Spatial Point Pattern Analysis of Valley Fever in the Phoenix Metropolitan Area

John Ginos[a], Beckett Sterner[b,*] and Petar Jevtić[a]

[a]School of Mathematical and Statistical Sciences, Arizona State University, 1151 S Forest Ave, Tempe, 85281, Arizona, USA
[b]School of Life Sciences, Arizona State University, 1151 S Forest Ave, Tempe, 85281, Arizona, USA

ARTICLE INFO

Keywords:
Valley Fever
Spatial Point Pattern
Coccidiodes
Phoenix Metropolitan Area

ABSTRACT

Using a novel Arizona Medicaid data set, we model Valley Fever (VF) cases for each semester during the period 2013-2023 in the Phoenix Metropolitan Area (study region) as a spatial point process. We estimate the intensity function of VF cases, observed at residential locations of patients, as a function of environmental covariates available at high spatial resolutions, including land cover type, change in land cover type, Normalized Difference Vegetation Index (NDVI), change in NDVI, and using a recently developed Habitat Suitability Index (HSI) for the *Coccidiodes* fungus. Models with all covariates included and quadratic terms for NDVI, NDVI difference, and HSI tend to perform the best in terms of AIC. Among environmental factors considered, all except (current semester) NDVI and the quadratic term for HSI, appear to be statistically significant and useful for estimating the intensity VF cases in 15 or more semesters. Moreover, we find that NDVI appears to have a nonlinear relationship with the intensity of VF cases. Nevertheless, relative intensity plots comparing the fitted intensity from each model to the empirical intensity suggest that even our best model greatly underestimates the intensity in large portions of the study region. Our exploratory analysis in this study demonstrates that spatially localized environmental data is useful for examining VF incidence and may provide unique insights unavailable at larger spatial scales.

1. Introduction

Valley Fever (VF) is a fungal pathogen endemic to North America that shows a growing geographic range and number of human cases, especially in the western United States (see [19] and [17]). Research identifying environmental factors that influence VF incidence are important to public health authorities and communities. Among the environmental factors identified in recent work on VF incidence and seasonal variability are temperature, precipitation, air quality, and soil properties (see [21], [14], and [32]; see also [25] for a succinct summary of variables and methods used in the analysis of VF data). Often, these studies aggregate VF cases to administrative geographical units like census tracts, counties, or cities (e.g. see [18], [21], and [25], respectively). Aggregation of spatially referenced location data to larger geographical units may present problems for data analysis including the modifiable areal unit problem [11], in which results from spatial data analysis are greatly impacted by the scale of aggregation. A related problem when analyzing VF specifically is discussed in [10]. The authors report that percentages of land cover type in geographical units showed drastic variation depending on the spatial scale of units used (e.g. census tract, block group, and zip code; see [10]). We seek to avoid this problem to the greatest extent possible by considering the VF case locations (by residence) at the point level and using higher resolution covariate data.

We present the first spatial point process model[1] using residential addresses of human VF cases from a dataset of Arizona Medicaid patients in and around the Phoenix Metropolitan Area during 2013 to 2023 (hereafter referred to as the study period). In 2023, Maricopa County, Arizona, which includes the greater part of the Phoenix Metropolitan Area accounted for roughly 38% of VF cases according to provisional data from the Centers for Disease Control and Prevention (CDC) as well as figures reported by Maricopa county; see [12] and [22]. To our knowledge, the most granular analysis was a logistic regression approach that used presence/absence data derived from residential addresses of those with VF and pseudo-absence data randomly generated based on the population of geographical units [10]. Our approach incorporates covariate information uniformly from locations throughout the entire study area (where covariates are not NA) rather than at observed VF cases by residence and pseudo-absence locations alone. We

*Corresponding author. Email: bsterne1@asu.edu
[1]All code, covariate data and simulated VF data will be available at https://github.com/jginos/VF-SPP-Analysis.

treat the residential addresses of VF cases as a spatial point pattern and employ well-established techniques[2] to analyze factors that influence the spatially inhomogeneous intensity of VF cases. This approach also allows us to obtain fitted intensities for the whole study window (and at the locations of VF cases themselves) which can be used to determine where the model fit breaks down.

We find evidence supporting spatially localized effects of environmental covariates, including land cover category (including annual changes in land cover category), vegetation level, change in vegetation, and soil habitat suitability for the *Coccidiodes* fungal pathogen. The support we find for biotic spatial variables such as vegetation level and change is consistent with the enzoonotic hypothesis that rodents play an important role in spreading the disease, although we are not able to directly investigate the importance of wind versus animal-borne dissemination mechanisms as discussed in [25].

While the Medicaid dataset we analyzed has several limitations, our results suggest that VF exposure is associated with environmental covariates at a fine spatial scale and that additional environmental variation and/or additional nonlinear model structure may explain residual spatial clustering we observe in the data based on our best model.

2. Data and Methods

2.1. Valley Fever Data

Locations of residence for individuals diagnosed, treated, and/or tested for VF were obtained from the Arizona Health Care Containment System (AHCCCS; the Arizona state Medicaid agency) during the study period 2013 through 2023. The Arizona Medicaid (AHCCCS)[3] data used in this study were provided by the Arizona State University, Center for Health Information Research (CHiR). The data were aggregated by semester[4] and year with the first semester of each year defined as January 1st through June 30th; the remaining days of the year constitute the second semester. All individuals represented in the data set (hereafter simply patients) were identified with a member identification code associated with unique patients who were continuously enrolled during a given semester.

Demographic information for each patient including age and sex were included along with whether the individual was tested for, treated for, and/or diagnosed with VF in a given semester. Only the earliest diagnosis of VF was considered for patients with multiple (non-consecutive) diagnoses to avoid double counting unique diagnoses for individuals who may have unenrolled and re-enrolled in Medicaid during the study period. The demographic information was used to categorize the VF cases into four groups based on age and sex (see 2.3 for more details). Splitting the locations of VF cases into sub-patterns of spatial points allows us to consider how each pattern's spatial distribution is influenced by the same set of covariates and account for known heterogeneity in VF prevalence among such groups.

Due to the extreme contrast in the spatial density of VF cases between major Arizona metropolitan areas (like Phoenix and Tucson) and rural regions of the state and due to covariate data availability, we decided to limit our analysis to the Central Arizona-Phoenix Long-Term Ecological Research study area within metropolitan Phoenix[5] (hereafter study window; see Figure 1).

2.2. Spatial Covariate Data

As mentioned above, we are interested in how the intensity of VF cases is influenced by environmental covariates including the soil Habitat Suitability Index (HSI) of the *Coccidiodes* fungus (see [14]), NDVI, land cover category, and changes in land cover category that occur from one year to the next (we call this land cover change). Certain land cover changes are considered relevant for our purposes (see 2.2.5). Aside from environmental covariates, we also use the estimated Medicaid population density to account for the spatially varying population at risk within the study window.

[2] See chapter 9 of [7] and section 2.3.

[3] Hereafter, we refer to this data as Arizona Medicaid.

[4] The time series of semesterly VF data for the Medicaid population in Maricopa County had a 79% correlation with the series of all VF cases for Maricopa County as reported by the Arizona Department of Health Services (with 2023 data being considered provisional at the time of the analysis; see [4] for annual reports on Valley Fever Cases in Arizona Counties).

[5] While the CAP LTER study region is roughly rectangular in latitude/longitude coordinates, as will be seen below, the plots of covariate rasters and the study window used for the analysis contained herein are based on a bounding box constructed from areas where the NDVI rasters were defined (i.e. not NA) inside the study region, which, after projecting (using the WGS 84/Pseudo-Mercator projection) deviated from the original rectangular shape.

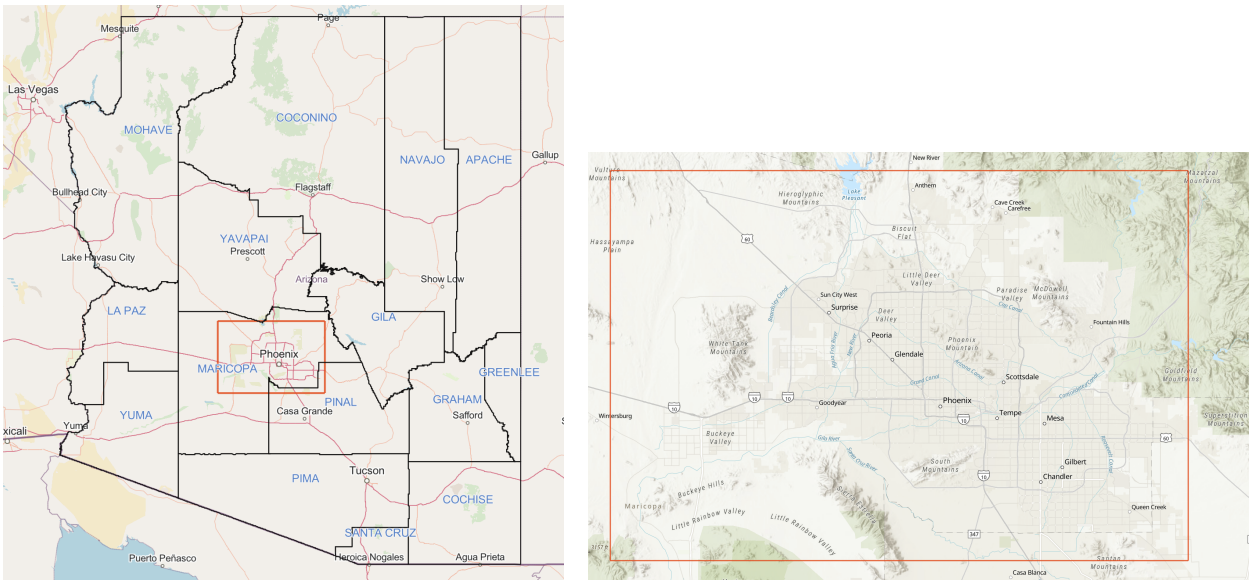

Figure 1: *Left: CAP LTER study window (red box) as situated within Arizona. Arizona county bounty borders (black) are shown along with county names (blue)(taken as a screenshot of the openstreetmap basemap available in ArcGIS Online [24]). County boundaries obtained from [5] and [2]. Right: CAP LTER study window (taken as a screenshot of the topographic basemap available in ArcGIS Online [16])*

All covariate data were rasterized[6] (if not already in raster format) and converted to a pixel image object for compatibility with the R package spatstat [8].

2.2.1. Medicaid Population Density (MPD) Estimates

We expect that the spatial distribution of VF cases varies with the Medicaid enrollment population density (MPD) which varies differently than the overall population density due, in part, to socioeconomic factors. Since we do not have access to the true MPD, we use the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates for Medicaid enrollment for the study period[7], respectively. Estimates are available at the census block group level; however, when aggregated to the county level, they underestimate the true Medicaid enrollment figures available from AHCCCS during the study period. To address this, we normalized all block group estimates (for all of Arizona) to sum to their respective county-level estimates[8] with data obtained for AHCCCS (for monthly AHCCCS enrollment data see [6]).

Areas where the Medicaid enrollment population is estimated to be zero present a numerical problem when modeling spatial point data using an offset to account for spatially varying exposure. To avoid zero-MPD areas, we set these block group estimates to one-fourth of the margin of error pertaining to the estimate[9]. The resulting Medicaid enrollment block group estimates are divided by block group area (in square miles) to obtain MPD.

To further reduce potential numerical issues related to outliers and to deal with the uncertainty in the estimates themselves, we rasterize (to 1,000-meter-resolution) and then smooth the MPD block group estimates using a Gaussian kernel. The standard deviation of the Gaussian kernel, often called the *bandwidth* for data smoothing purposes, is

[6] All rasters, including those for land cover, NDVI, land cover change, and HSI are projected using the WGS84/Pseudo Mercator projection and rescaled to reflect measurements in kilometers. NDVI was projected using the cubic spline method available as an option when using the R package Terra. Land cover data are projected using the modal pixel resampling scheme while land cover change was projected using the nearest neighbor scheme.

[7] Estimates regarding healthcare coverage by demographic group and at the block group level can be found ACS tables with ID B27010. This was used in conjunction with 2019 and 2022 TIGER/Line block group shapefiles provided by the U.S. Census Bureau to represent the 2010 and 2020 block group geometries [28], [29], [30]

[8] This was done for all census block groups in Arizona but only the subset of block groups contained in the study window were included in the analysis (this includes fractional portions of block groups near the boundaries of the study window)

[9] This fraction does not significantly impact the analysis as block groups with an estimated Medicaid enrollment population of zero not likely locations for VF cases and did not have large margins of error

chosen to be 2-kilometers to preserve local features in the MPD estimates but avoid large outliers in the population estimates. Finally, we normalize the smoothed MPD estimates (SMPDE) for use as an offset term in modeling VF cases. This is done by dividing all resulting pixel values by the sum of all pixel values in the study window. Such normalization reduces the remaining large-scale spatial variation in the offset term to avoid numerical problems when model fitting while maintaining positive values at each pixel location. In contrast, when treated as a model covariate (rather than an offset), we standardize the SMPD in the usual way by subtracting each pixel in the covariate image by the pixel mean and dividing by the pixel standard deviation.

2.2.2. Normalized Difference Vegetation Index (NDVI)

Raster data for the study region were obtained at the 30-meter resolution from the Evironmental Data Initiative [27]. Seasonal NDVI rasters for the study period and the year 2012 were used for the analysis. The average NDVI (by raster cell/pixel value) for winter and spring[10] of a given year was used for semester 1 covariates while the average NDVI for the summer and fall were used as semester 2 covariates during the study period.

NDVI is known have a lagged effect on VF cases for up to a year [26]. Thus, once the NDVI averages were obtained for each semester, a *differenced* NDVI covariate was obtained by differencing the NDVI for each semester with the NDVI from the prior semester. For example, for the first semester, the differenced NDVI covariate is calculated by computing the pixel-wise difference between the winter-spring average NDVI for 2013 and the fall-summer average NDVI for 2012[11]. Figure 2 shows the normalized difference of fall NDVI rasters from 2012 and 2023. The normalized difference, given by $\frac{\mathrm{NDVI}_{2023}-\mathrm{NDVI}_{2012}}{\mathrm{NDVI}_{2023}+\mathrm{NDVI}_{2012}}$, provides a usful pixel-wise calculation for inspecting changes in from the semester prior to and near the end of the study period. Note the significant areas of change in the west and southeastern portions of the study window. The increase in population and urban development in these areas may explain the apparent decreases in NDVI.

2.2.3. Coccidiodes Habitat Suitability Index (HSI)

Recent work by [14] showed that certain soil characteristics provide suitable conditions for the *Coccidiodes* fungus. The authors[12] created the Habitat Suitability Index to provide a spatially granular method of measuring where exposure to VF is a greater risk. The index takes values between 0 and 1 with values closer to 1 indicating increased suitability. The index is, in part, derived from soil survey data available and climatic (environmental) data. A resulting 10-meter resolution raster is obtained from the authors' predicted habitat suitability index within the study region[13]. Figure 3 shows the HSI raster for the study window; the large block of NA pixels in the northeast corner is due to unavailable soil survey data for the Tonto National Forest. We note that we use a single HSI covariate raster for the entire study period (the raster does not vary by year and semester).

2.2.4. Land Cover Category

The Multi-Resolution Land Characteristics (MRLC) Consortium provides data from the National Land Cover Database (NLCD) for use by the public (see [31]). From the available data, we have obtained annual land cover characteristics at the 30-meter resolution for the study window for 2013 through 2023. We are interested in whether four land cover categories (LCCs), namely open-area development[14] (LCC1), low intensity development areas (LCC2), medium intensity development (LCC3) areas, and high intensity development (LCC4) areas are associated with an increase in the spatial intensity of VF cases compared with the base category "Other" which includes undeveloped areas of all remaining classifications found in the study window.

As we mentioned above, the medicaid population density data contains inherent uncertainty. Thus, using land cover categories as a covariate is helpful as another measure to account for the spatially varying population density of the study window. While it is known that vectors of *Coccidiodes* are more likely to live in areas with less development, we recall that our data consists of the location of residence for associated VF infections. Thus, hypothesize a positive association between the spatial intensity of VF cases and areas of higher intensity development due to population

[10] Winter rasters include Landsat imagery from December 21st the prior year through March 19th of the current year. Spring rasters include imagery from March 20th through June 20th.

[11] Differenced NDVI covariates were obtained after conversion from a raster to a pixel image for compatibility with the R package spatstat

[12] We would like to thank Robert R. Dobos for help with obtaining the HSI data for this study.

[13] Using gridded Soil Survey Geographic Database (SSURGO) data and a table with HSI values containing a map unit key, the HSI raster layer was created using ArcGIS Pro 3.3.2 [15].

[14] According to the NLCD class legend and description "...These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes." [23]

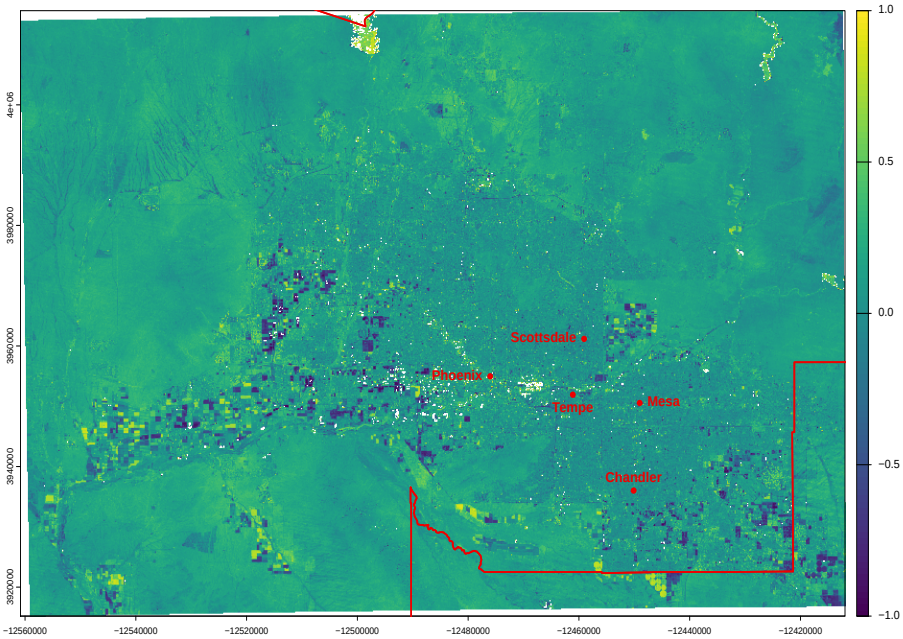

Figure 2: *Normalized difference (by pixel) between fall NDVI of 2023 and the fall NDVI of 2012 (the year prior to the start of the study period; 2012 NDVI data were used for creating a differenced NDVI covariate). Bright yellow areas denote an increase in NDVI over this period while dark purple areas represent a decrease in NDVI. Blank spaces are NA. Phoenix and a few nearby city centroids labeled in red along with county borders.* ***Source: The NDVI data used to create this figure was obtained from [27]***

concentration in these areas. It is also possible that those who live on less developed land (low-intensity development; LCC1) may be at a greater risk of VF exposure (especially land with high HSI soil) where they live.

2.2.5. Land Cover Change

From the same source as the land cover data, we have obtained data on the change in the land cover categories. Construction efforts for new development and/or renovation of existing structures or land involve excavation and possibly other soil-disturbing activities. With displaced soil comes increased risk for VF among both construction workers and those who live near the development site. To consider the affect that such development sites can have on the intensity of VF cases within the target population, we use land cover change data from the NLCD made available by the MRLC. The 30-meter resolution raster images classify 30-meter pixels in the study window based on whether a change from one land cover category to another (from annual land cover data) occurred from one product year to another (product years are when imagery is produced for public use).

Land cover change rasters were aggregated using a modal aggregation scheme[15], to an approximately 975-meter resolution. Since we are interested only in relevant land cover changes from a non-development category to a development category (open, low, medium and high intensity development, categories 1-4 as mentioned in 2.2.4), prior to aggregation, we chose to make the raster binary, indicating only pixels where a change occurred. Aggregated pixel values at the coarser 975-meter resolution were identified as a 'change' pixel if any of the subpixels from the original rasters contained a relevant change as described above.

2.3. Spatial Point Pattern Analysis

For individuals infected with VF, it may be difficult or impossible to identify the location of exposure to the *Coccidiodes* fungus. However, the locations of residence of patients infected with VF may offer valuable insight into the spatio-temporal dynamics of the disease. For example, in locations where soil and other environmental conditions are

[15] When projecting to the WGS 84 coordinate system, a nearest neighbor scheme was used for sampling pixel values

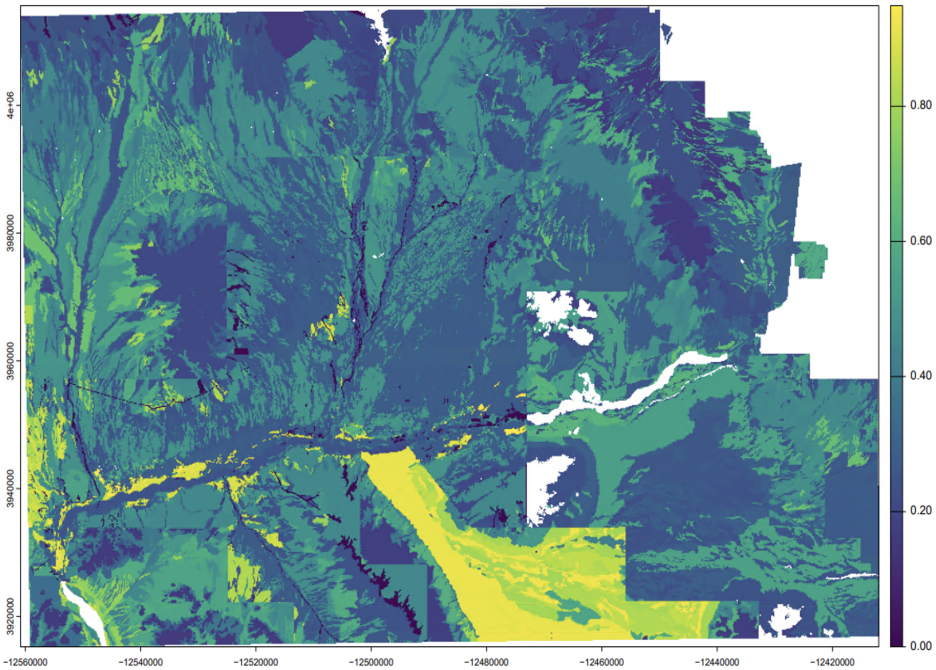

Figure 3: *Coccidiodes habitat suitability shown at the 10-meter resolution for the study window* ***Source:*** *Data obtained with the help of Robert R. Dobos and links to the original data sources provided in [14].*

right, nearby soil disturbance due to new construction, dust storms or other activity may infect multiple individuals in close proximity. Analyzing residential locations as a spatial point pattern allows us to understand the local conditions that increase the likelihood of VF exposure. Among the benefits of this approach is being able to extract covariate effects on disease intensity at a much smaller scale than if data were aggregated, while avoiding the so-called modifiable areal unit problem inherent in spatial data analysis. Treating VF cases as a spatial point pattern allows us to leverage environmental data collected at much finer spatial resolutions, which is especially useful when the spatial distribution of VF cases exhibits high levels of inhomogeneity over relatively small spatial scales (e.g. kilometers). We seek to estimate the intensity function of the VF cases over the study window as a function of high-resolution environmental covariate data. Loosely speaking, the intensity function of a spatial point process can be used as a measure of the expected number of cases in any bounded subregion of the study window W (including the entirety of W itself; see [7], chapter 6, section 6.3 for specifics on the intensity function of spatial point processes).

By analyzing the VF cases at the point level on the scale of kilometers, we may construct a smoothly varying surface of the intensity of VF cases. Specifically, we treat the VF cases as a spatially inhomogeneous, 'multitype' Poisson process. A 'multitype' process in our case refers to the categorization of VF cases based on qualitative attributes (or types) of the points called marks; we analyze the pattern as a multivariate spatial point process. In our case, each VF case may have one of four mark categories based on the age and sex of the person infected. The number of Arizona Medicaid patients during each semester of the study period are shown in Figure 4 and Table 1.

The age categories were determined by considering the distribution of VF cases by age as given in the Arizona Department of Health Services 'Valley Fever Annual Report' [3] over multiple years. It should also be noted that Medicaid eligibility, which begins at age 65, likely affects the Medicaid enrollment populations among older individuals.

Duplicate points, where both the mark category and location coincide (e.g. in the case of multiple people infected with VF in the same home), cause a technical violation of the assumption of a spatial Poisson process. Thus, to deal duplicated points, we randomly shuffle the indices of entire data set and take only the first mark category appearing in the shuffled data set at each location.

The MPD estimates vary dramatically within the study window and it is well known that environmental factors influence the abundance of VF cases in Maricopa County [21]. Thus, we model the spatially inhomogeneous

Semester	VF Cases
2013 1	140 (139)
2013 2	151 (150)
2014 1	243 (234)
2014 2	201 (190)
2015 1	197 (191)
2015 2	565 (552)
2016 1	584 (568)
2016 2	566 (548)
2017 1	576 (557)
2017 2	575 (559)
2018 1	620 (607)
2018 2	547 (536)
2019 1	550 (537)
2019 2	708 (688)
2020 1	659 (634)
2020 2	784 (766)
2021 1	857 (829)
2021 2	800 (775)
2022 1	710 (687)
2022 2	727 (701)
2023 1	723 (702)
2023 2	823 (800)
Total	12306 (11950)

Table 1
The total number of cases including residences with multiple reported cases. In parenthesis: the total number of cases used in this analysis (which involves the removal of duplicates, or points with identical attributes and locations; see 2.3). Recall that we define a case as anyone diagnosed and/or treated with VF. ***Source:*** *Arizona Medicaid (AHCCCS) 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

intensity of VF cases as a function of environmental and non-environmental covariates. Furthermore, since VF is not communicable among humans, the independence assumption that follows from modeling the VF case as a spatial Poisson process (accounting for the spatially inhomogeneous intensity) is a reasonable one.

For ease of comparison among fitted model coefficients, we standardize (pixel-wise) each non-categorical covariate by subtracting the mean pixel value and dividing by the standard deviation of pixel values within the study window. Additionally, to account for uncertainty in the underlying MPD estimates and as a method of facilitating model fitting, we smooth the MPD estimates (SMPD) over the study window using a Gaussian kernel with standard deviation equal to two kilometers. After smoothing MPD, we normalize the SMPD by diving each pixel by the sum over all pixels in the study window.

For year $i = 2013, \dots, 2023$ and semester $j = 1, 2$, we specify log-intensity of VF cases for age-sex category $k = 1, \dots, 4$, denoted by $\log \lambda_{i,j}(u, \boldsymbol{\theta}, k)$, at spatial location $u \in W$ as a linear function of environmental covariates. Here we let $\boldsymbol{\theta} \coloneqq (\gamma_{i,j}, \alpha_{i,j}(k), \boldsymbol{\beta}_{i,j})$ where $\boldsymbol{\beta}_{i,j}$ is a vector of model coefficients with length depending on the number of covariates included in the model and $\alpha_{i,j}(k)$ is the age-sex specific intercept term that depends on k. We first consider a log-linear base model with only an offset term and age-sex-specific intercept terms based on the marks of the VF cases. This model takes the form

$$\log \lambda_{i,j}(u, \boldsymbol{\theta}, k) = \gamma_{i,j} + \alpha_{i,j}(k) + \log SMPD_i(u) \tag{M1}$$

where we once again emphasize that u represents a spatial location in the study window W.

We build upon the base model by including land cover and land cover change covariates in the linear predictor of the log-intensity. The resulting intensity function is given by

$$\begin{aligned} \log \lambda_{i,j}(u, \boldsymbol{\theta}, k) = \gamma_{i,j} &+ \alpha_{i,j}(k) + \log SMPD_i(u) \\ &+ \beta_{i,j,1} LC_i(u) + \beta_{i,j,2} LCH_i(u). \end{aligned} \tag{M2}$$

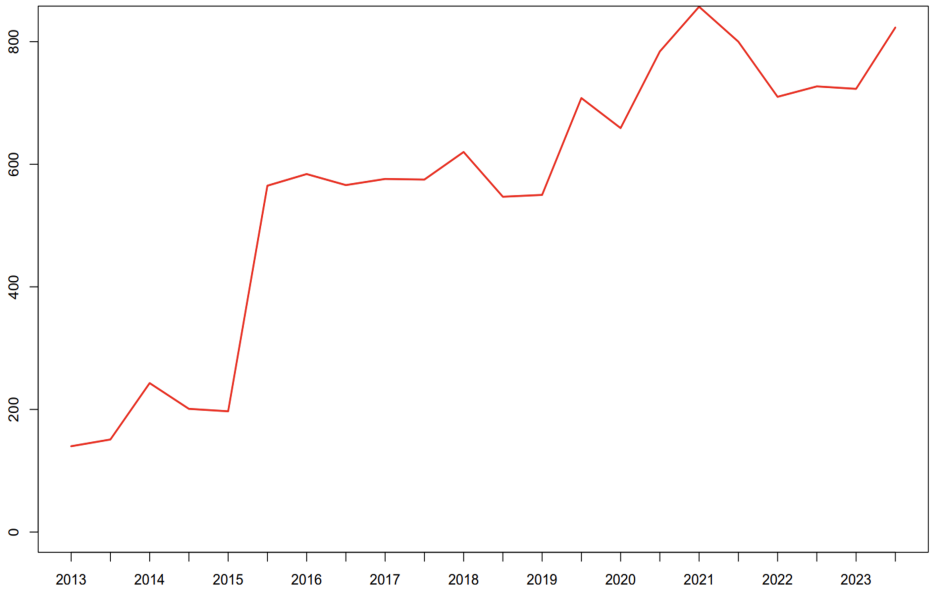

Figure 4: *Semesterly VF cases (with duplicates) among Arizona Medicaid patients during the study period.* ***Source:*** *AHCCCS 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

To understand the effect of HSI and NDVI on the intensity of VF cases, we add linear terms for these covariates to the structure of model M2 to get

$$\begin{aligned}\log \lambda_{i,j}(u,\boldsymbol{\theta},k) = \gamma_{i,j} &+ \alpha_{i,j}(k) + \log SMPD_i(u) \\ &+ \beta_{i,j,1}LC_i(u) + \beta_{i,j,2}LCH_i(u) \\ &+ \beta_{i,j,3}HSI(u) + \beta_{i,j,4}\left(NDVI_{i,j}(u) - NDVI_{prev.sem}(u)\right) \\ &+ \beta_{i,j,5}NDVI_{i,j}(u)\end{aligned} \tag{M3}$$

where $NDVI_{prev.sem}(u) := NDVI_{i-1,j+1}(u)$ if $j = 1$ and $NDVI_{prev.sem}(u) := NDVI_{i,j-1}(u)$ if $j = 2$. Note that for the NDVI difference covariate in 2013, semester 1, we also use the summer-fall average NDVI for 2012, one year before the start of the study period.

Preliminary nonparametric analysis of the relationship between NDVI, HSI, and NDVI difference revealed a nonlinear relationship between these covariates and the intensity of VF cases. Thus, we consider lastly, a log-quadratic (quadratic in HSI and NDVI covariates) model of intensity given by

$$\begin{aligned}\log \lambda_{i,j}(u,\boldsymbol{\theta},k) = \gamma_{i,j} &+ \alpha_{i,j}(k) + \log SMPD_i(u) \\ &+ \beta_{i,j,1}LC_i(u) + \beta_{i,j,2}LCH_i(u) \\ &+ \beta_{i,j,3}HSI(u) + \beta_{i,j,4}\left(NDVI_{i,j}(u) - NDVI_{prev.sem}(u)\right) \\ &+ \beta_{i,j,5}NDVI_{i,j}(u) + \beta_{i,j,6}\left(HSI(u)\right)^2 \\ &+ \beta_{i,j,7}\left(NDVI_{i,j}(u) - NDVI_{prev.sem}(u)\right)^2 + \beta_{i,j,8}\left(NDVI_{i,j}(u)\right)^2.\end{aligned} \tag{M4}$$

Note that SMPD estimates are used as an offset term to ensure that the dependence of VF cases on covariates is considered after accounting for the spatially varying at-risk population for VF.

The R package spatstat provides a robust toolkit for the analysis of spatial point patterns. Exploratory analysis, statistical inference on the covariates, and model validation were conducted mainly using tools from the spatstat package. We have used the spatstat function ppm to fit the models. We note that we have increased the resolution of the quadrature scheme used in model fitting from its default setting to a 250×250 grid[16] of dummy points. Furthermore,

[16] While the entire grid is 250×250 dummy points, those points located where any covariate was labeled NA were removed from the quadrature scheme used to estimate the intensity over the window of study.

	AIC: (4)	AIC: (3)	AIC: (2)	AIC: (1)
2013 1	**1490.8261**	1502.654983	1498.095707	1530.525917
2013 2	1625.430587	1617.534461	**1606.6054**	1651.456671
2014 1	**2301.0941**	2326.963181	2343.968079	2343.830204
2014 2	**1916.3228**	1944.763234	1956.011895	1962.516841
2015 1	2007.380156	2008.619041	**1986.3273**	2010.535946
2015 2	**4530.7641**	4587.026617	4546.869897	4621.828938
2016 1	4730.961186	**4690.4835**	4730.490519	4790.093048
2016 2	**4740.3089**	4787.640264	4778.695762	4830.225735
2017 1	**4616.0556**	4624.359884	4638.122792	4698.977958
2017 2	**4652.2217**	4711.542317	4727.969714	4787.417302
2018 1	5252.308698	5167.782327	**5130.8837**	5217.765404
2018 2	**4532.1**	4623.836849	4602.096133	4671.321491
2019 1	**4544.8476**	4551.401777	4596.731665	4642.235404
2019 2	5515.855873	5505.192584	**5499.8811**	5610.138492
2020 1	**5298.7449**	5303.879198	5360.385088	5386.454817
2020 2	6110.764687	6110.931191	**6110.2537**	6222.545192
2021 1	6643.335278	**6636.331**	6646.373956	6727.697617
2021 2	6204.369494	6242.732924	**6196.6494**	6313.311271
2022 1	**5594.7491**	5689.829514	5677.789615	5761.771188
2022 2	**5616.5739**	5678.400042	5653.380299	5790.46562
2023 1	5740.364042	**5704.4113**	5719.981215	5830.867812
2023 2	**6232.4157**	6295.364232	6290.720454	6392.168044

Table 2
AIC scores for models 1-4. Bold values indicate the model with lowest AIC in a given year and semester.

our model results involve randomization as we have used the Huang and Ogata model improvement procedure ([20]) to obtain more accurate parameter estimates[17].

3. Model Results

Estimated coefficients for all four models appear in Tables 6, 5, 4, and 3 located in the appendix, while table 2 displays AIC scores for models M1-M4 for all twenty-two semesters. As expected, the most complex model (M4) outperforms the remaining models for most (thirteen of twenty-two) semesters in terms of Aikake Information Criterion (AIC). Notably, all semesters show improved model fit when incorporating covariate effects. This is confirmed by likelihood ratio test (LRT) results as show in in 7 in the appendix, which compare models M1 and M4 for all semesters.

Since our goal is to understand how all environmental factors together affect the intensity of VF cases, we are primarily interested in the interpretation of results for models M3 and M4. Note that for model M3, given an estimated coefficient, β for a standardized, continuous covariate (like HSI or NDVI/NDVI difference), an increase by one standard deviation in the covariate is associated with a change in the intensity of VF cases by a factor of e^{β} per square kilometer. For model M4, a unit change in the standard deviation of the continuous covariate depends on the value of the estimated coefficients and a particular value of the covariate. In 2013, all other covariates being held constant[18], at the mean value of HSI, which is about 0.4 over W, a one standard deviation (which is 0.187 over W) increase in HSI increases the intensity of VF cases by a factor of $e^{\beta_{2013,1,3}+\beta_{2013,1,6}(2*0+1)} = e^{0.517-0.405} \approx 1.12$ or by 12% per square kilometer. The effect of a one standard deviation increase at the mean HSI is associated with an increase in VF case intensity in all semesters except the first semester of 2015 and the second semester of 2022. The greatest effect at the mean HSI is observed in the first semester of 2019, where we estimate a 63% increase in intensity per square kilometer for a one standard deviation increase in HSI while the average effect of one standard deviation increase at the mean value of HSI is an increase in intensity of about 28% per square kilometer. However, it should be noted that we estimate a statistically significant linear effect from HSI in most semesters, the quadratic term in the first semester of 2019 (where we observed the maximum marginal effect of HSI) and indeed in most semesters is estimated to be not significantly different from

[17]See documentation of the ppm function in spatstat for details.
[18]When interpreting model parameters, we assume all other covariates are constant.

zero. This provides evidence for the utility of HSI as a useful environmental factor to consider when modeling VF incidence at fine spatial scales. We refer to Figure 6 for a visual guide to the marginal effects of the continuous covariates when the values of the other covariates (and population offset) are fixed at specified values/categories).

Time series plots of coefficients for HSI, NDVI, NDVI difference, and their quadratic terms are given in Figure 5. There is an apparent relationship between the average NDVI for each semester and the trajectory in the coefficients for NDVI and differenced NDVI over the study period. The pattern suggests that when the average NDVI is higher, the estimated log-linear effect of NDVI on the intensity of VF cases is decreased. A similar, but inverted relationship appears to exist between the effect of NDVI difference and average NDVI. This may be due to other environmental factors associated with NDVI like soil moisture and precipitation that are known to influence the number of VF cases in the study region (see [21] and [26]). The changing signs for the NDVI and NDVI difference coefficients are also likely related to the dynamic effect of NDVI on VF incidence as discussed in [26]. Despite the dynamic nature of the relationship between NDVI and VF incidence, the negative and consistently significant quadratic terms in the model suggest that a nonlinear relationship is more appropriate. Moreover, values of NDVI within $0.2 - 0.4$ seem to have the most pronounced effect on intensity as can be seen in Figure 6.

Another interesting feature in Figure 5 is the increasing volatility in estimates for the coefficient of NDVI difference following the second semester of 2017. Moreover, from the first semester of 2018 through the end of 2023, in the same series, it can be observed that the coefficients are statistically significant at the 95% level or greater. While a notable increase in all VF cases (i.e. the case count not restricted to only those enrolled in Medicaid) was observed in Maricopa county during this same period [9], no such spike appears within the number of cases used in this study (see Table 4). An alternative explanation is that volatility in environmental factors that affect NDVI like precipitation and temperature are contributing to the affect of NDVI as a model covariate. A similar pattern of variation in the average NDVI for the region W for the period 2016-2020 as well.

Recall that for factor-valued covariates land cover and land cover change, the coefficient β reflects the intensity of VF cases per square kilometer for each category relative to the base category "Other" (which is treaed as having coefficient $\beta = 0$).

As expected, the intensities of VF cases among individuals over 45 years of age (for both sexes) is greater relative to the intensity of males under 45 years; this can be seen by inspection of the age-sex specific intercept coefficients in Table 3 in the 4. No clear pattern exists, in sign or statistical significance, for the estimated intercept term of the intensity of VF cases among females under 45 years relative to males under 45. The magnitude of estimated intercept terms for individuals over 45 years of age do not appear compare in the same way for each semester. These results suggest that, while there is a greater intensity of VF cases among individuals over 45 years of age of both sexes relative to the intensity of males under 45 years, no such pattern exists regarding the intensities of VF cases males and females. Moreover, the the age-sex-specific intercept terms for both males and females above 45 years seem to be larger for earlier semesters when compared to later semesters. This is likely due to changing Medicaid enrollment standards and demographic shifts during the study period.

For most semesters and all land cover categories reported, large positive effects are evident, with the exception of the outliers in 2013 for areas in LCC4. Taking the average over the twenty-two semesters, areas in LCC2 and LCC3 have intensities that are $e^{2.33} \approx 10$ and $e^{2.9} \approx 18$ times greater than areas classified according to the base land cover category "Other", respectively. This is roughly in line with expectations. The large positive effects on intensity of LCC2 and LCC3 relative to the base "Other" category are likely due to population-related factors not accounted for in the model. Estimates of the coefficients for land cover change are also in line with expectations. Areas with new development are estimated to have an intensity of VF cases of about 1.25 to 3.47 times that of areas with no change in development when considering the eighteen semesters where the land cover change estimates were statistically significant at the 95% level or greater.

3.1. Relative Intensity Analysis

Relative intensity[19] estimates are a simple way to show congruence between fitted spatial point process models and a smooth, nonparametric estimate of intensity of VF cases for a given semester. Let $X_{i,j}$ denote the unmarked spatial point process of VF cases (i.e. the superposition of all four processes of VF cases, one for each age-sex category) in year i, semester j and let $\lambda_{i,j}(u)$ denote the true intensity of $X_{i,j}$ at $u \in W$. Note that $X_{i,j}$ would be the unobserved process generating VF cases among the Medicaid subpopulation and the process for which we are interested in learning

[19] See [7], chapter 11, section 11.2.1 for a description of this approach.

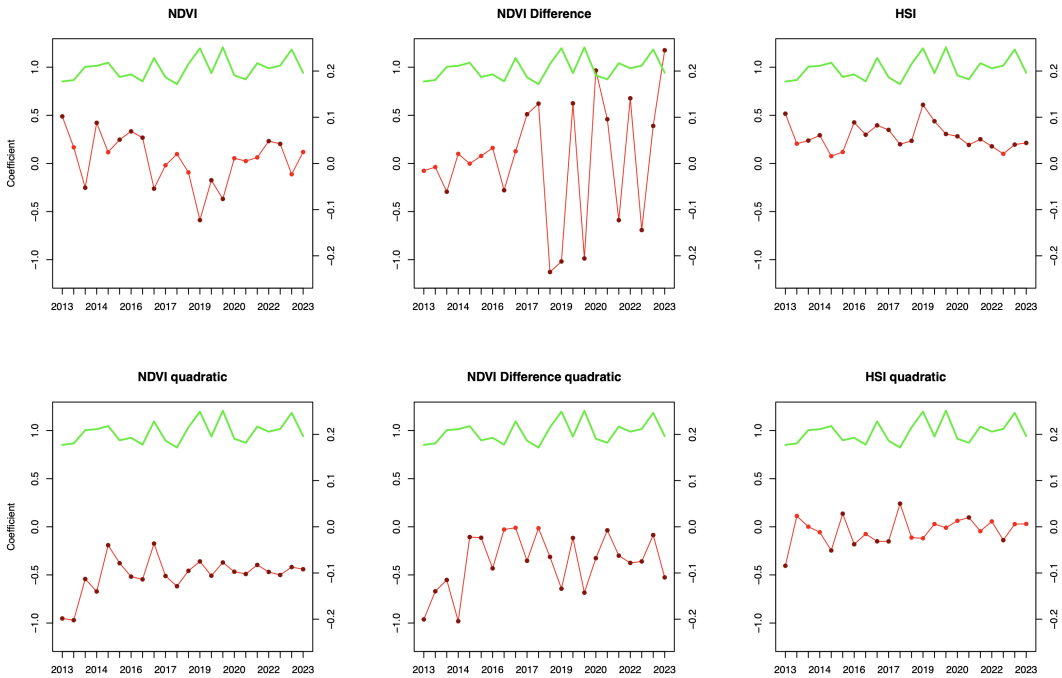

Figure 5: *Time series of coefficient estimates from model M4 for all semesters. The secondary vertical axis pertains to the pixel-wise average NDVI (green curve). Dark red points indicate semesters where the coefficient was significant at the 95% level or greater (see Table 3).*

the intensity function. The relative intensity of $X_{i,j}$ and a fitted model for the intensity of $X_{i,j}$ is given by $r_{i,j}(u) = \frac{\lambda_{i,j}(u)}{\hat{\lambda}(u)}$ where $\hat{\lambda}_{i,j}(u, \theta) := \sum_{k=1}^{4} \hat{\lambda}_{i,j}(u, \theta, k)$ denotes the estimated intensity function of $X_{i,j}$. We estimate the relative intensity based on an observation of $X_{i,j}$ by

$$\hat{r}_{i,j}(u) = \frac{1}{e(u)} \sum_{l=1}^{n} \frac{1}{\hat{\lambda}(x_l, \theta)} \kappa\left(u - x_l\right)$$

where $x_l, l = 1, \ldots, n$ denotes the locations of n VF cases in an observation of $X_{i,j}$, κ is a smoothing kernel, and $e(u)$ is an edge correction[20] of the estimate to deal with the bias arising from fact that we only observe the pattern in the window W of finite area[21] [8]. We use an adaptive-bandwidth kernel density approach with each point in the data set weighted according to the inverse of fitted intensity of models M1=M4. Figures 7, 8, 9 and 10 showed the estimated relative intensities for each model in both semesters of 2016 and 2022 (similar estimated relative intensity patterns among the different models hold for the other semesters which are not shown here). Nonparametric, weighted intensity estimates were estimated using a kernel density approach with an adaptive bandwidth via the spatstat function adaptive.density to account for the dramatic spatial inhomogeneity in VF cases in the study window. A global bandwidth of 5 kilometers was chosen for use as a pilot bandwidth which is passed to the spatstat function bw.abram.ppp[22] (see [8] and [7], especially chapter 6, for details on nonparametric intensity estimation). In the estimates of relative intensity we replace outlier weights (which occurred when fitted intensities were less than 0.005) with the inverse of the minimum value of the fitted intensities at observed VF cases that were greater than or equal to 0.005. This was done to limit the visual affect of outliers and still allow for comparisons of clusters in locations where the fitted models under- or overestimate intensity of VF cases. More severe underestimation outliers were more common in the relative intensities for model M4 but the spatial area locations with an estimated relative intensity greater than or equal to 2 appear to be slightly smaller

[20] We used the spatstat default, adaptive bandwidth edge correction given by $e(u) = \int_W \kappa_{h(u)}(u - v)dv$ for $u \in W$. Here $h(u)$ denotes a spatially varying bandwidth, see section 2.3 in [13] for more details on edge correction factors in adaptive bandwidth kernel density estimation.

[21] see chapter 6, section 6.5.1 and chapter 11, section 11.2.1 in [8] for more details on edge correction factors and relative intensity estimates for model validation, respectively.

[22] See also [1] and [13] for details on how this method works.

for the models with covariates included when compared to the offset only model. Indeed, in 18 of the 22 semesters, model M4 had the highest percentage of 62, 500 pixels that had an estimated relative intensity of $0.5 - 2$ which indicate areas where the model did not under/overestimate intensity by a factor of two or more. Model M4 had an average of 4%, 1% and 1% more pixels in range $0.5 - 2$ than models M1, M2, and M3, respectively[23].

The percentage of points treated as outliers when estimating and visualizing the relative intensity was also larger for the offset only model in 15, 12, and 13 of the 22 semesters for models M4, M3, and M2, respectively.

In Figures 7, 8, 9, and 10 we can see that hot spots of higher relative intensity are present for all models, albeit in different locations of the study window in some cases. Note that near the perimeter of the study window, all of the models tend to overestimate intensity, however, the region in the northeast corner of the study area is mostly Tonto National Forest and no cases can be observed here, thus we may ignore this region. Furthermore, the underestimation of cases near the perimeter is likely due to the general lack of observed cases in portions of these regions and is subject to higher uncertainty.

4. Discussion and Conclusion

While our approach provides evidence at the spatial point level for environmental covariate effects on VF incidence, even our best model under/overestimates VF case intensity in large portions of the study region and especially in areas at the boundary of the more densely populated region of the study area. There are multiple possible reasons for this; we cite a couple of reasons here and suggest future areas of research. First, while working with Medicaid data provides unique insight into the economic and public health burden on the Arizona Medicaid system caused by VF, the Arizona Medicaid population is only a subpopulation of those who are susceptible to VF infection. Moreover, the spatially varying Medicaid population is not simply proportional to the overall population in the study area. This limits the areas where we may observe VF cases and may lead to an underrepresentation of VF cases in regions where the covariates may have strong localized effects. Further, we are using only estimates of the spatially varying Medicaid population. Indeed, some of the census block groups used for the estimates have relatively large margins of error.

A second limitation is that, aside from the fact that residential addresses of those treated and/or diagnosed with VF do not necessarily reflect the nature of the covariates where the individual was exposed to the *Coccidiodes* fungus, large outliers in the relative intensities of the fitted models may suggest nonlinear covariate effects beyond the log-quadratic relationships considered herein. Additionally, large outliers may indicate spatially varying environmental factors not considered herein and/or interactions between environmental factors and individual behavior (e.g. gardening, small-scale excavation, and digging on residential properties). Thus, methods for considering unobserved spatial 'random effects' should be considered[24] in future analysis.

Our exploratory analysis in this paper has highlighted the utility of spatial point pattern analysis techniques in understanding VF incidence, however, since our data is temporally aggregated to semesters, this poses similar challenges to the spatial aggregation of VF data. If the VF cases were aggregated at, say, monthly intervals, we would observe different spatial point patterns which may reveal effect sizes that differ from those estimated at the semester-based temporal scale. Temporal aggregation of covariate data also plays a critical role in analyzing VF case intensity. It has been shown, for example, that one-month lagged precipitation is useful for predicting VF incidence at the county level ([21]). Monthly variation in NDVI, which is affected by precipitation, may have a monthly (or smaller temporal scale) lag effect that we do not observe. To address this, higher frequency environmental covariate data could be analyzed in a spatiotemporal point pattern model for VF cases that are identified by location as well as date/month of diagnosis/first symptoms.

Acknowledgements

This work was supported under the National Institutes of Health, United States grant DMS-1615879.

This study was made possible by an ongoing partnership between Arizona State University's Center for Health Information Research (CHiR) and the Arizona Healthcare Cost Containment System (AHCCCS). Special thanks to CHiR for providing valuable technical support to the study.

[23] It should be noted that these percentages were calculated after outlier weights were adjusted as described earlier.

[24] See [7], chapter 12 for a discussion of relevant methods (e.g. Cox processes).

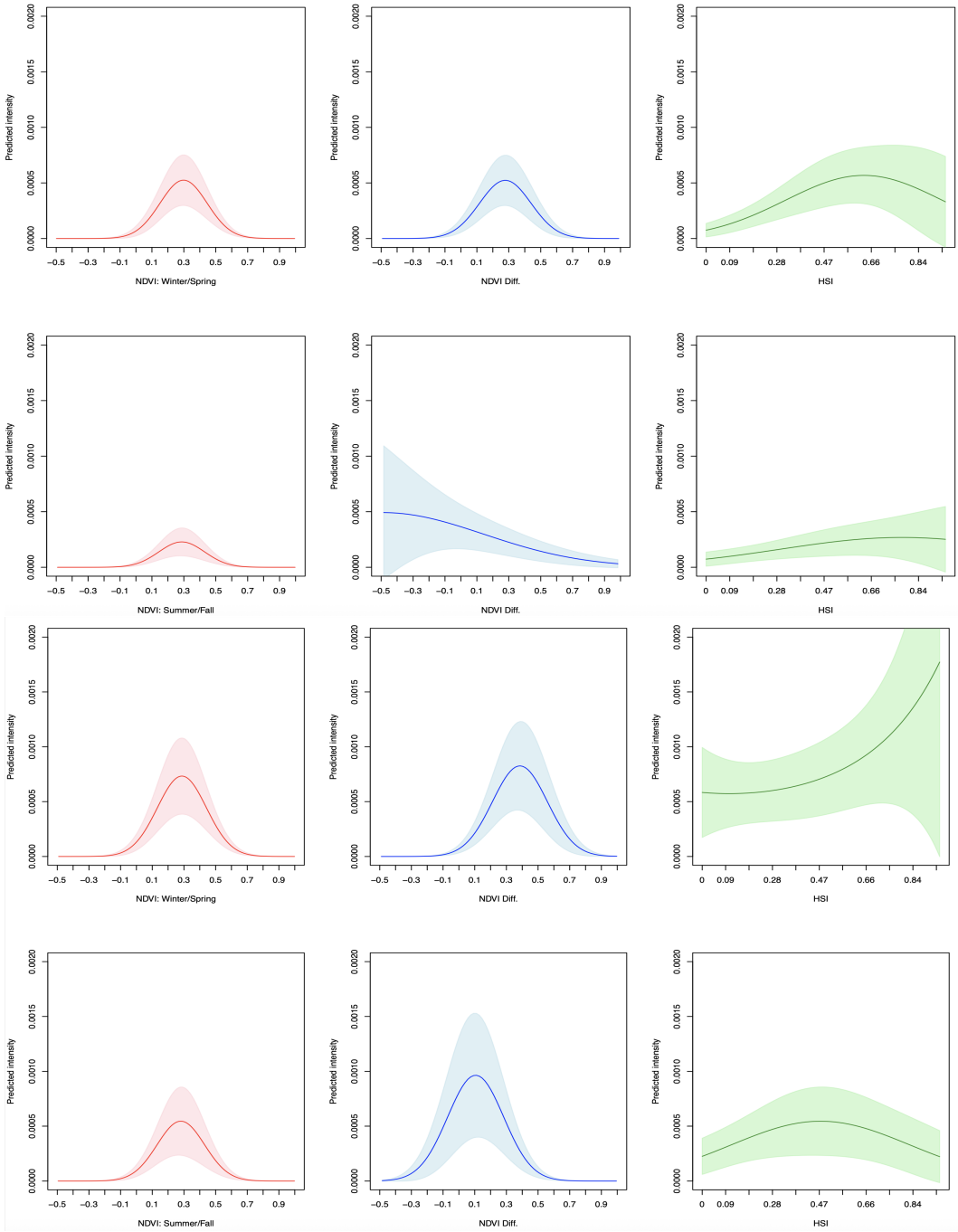

Figure 6: *Marginal effect of NDVI, NDVI difference and HSI on intensity of VF cases based on model M4. Top row: 2016 semester 1. Second row from top: 2016 semester 2. Third row from top: 2022 semester 1. Bottom row: 2022 semester 2. Effects are calculated by setting other continuous covariates to their* 0.75 *quantile. Land cover and land cover change are left as their base categories, "Other" and "No Change", respectively.*

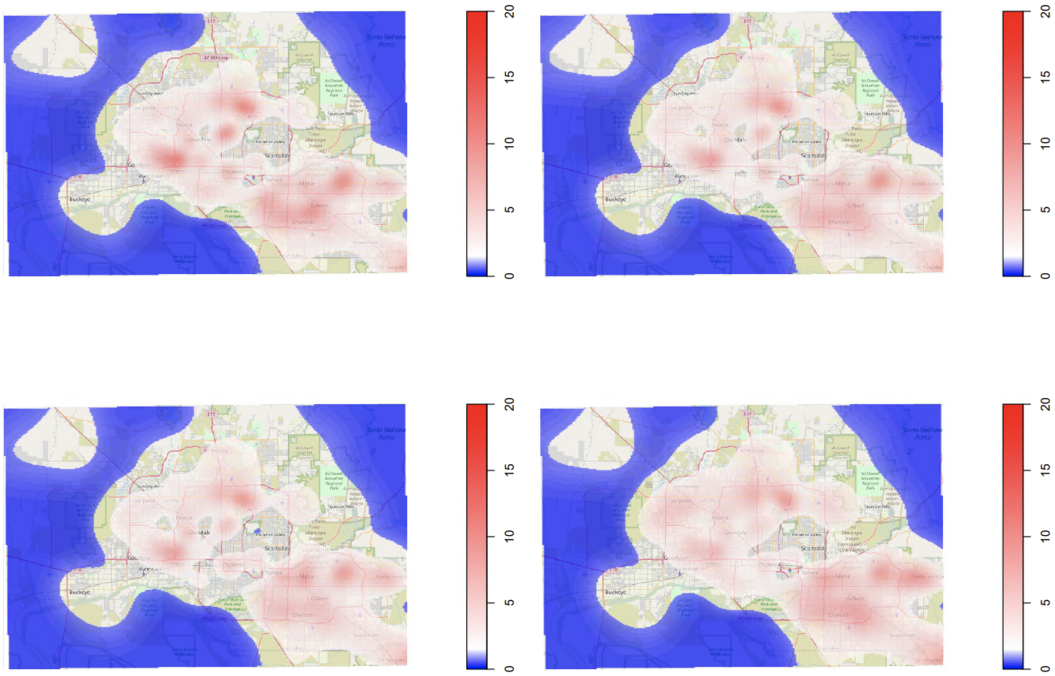

Figure 7: *Estimated relative intensities for 2016, semester 1. Weights less than* 0.005 *were replaced with the minimum weight greater than or equal to* 0.005*. Relative intensities below 2 and above 0.5 are not shown in color to highlight where the models underestimate or overestimate the intensity of VF cases as show in red and blue, respectively. Top left: Model M4. Top right: Model M3. Bottom left: Model M2. Bottom right: Model M1.* ***Source:*** *AHCCCS, 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

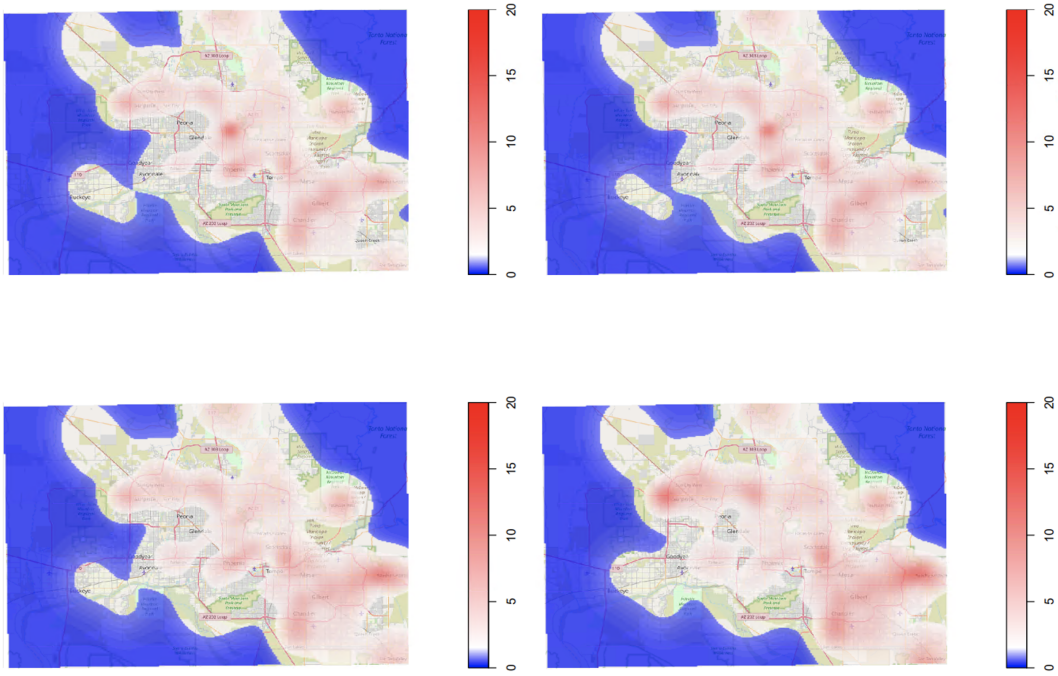

Figure 8: *Estimated relative intensities for 2016, semester 2. Weights less than* 0.005 *were replaced with the minimum weight greater than or equal to* 0.005*. Relative intensities below 2 and above 0.5 are not shown in color to highlight where the models underestimate or overestimate the intensity of VF cases as show in red and blue, respectively. Top left: Model M4. Top right: Model M3. Bottom left: Model M2. Bottom right: Model M1.* ***Source:*** *AHCCCS, 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

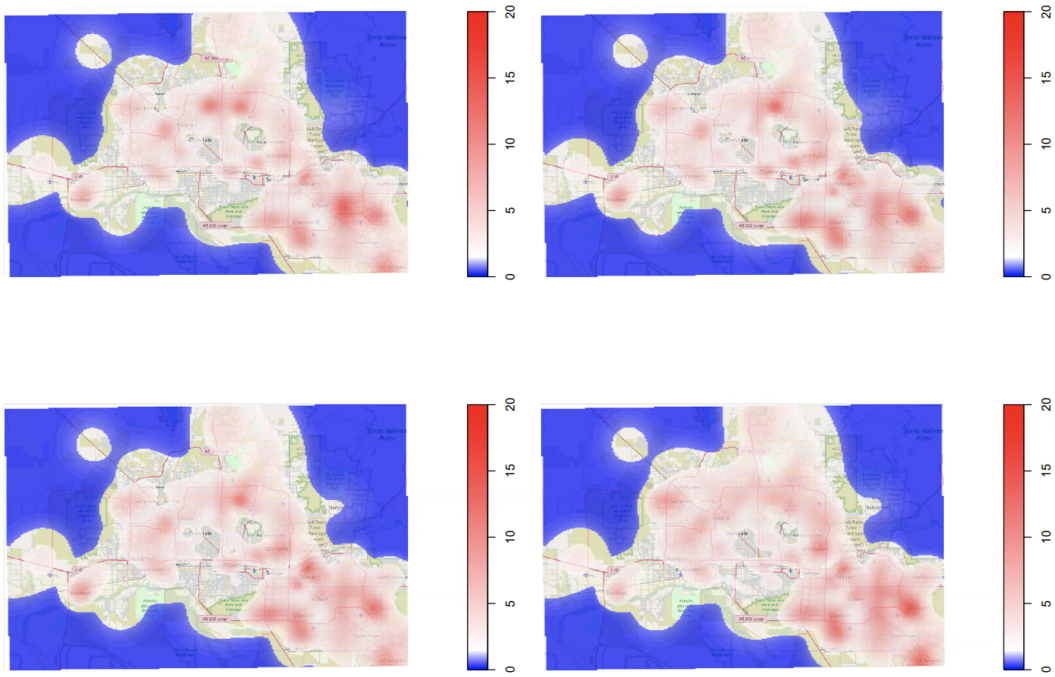

Figure 9: *Estimated relative intensities for 2022, semester 1. Weights less than* 0.005 *were replaced with the minimum weight greater than or equal to* 0.005*. Relative intensities below 2 and above 0.5 are not shown in color to highlight where the models underestimate or overestimate the intensity of VF cases as show in red and blue, respectively. Top left: Model M4. Top right: Model M3. Bottom left: Model M2. Bottom right: Model M1.* ***Source:*** *AHCCCS, 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

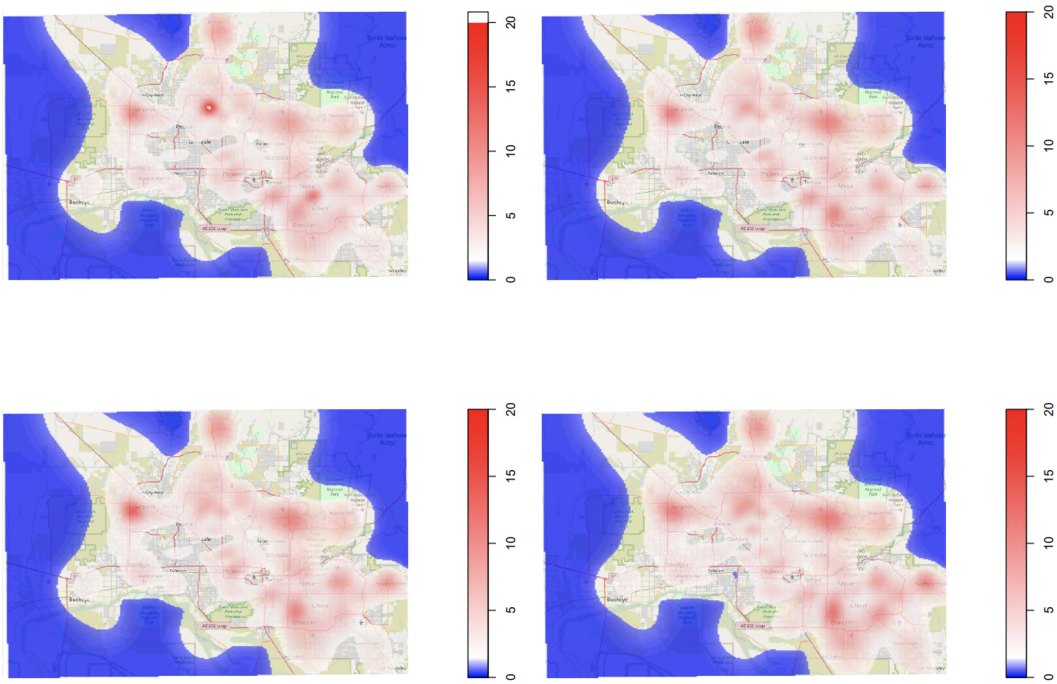

Figure 10: *Estimated relative intensities for 2022 semester 2. Weights less than* 0.005 *were replaced with the minimum weight greater than or equal to* 0.005*. Relative intensities below 2 and above 0.5 are not shown in color to highlight where the models underestimate or overestimate the intensity of VF cases as show in red and blue, respectively. Top left: Model M4. Top right: Model M3. Bottom left: Model M2. Bottom right: Model M1.* ***Source:*** *AHCCCS, 2023. CHiR created patient record files from Arizona Medicaid (AHCCCS) data that were used for this analysis.*

References

[1] Ian S Abramson. On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*, pages 1217–1223, 1982.

[2] ArcGIS Online Contributor. AZ Counties V2 [Feature layer]. https://services6.arcgis.com/l7uujk4hHifqabRB/arcgis/rest/services/AZ_Counties_V2/FeatureServer, 2025. Accessed via ArcGIS Pro on July 13, 2025.

[3] Arizona Department of Health Services. Valley fever surveillance reports, 2013–2023. https://www.azdhs.gov/documents/preparedness/epidemiology-disease-control/valley-fever/reports/valley-fever-2023.pdf, 2023. Annual reports accessed between 2013 and 2025 from the Arizona Department of Health Services. Accessed: 2025-06-25.

[4] Arizona Department of Health Services. Valley fever surveillance reports, 2013–2023. https://www.azdhs.gov/preparedness/epidemiology-disease-control/valley-fever/index.php#data, 2024. Annual reports from 2013 to 2023 available under the "Data and Reports" section. Example: 2022 report accessed at https://www.azdhs.gov/documents/preparedness/epidemiology-disease-control/valley-fever/reports/valley-fever-2022.pdf.

[5] Arizona Geographic Information Council. Arizona counties shapefile. https://azgeo-open-data-agic.hub.arcgis.com/datasets/545fc0393f0747cb9665848a75f9b23c_0/explore, 2025. Shapefile accessed from the AZGeo Open Data Portal on 2025-06-25.

[6] Arizona Health Care Cost Containment System (AHCCCS). Ahcccs monthly enrollment reports, 2013–2023. https://archive.azahcccs.gov/, 2024. Accessed December 20, 2024. Navigate to: Resources > Reports > Population Statistics to locate monthly enrollment data.

[7] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015.

[8] Adrian Baddeley and Rolf Turner. spatstat: An r package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005.

[9] Carla P Bezold. Notes from the field: increase in coccidioidomycosis—arizona, october 2017–march 2018. *MMWR. Morbidity and Mortality Weekly Report*, 67, 2018.

[10] Heidi E Brown, Mu Wangshu, Khan Mohammed, Tsang Clarisse, Liu Jian, and Tong Daoqin. Spatial scale in environmental risk mapping: A valley fever case study. *Journal of Public Health Research*, 6(2):jphr–2017, 2017.

[11] Michael Buzzelli. Modifiable areal unit problem. *International encyclopedia of human geography*, page 169, 2019.

[12] Centers for Disease Control and Prevention. Valley fever: Rates and trends of reported cases. https://www.cdc.gov/valley-fever/php/statistics/index.html, 2025. Text-based infographic retrieved on 2025-06-25 from CDC Valley Fever statistics page.

[13] Tilman M Davies and Adrian Baddeley. Fast computation of spatially adaptive kernel estimates. *Statistics and Computing*, 28(4):937–956, 2018.

[14] Robert R. Dobos, Kaitlin Benedict, Brendan R. Jackson, and Orion Z. McCotter. Using soil survey data to model potential coccidioides soil habitat and inform valley fever epidemiology. *PLOS ONE*, 16(2):1–17, 02 2021.

[15] Esri. ArcGIS Pro: Version 3.3.2. https://pro.arcgis.com/, 2024. Environmental Systems Research Institute, Redlands, CA, USA.

[16] Esri. Screenshot of arcgis online topographic basemap showing the [phoenix metropolitan area]. https://www.arcgis.com/home/webmap/viewer.html, 2025. Basemap: Esri Topographic. Map data © Esri, USGS, NOAA, EPA, NPS, and others. Screenshot captured on 2025-06-25. See ArcGIS Terms of Use for basemap licensing.

[17] Morgan E Gorris, Kathleen K Treseder, Charles S Zender, and James T Randerson. Expansion of coccidioidomycosis endemic regions in the united states in response to climate change. *Geohealth*, 3(10):308–327, 2019.

[18] Jennifer R Head, Gail Sondermeyer-Cooksey, Alexandra K Heaney, Alexander T Yu, Isabel Jones, Abinash Bhattachan, Simon K Campo, Robert Wagner, Whitney Mgbara, Sophie Phillips, et al. Effects of precipitation, heat, and drought on incidence and expansion of coccidioidomycosis in western usa: a longitudinal surveillance study. *The Lancet Planetary Health*, 6(10):e793–e803, 2022.

[19] Madelynn Howard, Christie Sayes, John P Giesy, and Yang Li. Valley fever under a changing climate in the united states. *Environment International*, page 109066, 2024.

[20] Fuchun Huang and Yosihiko Ogata. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, 8(3):510–530, 1999.

[21] Daniel R Kollath, Joseph R Mihaljevic, and Bridget M Barker. Pm10 and other climatic variables are important predictors of seasonal variability of coccidioidomycosis in arizona. *Microbiology Spectrum*, 10(2):e01483–21, 2022.

[22] Maricopa County Department of Public Health. Confirmed valley fever cases in maricopa county as of may 31, 2025. https://www.maricopa.gov/5813/Valley-Fever, 2025. Public health surveillance figure retrieved on 2025-06-25.

[23] Multi-Resolution Land Characteristics (MRLC) Consortium. National land cover database (nlcd) class legend and description. https://www.mrlc.gov/data/legends/national-land-cover-database-class-legend-and-description, n.d. Accessed: 2025-06-25.

[24] OpenStreetMap contributors. Screenshot of arcgis online openstreetmap basemap showing the state of arizona. https://www.arcgis.com/home/webmap/viewer.html, 2025. Basemap: OpenStreetMap via ArcGIS Online. Screenshot captured on 2025-06-25. Data © OpenStreetMap contributors, licensed under ODbL (https://opendatacommons.org/licenses/odbl/).

[25] W Tanner Porter, Lalitha Gade, Parker Montfort, Joseph R Mihaljevic, Jolene R Bowers, Andrew Willman, Brian A Klimowski, Bonnie J LaFleur, Rebecca H Sunenshine, Jennifer Collins, et al. Understanding the exposure risk of aerosolized coccidioides in a valley fever endemic metropolis. *Scientific Reports*, 14(1):1311, 2024.

[26] Patrick KR Stacy, Andrew C Comrie, and Stephen R Yool. Modeling valley fever incidence in arizona using a satellite-derived soil moisture proxy. *GIScience & remote sensing*, 49(2):299–316, 2012.

[27] M. Stuhlmacher, D. Jenerette, and A. Brazel. Normalized difference vegetation index (ndvi) data for the central arizona-phoenix lter, 2013–2023. https://portal.edirepository.org/nis/mapbrowse?scope=knb-lter-cap&identifier=720, 2019. NDVI data downloaded for the years 2013–2023 from the Environmental Data Initiative (EDI) portal. Accessed: 2025-03-13.

[28] U.S. Census Bureau. TIGER/Line Shapefile, 2019, 2010 Census Block Group National. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2019.html, 2019. Accessed: 2025-07-13.

[29] U.S. Census Bureau. TIGER/Line Shapefile, 2022, 2020 Census Block Groups, Arizona. https://www2.census.gov/geo/tiger/TIGER2022/BG/tl_2022_04_bg.zip, 2022. Accessed via ArcGIS Pro from the Maricopa Association of Governments (MAG) feature layer on 2025-07-13.

[30] U.S. Census Bureau. American community survey (acs) 5-year estimates: Table b27010 – health insurance coverage status by sex by age. Accessed via *data.census.gov*, 2023. ACS 5-Year Tables B27010 (ACS5Y2013.B27010 to ACS5Y2023.B27010), block group-level data, accessed November, 2024. See also https://data.census.gov/table/ACSDT5Y2023.B27010?q=B27010 to find and select the data from the Census Bureau website.

[31] U.S. Geological Survey. National land cover database (nlcd) land cover and change data, phoenix metro area, 2013–2023. https://www.mrlc.gov/viewer/, 2023. Data downloaded for the Phoenix metropolitan area using the MRLC Viewer. Includes NLCD Land Cover and Land Cover Change data from 2013 to 2023. Produced by the Multi-Resolution Land Characteristics (MRLC) Consortium. Accessed: 2025-06-25.

[32] Elizabeth A Weaver and Korine N Kolivras. Investigating the relationship between climate and valley fever (coccidioidomycosis). *EcoHealth*, 15:840–852, 2018.

Appendix

	(Intercept)	Female >= 45	Male >= 45	Female <45	(LCC1)	(LCC2)	(LCC3)	(LCC4)
2013 1	1.701**	1.115**	0.867*	0.189	1.735**	2.102***	2.788***	-13.83***
2013 2	-0.633	0.884**	1.135***	0.01	3.434***	4.004***	4.739***	-11.526***
2014 1	1.997***	1.242***	1.06***	-0.036	0.341	1.067**	2.11***	0.744
2014 2	1.312**	0.525	1.063***	-0.192	1.167*	2.606***	3.205***	2.04***
2015 1	1.207**	0.877***	1.145***	0.155	1.941**	1.742***	2.894***	1.632**
2015 2	2.164***	1.123***	1.298***	-0.102	0.894***	1.835***	2.759***	1.595***
2016 1	2.4***	0.775***	0.589***	-0.755***	1.493***	2.255***	3.374***	1.19***
2016 2	1.727***	1.173***	1.261***	0.36*	2.004***	2.838***	3.297***	2.384***
2017 1	2.604***	0.982***	1.118***	-0.586**	0.792**	2.055***	2.363***	1.216***
2017 2	2.382***	0.771***	0.988***	-0.497*	1.493***	2.545***	2.661***	1.664***
2018 1	1.909***	0.969***	0.9***	0.233	1.102***	2.434***	3.111***	1.363***
2018 2	3.019***	0.753***	0.848***	-0.06	-0.141	1.786***	1.806***	1.322***
2019 1	1.396***	0.931***	0.731***	-0.141	0.671*	2.624***	3.209***	1.548***
2019 2	2.406***	0.895***	0.69***	0.209	0.504	1.84***	2.785***	0.838**
2020 1	3.277***	0.413**	0.545***	-0.005	0.039	1.593***	1.919***	0.349
2020 2	3.09***	0.276**	0.377***	-0.236	0.769**	1.747***	2.417***	1.097***
2021 1	2.601***	0.814***	0.789***	-0.136	0.909***	2.502***	2.946***	2.021***
2021 2	2.818***	0.706***	0.574***	-0.114	0.85***	2.592***	2.799***	1.693***
2022 1	2.52***	0.514***	0.191	-0.136	2.078***	3.068***	3.108***	2.624***
2022 2	2.748***	0.511***	0.495***	-0.249	1.877***	2.831***	3.219***	1.639***
2023 1	1.263***	0.706***	0.568***	0.425**	2.451***	4.087***	4.559***	2.47***
2023 2	3.776***	0.138	0.397***	-0.263*	0.029	1.108***	1.731***	0.126
	LC_Change	NDVI	NDVI^2	HSI	HSI^2	NDVI_Diff	NDVI_Diff^2	
2013 1	0.681*	0.489**	-0.952***	0.517***	-0.405**	-0.077	-0.962***	
2013 2	0.842**	0.167	-0.97***	0.205	0.112	-0.038	-0.67***	
2014 1	0.47*	-0.252*	-0.542***	0.238*	0.001	-0.294**	-0.552***	
2014 2	0.354	0.422**	-0.672***	0.293*	-0.056	0.1	-0.981***	
2015 1	0.591**	0.116	-0.191***	0.075	-0.246*	-0.002	-0.107***	
2015 2	0.303*	0.247**	-0.378***	0.119	0.136*	0.077	-0.114***	
2016 1	0.825***	0.334***	-0.518***	0.426***	-0.181*	0.161	-0.432***	
2016 2	0.811***	0.267***	-0.546***	0.298***	-0.075	-0.279**	-0.028	
2017 1	0.095	-0.262**	-0.174***	0.396***	-0.151*	0.126	-0.01	
2017 2	0.309*	-0.019	-0.511***	0.349***	-0.151*	0.511***	-0.353***	
2018 1	1.245***	0.097	-0.617***	0.199***	0.241***	0.622***	-0.015	
2018 2	-0.006	-0.093	-0.457***	0.235***	-0.112	-1.128***	-0.313***	
2019 1	0.758***	-0.589***	-0.36***	0.609***	-0.119	-1.019***	-0.643***	
2019 2	0.32**	-0.175*	-0.508***	0.439***	0.029	0.624***	-0.116***	
2020 1	0.397**	-0.369***	-0.371***	0.306***	-0.009	-0.987***	-0.685***	
2020 2	0.349***	0.054	-0.467***	0.282***	0.062	0.966***	-0.326***	
2021 1	0.593***	0.024	-0.49***	0.193***	0.097*	0.459***	-0.037***	
2021 2	0.222*	0.062	-0.396***	0.252***	-0.045	-0.59***	-0.3***	
2022 1	0.501***	0.232***	-0.469***	0.178**	0.056	0.677***	-0.375***	
2022 2	0.487***	0.203**	-0.501***	0.1	-0.139*	-0.692***	-0.36***	
2023 1	0.413***	-0.113	-0.418***	0.196***	0.028	0.39***	-0.086***	
2023 2	0.162	0.119	-0.441***	0.213***	0.03	1.176***	-0.526***	

Table 3
Model (4) results.

	(Intercept)	Female >= 45	Male >= 45	Female <45	(LCC1)	(LCC2)
2013 1	0.376	1.002**	0.831*	-0.126	1.872***	2.607***
2013 2	-1.408**	1.207***	1.139***	-0.028	3.355***	4.152***
2014 1	0.587	1.481***	1.426***	0.273	0.909*	2.039***
2014 2	0.472	0.753**	1.227***	0.021	0.786	2.639***
2015 1	0.311	0.837***	0.874***	0.125	2.088***	2.744***
2015 2	1.538***	1.109***	1.25***	-0.456*	0.837**	2.402***
2016 1	1.799***	0.876***	0.708***	-0.199	1.004***	2.476***
2016 2	1.302***	0.968***	0.993***	0.253	1.67***	3.223***
2017 1	2.178***	1.094***	1.173***	-0.205	0.663**	2.184***
2017 2	1.993***	0.789***	0.949***	-0.287	0.904***	2.568***
2018 1	2.001***	1.131***	0.923***	0.256	0.182	2.163***
2018 2	2.154***	0.737***	0.982***	-0.13	0.161	2.462***
2019 1	1.15***	0.857***	0.772***	0.194	0.914***	2.933***
2019 2	2.13***	0.592***	0.39**	-0.353**	0.575*	2.554***
2020 1	2.728***	0.75***	0.714***	0.199	0.395	1.812***
2020 2	2.532***	0.332**	0.36**	-0.223	1.043***	2.234***
2021 1	2.784***	0.577***	0.652***	-0.15	0.287	2.211***
2021 2	2.776***	0.712***	0.655***	-0.224	0.237	2.446***
2022 1	1.602***	0.676***	0.4***	-0.123	1.454***	3.67***
2022 2	1.718***	0.527***	0.536***	-0.186	1.745***	3.243***
2023 1	1.356***	0.624***	0.576***	0.08	2.034***	3.657***
2023 2	2.772***	0.464***	0.671***	-0.171	0.374	2.21***
	(LCC3)	(LCC4)	LC_Change	NDVI	HSI	NDVI_Diff
2013 1	3.437***	-13.187***	0.489*	0.059	0.459***	0.245
2013 2	4.89***	-12.487***	0.64**	-0.112	0.093	-0.17
2014 1	3.098***	1.687***	0.256	-0.336***	0.217	0.235*
2014 2	3.492***	2.198***	0.17	-0.049	0.239*	-0.208
2015 1	3.682***	2.015***	0.436*	0.099	0.145	0.002
2015 2	3.296***	1.926***	0.627***	-0.171**	0.147**	-0.063
2016 1	3.387***	1.73***	0.24	-0.116	0.308***	0.234***
2016 2	3.698***	2.512***	0.648***	-0.147*	0.289***	-0.477***
2017 1	2.555***	1.283***	0.067	-0.223***	0.26***	0.147
2017 2	2.931***	1.643***	0.184	-0.29***	0.262***	-0.043
2018 1	2.911***	1.433***	0.61***	-0.318***	0.327***	0.297***
2018 2	2.588***	1.254***	0.062	-0.236***	0.364***	-0.39***
2019 1	3.689***	1.567***	0.385***	-0.339***	0.298***	-0.005
2019 2	3.106***	0.984***	0.272*	-0.348***	0.372***	0.35***
2020 1	2.26***	0.211	0.387***	-0.128*	0.213***	-0.103
2020 2	2.843***	1.695***	0.44***	-0.219***	0.342***	0.281***
2021 1	2.702***	1.809***	0.408***	-0.171**	0.198***	0.173**
2021 2	2.514***	1.256***	0.178	-0.38***	0.222***	-0.082
2022 1	3.648***	2.537***	0.376***	-0.124*	0.279***	0.447***
2022 2	3.668***	2.305***	0.381**	-0.222***	0.074	-0.33***
2023 1	4.174***	2.291***	0.26*	-0.125*	0.205**	0.345***
2023 2	2.63***	0.752**	0.138	-0.191***	0.106*	0.169*

Table 4
Model (3) results.

	(Intercept)	Female >= 45	Male >= 45	Female <45	(LCC1)	(LCC2)
2013 1	0.283	1.198***	1.059***	0.175	1.314**	2.269***
2013 2	-1.131**	1.024***	1.168***	0.114	3.234***	3.702***
2014 1	0.669	1.142***	1.196***	0.007	0.855**	1.983***
2014 2	0.265	0.592	1.121***	0.006	0.907*	2.977***
2015 1	0.683*	0.934***	1.009***	0.007	1.853***	2.069***
2015 2	1.664***	1.086***	1.143***	-0.068	1.072***	2.113***
2016 1	1.261***	1.055***	0.787***	-0.19	1.424***	2.953***
2016 2	1.536***	0.831***	1.004***	0.099	1.349***	2.626***
2017 1	2.214***	0.914***	1.044***	-0.328	0.704**	2.207***
2017 2	1.907***	0.566***	0.655***	-0.645***	1.475***	2.744***
2018 1	1.888***	1.272***	0.985***	0.324*	0.755**	2.177***
2018 2	2.384***	0.728***	0.822***	-0.211	0.682**	2.057***
2019 1	0.976***	0.754***	0.855***	-0.283	1.667***	3.054***
2019 2	1.879***	0.66***	0.608***	0.072	1.269***	2.625***
2020 1	2.426***	0.667***	0.754***	0.309*	0.799***	2.082***
2020 2	2.617***	0.436***	0.375***	-0.277*	1.305***	2.319***
2021 1	2.359***	0.566***	0.593***	-0.12	0.963***	2.499***
2021 2	2.925***	0.646***	0.697***	-0.114	0.261	1.91***
2022 1	1.71***	0.619***	0.464***	-0.139	1.643***	3.172***
2022 2	2.114***	0.652***	0.503***	-0.077	1.635***	2.621***
2023 1	1.219***	0.646***	0.303**	-0.113	2.124***	3.65***
2023 2	2.573***	0.404***	0.641***	-0.106	0.721***	2.361***
	(LCC3)	(LCC4)	LC_Change			
2013 1	3.113***	-13.642***	0.434			
2013 2	4.561***	-12.028***	0.581**			
2014 1	3.232***	1.132**	0.8***			
2014 2	3.76***	1.941***	0.51*			
2015 1	3.005***	1.396**	0.486*			
2015 2	3.059***	1.186***	0.472***			
2016 1	3.839***	1.831***	0.306*			
2016 2	3.304***	1.667***	0.652***			
2017 1	2.619***	1.303***	0.021			
2017 2	3.265***	1.492***	0.408***			
2018 1	2.854***	0.71**	0.67***			
2018 2	2.54***	0.903**	0.325**			
2019 1	4.07***	1.502***	0.821***			
2019 2	3.459***	1.258***	0.378***			
2020 1	2.782***	0.052	0.449***			
2020 2	2.931***	1.59***	0.426***			
2021 1	3.081***	1.507***	0.625***			
2021 2	2.409***	0.989***	0.274**			
2022 1	3.586***	2.433***	0.354**			
2022 2	3.181***	1.241***	0.388***			
2023 1	4.219***	2.314***	0.516***			
2023 2	2.98***	1.092***	0.304**			

Table 5
Model (2) results.

	(Intercept)	Female >= 45	Male >= 45	Female <45
2013 1	2.817***	1.226***	1.033**	0.075
2013 2	2.957***	1.01***	1.095***	0.048
2014 1	3.399***	1.049***	1.069***	-0.109
2014 2	3.21***	0.754**	1.239***	0.038
2015 1	3.338***	0.8**	0.997***	-0.042
2015 2	4.135***	1.218***	1.278***	-0.164
2016 1	4.428***	1.004***	0.795***	-0.146
2016 2	4.286***	0.971***	1.024***	0.1
2017 1	4.32***	1.015***	1.081***	-0.333
2017 2	4.48***	0.794***	0.885***	-0.392*
2018 1	4.367***	1.06***	0.941***	0.136
2018 2	4.43***	0.765***	0.842***	-0.124
2019 1	4.4***	0.822***	0.843***	0.004
2019 2	4.78***	0.712***	0.595***	-0.071
2020 1	4.651***	0.628***	0.665***	0.152
2020 2	5.063***	0.418***	0.442***	-0.221
2021 1	5.008***	0.611***	0.571***	-0.165
2021 2	4.843***	0.703***	0.727***	0.025
2022 1	4.887***	0.594***	0.415***	-0.161
2022 2	4.816***	0.649***	0.573***	-0.034
2023 1	4.8***	0.688***	0.54***	0.065
2023 2	5.102***	0.295*	0.529***	-0.192

Table 6
Model (1) results.

	Deviance	P-value	Significance
2013 1	-94.4986	2.18E-15	***
2013 2	-98.03475	4.37E-16	***
2014 1	-93.29907	3.75E-15	***
2014 2	-95.44943	1.42E-15	***
2015 1	-51.65486	3.15E-07	***
2015 2	-188.19767	2.07E-34	***
2016 1	-236.22911	2.12E-44	***
2016 2	-201.53942	3.56E-37	***
2017 1	-139.82335	1.77E-24	***
2017 2	-195.77712	5.59E-36	***
2018 1	-243.02102	8.07E-46	***
2018 2	-205.57273	5.18E-38	***
2019 1	-241.92584	1.37E-45	***
2019 2	-289.83421	1.21E-55	***
2020 1	-216.34161	2.98E-40	***
2020 2	-278.49724	2.93E-53	***
2021 1	-247.71976	8.39E-47	***
2021 2	-233.3544	8.46E-44	***
2022 1	-231.40941	2.15E-43	***
2022 2	-268.55572	3.60E-51	***
2023 1	-247.12996	1.11E-46	***
2023 2	-273.39994	3.46E-52	***

Table 7
Likelihood ratio test results comparing models (4) and (1) for data in all semesters.