022). Changing Diagnostic Testing Practices for Foodborne Pathogens, Foodborne Diseases Active Surveillance Network, 2012–2019. Open Forum Infectious Diseases, 9(8), ofac344. 10.1093/ofid/ofac344 [PubMed: 35928506]

Ribera LA, Palma MA, Paggi MS, Knutson RD, Masabni JG, & Anciso JR (2012). Economic Analysis of Food Safety Compliance Costs and Foodborne Illness Outbreaks in the United States. Horttechnology, 22, 150–156.

Rose TC, Adams NL, Whitehead M, Wickham S, O'Brien SJ, Hawker J, Taylor-Robinson DC, Violato M, & Barr B (2020). Neighbourhood unemployment and other socio-demographic predictors of emergency hospitalisation for infectious intestinal disease in England: A longitudinal ecological study. Journal of Infection, 81(5), 736–742. 10.1016/j.jinf.2020.08.048 [PubMed: 32888980]

Ruth M, & Ibarrarán ME (2009). Distributional Impacts of Climate Change and Disasters: Concepts and Cases.

Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, & Griffin PM (2011). Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis, 17(1), 7–15. 10.3201/eid1701.p11101 [PubMed: 21192848]

Scallan E, Hoekstra RM, Mahon BE, Jones TF, & Griffin PM (2015). An assessment of the human health impact of seven leading foodborne pathogens in the United States using disability adjusted life years. Epidemiology and Infection, 143(13), 2795–2804. 10.1017/S0950268814003185 [PubMed: 25633631]

Shah HJ, Jervis RH, Wymore K, Rissman T, LaClair B, Boyle MM, Smith K, Lathrop S, McGuire S, Trevejo R, McMillian M, Harris S, Zablotsky Kufel J, Houck K, Lau CE, Devine CJ, Boxrud D, & Weller DL (2024). Reported incidence of infections caused by pathogens transmitted commonly through food: Impact of increased use of culture-independent diagnostic tests - Foodborne Diseases Active Surveillance Network, 1996–2023. MMWR Morb. Mortal. Wkly. Rep, 73(26), 584–593. 10.15585/mmwr.mm7326a1 [PubMed: 38959172]

Shaw KS, Cruz-Cano R, Jiang C, Malayil L, Blythe D, Ryan P, & Sapkota AR (2016). Presence of animal feeding operations and community socioeconomic factors impact salmonellosis incidence rates: An ecological analysis using data from the Foodborne Diseases Active Surveillance Network (FoodNet), 2004–2010. Environ Res, 150, 166–172. 10.1016/j.envres.2016.05.049 [PubMed: 27290657]

Signs RJ, Darcey VL, Carney TA, Evans AA, & Quinlan JJ (2011). Retail Food Safety Risks for Populations of Different Races, Ethnicities, and Income Levels. Journal of Food Protection, 74(10), 1717–1723. 10.4315/0362-028X.JFP-11-059 [PubMed: 22004820]

Sirsat SA, Mohammad ZH, & Raschke I (2021). Safety and Quality of Romaine Lettuce Accessible to Low Socioeconomic Populations Living in Houston, TX. Journal of Food Protection, 84(12), 2123–2127. 10.4315/JFP-21-250 [PubMed: 34383915]

Thomas K, Hardy RD, Lazrus H, Mendez M, Orlove B, Rivera-Collazo I, Roberts JT, Rockman M, Warner BP, & Winthrop R (2019). Explaining differential vulnerability to climate change: A social science review. WIREs Climate Change, 10(2), e565. 10.1002/wcc.565

U.S. Department of Health and Human Services. (2023). Reduce infections caused by Salmonella — FS-04. https://health.gov/healthypeople/objectives-and-data/browse-objectives/foodborne-illness/reduce-infections-caused-salmonella-fs-04/data-methodology

Varga C, Pearl DL, McEwen SA, Sargeant JM, Pollari F, & Guerin MT (2013). Evaluating area-level spatial clustering of Salmonella Enteritidis infections and their socioeconomic determinants in the greater Toronto area, Ontario, Canada (2007 – 2009): a retrospective population-based ecological study. BMC Public Health, 13, 1078. 10.1186/1471-2458-13-1078 [PubMed: 24237666]

Wainer H (2019). The Most Dangerous Equation. https://www.americanscientist.org/article/the-most-dangerous-equation

Weisbecker A (2010). Few Healthy Food Choices in Urban Food Deserts. Food Safety News. https://www.foodsafetynews.com/2010/05/few-healthy-food-choices-in-urban-food-deserts/

Weller D, Belias A, Green H, Roof S, & Wiedmann M (2020). Landscape, Water Quality, and Weather Factors Associated With an Increased Likelihood of Foodborne Pathogen Contamination of New York Streams Used to Source Water for Produce Production. Front Sustain Food Syst, 3. 10.3389/fsufs.2019.00124

Weller D, Brassill N, Rock C, Ivanek R, Mudrak E, Roof S, Ganda E, & Wiedmann M (2020). Complex Interactions Between Weather, and Microbial and Physicochemical Water Quality Impact the Likelihood of Detecting Foodborne Pathogens in Agricultural Water. Front Microbiol, 11, 134. 10.3389/fmicb.2020.00134 [PubMed: 32117154]

Weller DL, Love TMT, & Wiedmann M (2021). Interpretability Versus Accuracy: A Comparison of Machine Learning Models Built Using Different Algorithms, Performance Measures, and Features to Predict E. coli Levels in Agricultural Water [Original Research]. Frontiers in Artificial Intelligence, 4. 10.3389/frai.2021.628441

Younus M, Hartwick E, Siddiqi AA, Wilkins M, Davies HD, Rahbar M, Funk J, & Saeed M (2007). The role of neighborhood level socioeconomic characteristics in Salmonella infections in Michigan (1997–2007): assessment using geographic information system. Int J Health Geogr, 6, 56. 10.1186/1476-072X-6-56 [PubMed: 18093323]

Highlights:

- Blueprint for use of counterfactual random forest to quantify health disparities
- Environment predisposed some counties toward higher salmonellosis incidence.
- Extreme weather events were associated with higher reported salmonellosis incidence
- Underreporting obfuscates associations between socioeconomic traits and incidence
- Demographic disparities likely reflect unmeasured structural inequities.

Scope Statement:

Advancements in disease reporting, detection, and statistical methods since the most recent U.S.-wide study on disparities in salmonellosis incidence was published necessitate an updated analysis. The updated analysis should use (i) a broader range of determinants of health (DOH), (ii) more recent data, and (iii) an analytical approach with strong confounding control. This study meets this need. Specifically, it uses counterfactual random forest analysis (CFRF) to identify potential disparities in salmonellosis incidence between counties with different DOH profiles, and characterize communities that are differentially disadvantaged toward a higher incidence. Because CFRF has not been used previously to characterize disparities in infectious disease incidence, this study provides a blueprint for how CFRF can be used to improve our understanding of health disparities and infectious disease epidemiology.

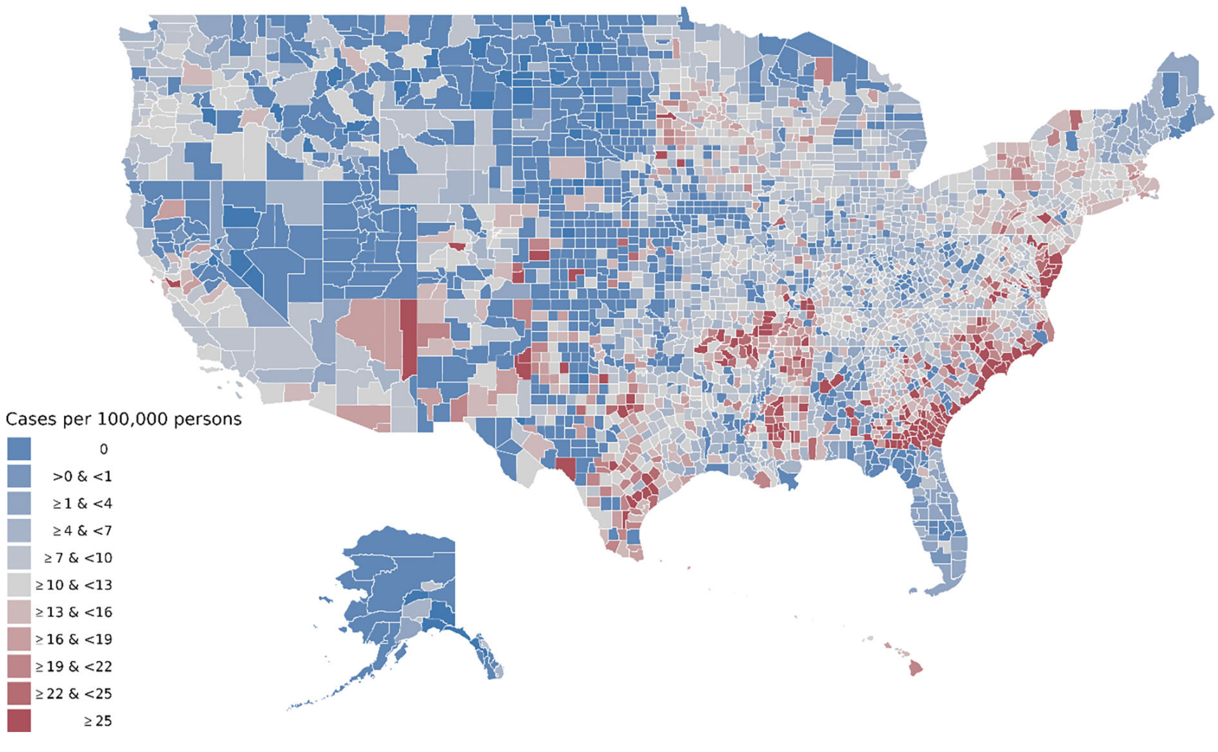

Figure 1:

Median annual, age-adjusted incidence of *Salmonella* cases reported to the Laboratory Enteric Diseases Surveillance System (LEDS) during 1997 through 2019.

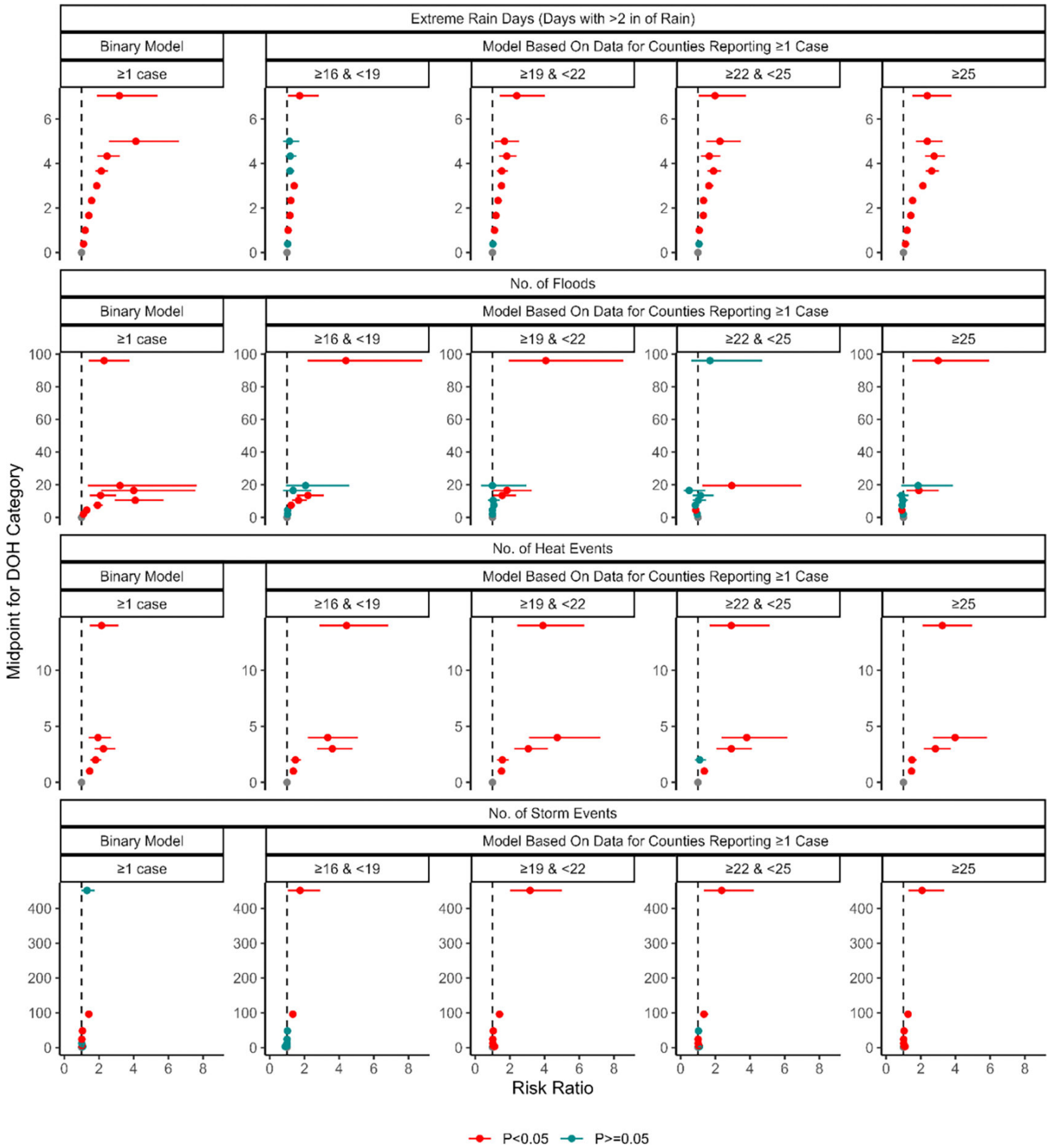

Figure 2:

Results from the hurdle CFRF for DOH related to extreme weather events. The first column shows results from the first part of the hurdle model and should be interpreted as the change in risk that a county reported ≥1 one case (versus no cases) if the county moved from the reference DOH group (in gray) to another group. The second to fifth columns show results from the second part of the hurdle model and should be interpreted as the change in risk that a county is in the given high-incidence groups (versus the reference) if the county moved from the reference DOH group (in gray) to another group. Significant associations are in red. Non-significant associations are in blue. The dot represents the risk ratio, and the bars represent its 95% confidence interval. Numerical data and data for other incidence groups can be found in Tables S2 and S3. Similar graphs for all the other DOH considered in the CFRF are in Supplemental Tables X-Y.

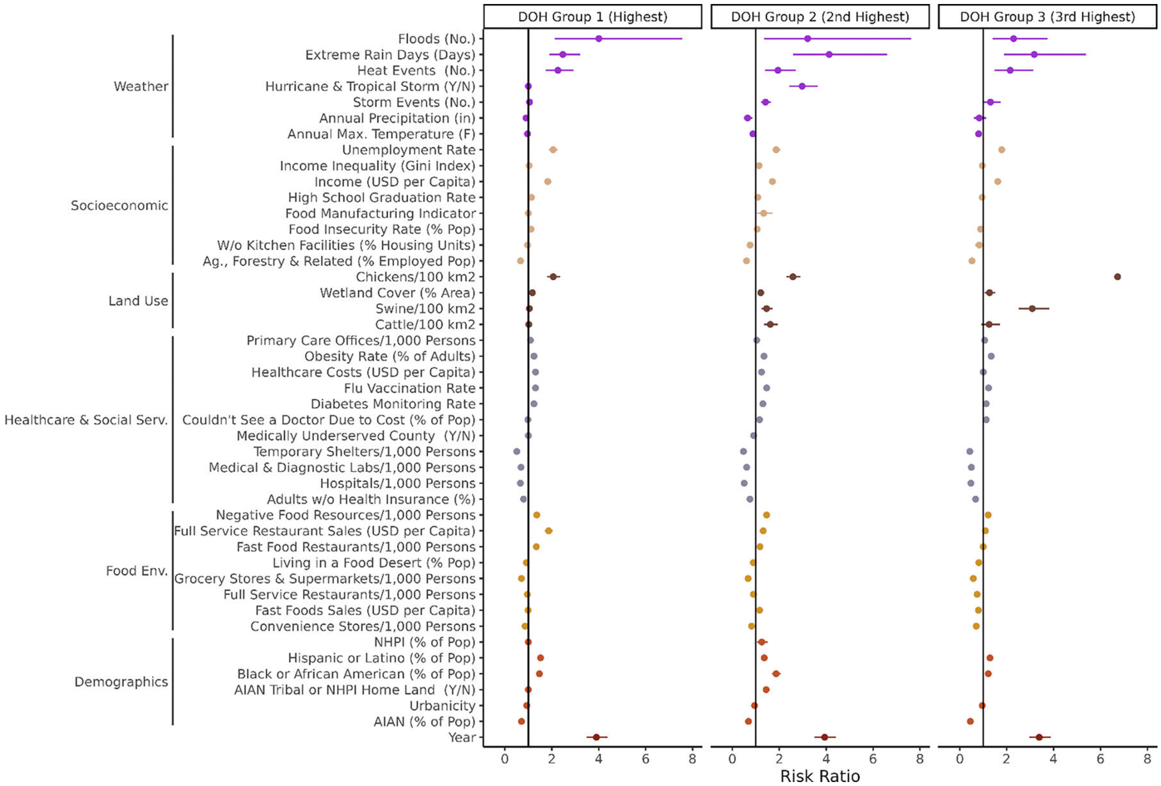

Figure 3:

Results from the first part of the hurdle CFRF. From left-to-right, the facets show risk of reporting ≥1 cases if a county moves from the reference DOH category to the highest (left), second highest (middle), or third highest (right) DOH categories. The dot represents the risk ratio, and the bars represent the 95% confidence interval. Numerical data and data for other DOH categories can be found in Table 2 and Figures S1–S13. Results are not shown for two DOH, population density and the percentage of surface waterways that were fecally-impaired, due to the magnitude of the risk ratio.

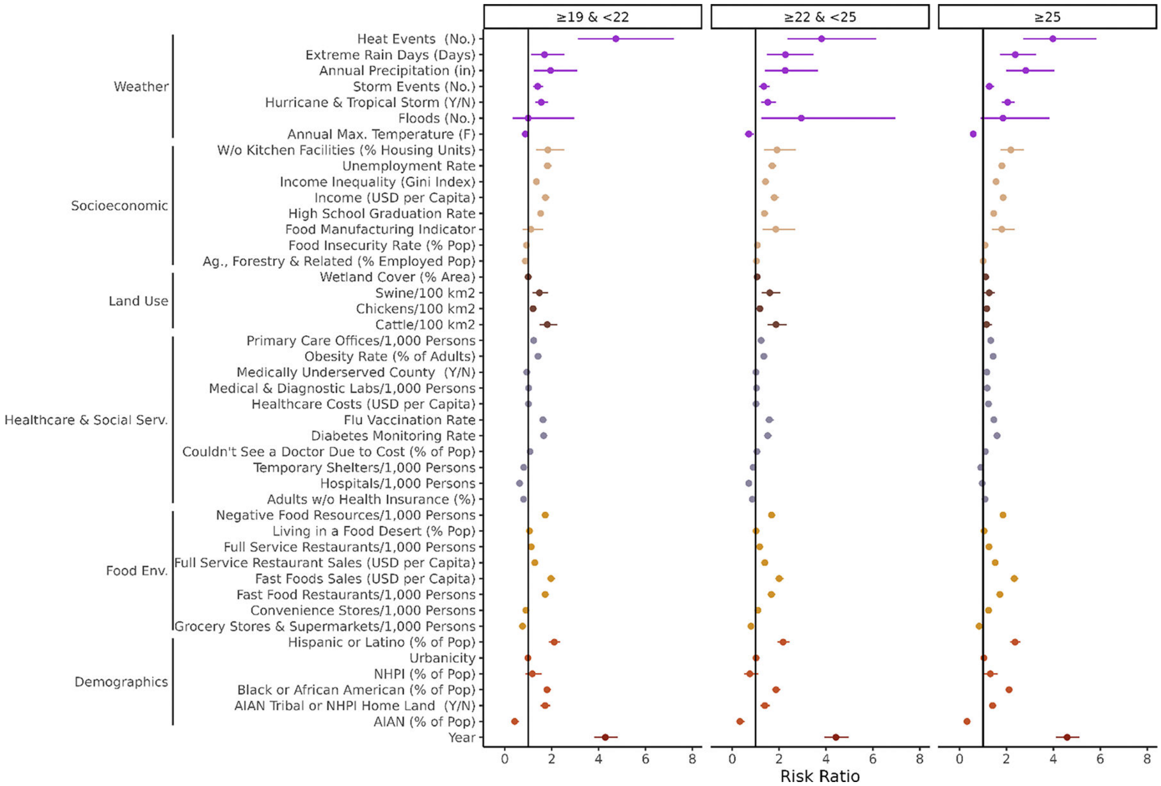

Figure 4:

Results from the second part of the hurdle CFRF. For counties reporting ≥1 case, risk of reporting a high salmonellosis incidence if a county moved from the reference DOH category to the second highest DOH category. From left-to-right, the facets show risk of a county reporting between 19 and 22 cases/100000 persons (left), 22 and 25 cases/100000 persons (middle), or ≥25 cases/100000 persons (right); the three highest incidence categories. The dot represents the risk ratio, and the bars represent the 95% confidence interval. Numerical data and data for other DOH categories can be found in Table 2 and Figures S1–S13. Results are not shown for two DOH, population density and the percentage of surface waterways that were fecally-impaired, due to the magnitude of the risk ratio.

Table 1:

Association between a county reporting ≥1 case for those determinants of health (DOH) that were inherently categorical (e.g., urbanicity) or were binary after being converted from a continuous to categorical (e.g., percent of residents that identified as Native Hawaiian or Other Pacific Islander (NHPI]).

DOH	RR [a]	95% CI [b]	*P*-value
Urbanicity (Ref = Large Central Metropolitan)			
Large Metropolitan Fringe (Suburban)	1.00	(0.96, 1.05)	0.966
Medium Metropolitan	0.96	(0.93, 1.00)	0.030
Small Metropolitan	0.93	(0.91, 0.96)	< 0.001
Micropolitan	0.95	(0.94, 0.96)	< 0.001
Non-core (Rural)	0.96	(0.96, 0.97)	< 0.001
Did a Hurricane or Tropical Storm Occur? (Ref = No)			
Yes	2.98	(2.43, 3.64)	< 0.001
Did the county include AIAN [c] tribal lands or an NHPI homeland? (Ref=No)			
Yes	1.44	(1.32, 1.57)	<0.001
Medically Underserved County (Ref = No)			
Yes	0.91	(0.88, 0.94)	< 0.001
Persons Identifying as NHPI (% of Pop; Ref = < 0.5%)			
≥0.5%	1.26	(1.05, 1.50)	0.013
Was the predominant employer in the county food manufacturing or processing? (Ref = No)			
Yes	1.34	(1.04, 1.72)	0.024

[a] Risk ratio

[b] Confidence Interval

[c] American Indian or Alaska Native

Table 2:

Association between a county being in one of the four higher incidence percentiles (≥16 cases/100,000 persons) versus the low incidence reference percentile (≥1.0 & <4.0 cases/100,000 persons) for select determinants of health (DOH) that were inherently categorical (e.g., was it a medically underserved county) or were binary after being converted from a continuous to categorical (e.g., percent of residents that identified as Native Hawaiian or Other Pacific Islander (NHPI]).

DOH	High Incidence Percentile (Cases/100,000 persons) [a]		RR [b]	95% CI [c]	P-value
Urbanicity (Ref = Large Central Metropolitan County)					
	≥16 & <19				
		Large Metropolitan Fringe (Suburban)	0.97	(0.94, 1.01)	0.094
		Medium Metropolitan	0.95	(0.92, 0.99)	0.013
		Small Metropolitan	0.98	(0.95, 1.02)	0.339
		Micropolitan	0.98	(0.96, 1)	0.105
		Non-core (Rural)	0.97	(0.96, 0.99)	< 0.001
	≥22 & <25				
		Large Metropolitan Fringe (Suburban)	0.97	(0.93, 1.01)	0.113
		Medium Metropolitan	0.94	(0.89, 0.99)	0.011
		Small Metropolitan	1.00	(0.97, 1.04)	0.825
		Micropolitan	0.99	(0.96, 1.01)	0.211
		Non-core (Rural)	0.98	(0.97, 0.99)	0.002
	≥22 & <25				
		Large Metropolitan Fringe (Suburban)	1.01	(0.97, 1.06)	0.585
		Medium Metropolitan	1.01	(0.96, 1.05)	0.804
		Small Metropolitan	1.04	(1, 1.08)	0.074
		Micropolitan	1.02	(0.99, 1.04)	0.147
		Non-core (Rural)	1.00	(0.99, 1.01)	0.832
	≥25				
		Large Metropolitan Fringe (Suburban)	1.04	(1.01, 1.07)	0.003
		Medium Metropolitan	1.06	(1.03, 1.08)	< 0.001
		Small Metropolitan	1.08	(1.05, 1.11)	< 0.001
		Micropolitan	1.04	(1.02, 1.05)	< 0.001
		Non-core (Rural)	1.01	(1.01, 1.02)	< 0.001
Did a Hurricane or Tropical Storm Occur? (Ref= No)					
	≥16 & <19	Yes	0.96	(0.8, 1.15)	0.661
	≥19 & <22	Yes	1.55	(1.30, 1.85)	< 0.001
	≥22 & <25	Yes	1.52	(1.24, 1.86)	< 0.001
	≥25	Yes	2.05	(1.79, 2.34)	< 0.001
Medically Underserved County (Ref= No)					
	≥16 & <19	Yes	0.91	(0.86, 0.95)	< 0.001
	≥19 & <22	Yes	0.94	(0.89, 0.98)	0.011
	≥22 & <25	Yes	1.01	(0.96, 1.07)	0.670
	≥25	Yes	1.16	(1.12, 1.20)	< 0.001

DOH	High Incidence Percentile (Cases/100,000 persons) [a]		RR [b]	95% CI [c]	P-value
Persons Identifying as NHPI (% of Pop; Ref = < 0.5%)					
	≥16 & <19	≥0.5%	1.36	(1.06, 1.74)	0.014
	≥19 & <22	≥0.5%	1.17	(0.87, 1.57)	0.290
	≥22 & <25	≥0.5%	0.75	(0.50, 1.12)	0.163
	≥25	≥0.5%	1.31	(1.05, 1.63)	0.016
Did the county include AIAN [d] tribal lands or an NHPI homeland? (Ref=No)					
	≥16 & <19	Yes	1.65	(1.48, 1.84)	< 0.001
	≥19 & <22	Yes	1.72	(1.52, 1.94)	< 0.001
	≥22 & <25	Yes	1.39	(1.20, 1.61)	< 0.001
	≥25	Yes	1.41	(1.27, 1.56)	< 0.001
Was the predominant employer in the county food manufacturing or processing? (Ref= No)					
	≥16 & <19	Yes	1.05	(0.74, 1.47)	0.803
	≥19 & <22	Yes	1.11	(0.75, 1.63)	0.607
	≥22 & <25	Yes	1.86	(1.28, 2.68)	0.001
	≥25	Yes	1.80	(1.38, 2.35)	< 0.001

[a]Reference incidence percentile was 1.0≤ and <4.0 cases/100,000 persons. The high incidence percentiles were those reporting ≥16 & <19, ≥19 & <22, ≥22 & <25 or ≥25 cases per 100,000 persons.

[b]Risk ratio

[c]Confidence Interval

[d]American Indian or Alaska Native