Published in final edited form as:

Comput Stat Data Anal. 2009 June 15; 53(8): 3001–3015. doi:10.1016/j.csda.2008.10.013.

Spatiotemporal and Spatial Threshold Models for Relating UV Exposures and Skin Cancer in the Central United States

Laura A. Hatfield, Richard W. Hoffbeck, Bruce H. Alexander, and Bradley P. Carlin

Laura A. Hatfield is graduate assistant and Bradley P. Carlin is Mayo Professor in Public Health in the Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, 55455. Richard W. Hoffbeck is Research Associate and Bruce H. Alexander is Associate Professor in the Division of Environmental Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55455

Abstract

The exact mechanisms relating exposure to ultraviolet (UV) radiation and elevated risk of skin cancer remain the subject of debate. For example, there is disagreement on whether the main risk factor is duration of the exposure, its intensity, or some combination of both. There is also uncertainty regarding the form of the dose-response curve, with many authors believing only exposures exceeding a given (but unknown) threshold are important. In this paper we explore methods to estimate such thresholds using hierarchical spatial logistic models based on a sample of a cohort of x-ray technologists for whom we have self-reports of time spent in the sun and numbers of blistering sunburns in childhood. A preliminary goal is to explore the temporal pattern of UV exposure and its gradient. Changes here would imply that identical exposure self-reports from different calendar years may correspond to differing cancer risks.

Keywords

Conditionally autoregressive (CAR) model; Erythematous exposure; Hierarchical model; Non-melanoma skin cancer

1 Introduction

Non-melanoma skin cancer (NMSC) is the most common cancer in the US, with estimated annual incidence over one million cases [2]. Comprised primarily of basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) in a ratio of 4:1, surveillance of these cancers is not undertaken nationally, because of the low morbidity and mortality compared to other cancers. Estimates of age-adjusted incidence for BCC are 407 per 100,000 in white men and 212 per 100,000 in white women. For SCC, the estimated annual rate is 100 to 150 per 100,000 in whites [32].

Exposure to ultraviolet (UV) radiation is associated with elevated risk of skin cancer. However, the exact mechanisms relating this exposure to the adverse health outcomes remain unclear. Early epidemiological evidence related skin cancer rates to ground-based measurements of UV

Correspondence author: Bradley P. Carlin, telephone: (612) 624-6646, fax: (612) 626-0660, email: brad@biostat.umn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

radiation [41]. In the earliest years of ground-based UV monitoring, no increase in exposure was measured, but in recent years, changes in ozone and UV exposure have been noted [1, 40].

Clear relationships exist between UV exposure and benign sun-related changes, including sunburn (erythema), freckling in childhood, moles (melanocytic naevi), and solar keratoses [7,14,18,25]. Exposure to sunlight interacts with individual characteristics related to sun sensitivity. Common markers are ethnic origin, skin pigmentation, eye color, hair color, and propensity to sunburn [4,21,33].

Different types of exposure are associated with risk of melanoma, BCC, and SCC. Chronic, high-level UV exposure is often associated with SCC, while both short-term burning episodes and chronic exposure contribute to BCC risk [3,20,24,26,²⁷,29,38,46,49]. However, one study showed that very high UV doses (measured at ground-based stations) were associated with greater risk of SCC versus BCC [37]. Another study using multivariate models showed SCC was associated with lifetime blistering sunburns and cumulative recreational sun exposure, while BCC was associated with cumulative exposure only [21].

UV radiation is a complex exposure variable. In addition to differences by geography, behavior, age, ethnicity, and occupation [13,15,19,36,45], methods for direct quantification of UV dose vary. Ground-based UV measurements are systematically 15-20% lower than satellite-based values during summer months [22,39]. The biologically effective spectrum does not vary substantially over the times during the day when sunburn may occur [23]. Fortunately, study participants generally provide reliable information on sun sensitivity and history of skin lesions [11].

There is disagreement concerning the relevant *ages* of exposure. One study found that BCC risk was associated with adolescent and childhood exposure, but inversely associated with lifetime recreational exposure [18]. In that same study, risk of SCC was unassociated with lifetime exposure, but a trend for greater risk with occupational sun exposure in the 10 years prior to diagnosis was observed [17]. The U.S. Radiologic Technologists (USRT) cohort showed that annual mean residential UV radiation exposure during adulthood (but not childhood) was associated with increased risk of SCC and BCC [48].

The precise form of the relationship between exposure and skin neoplasia is not known. A logistic model, which proposes a linear relationship between the logit of the adverse event probabilities and the predictor variables of interest, is typically used and often produces reasonably good agreement with other functional forms (e.g., the probit or the complementary log-log in place of the logit). However, a more difficult issue is the precise input to this logistic model. Some authors have suggested there may be a *threshold* below which no response is observed [38]. However, few studies have investigated UV dose thresholds for carcinogenesis in humans [35], though experimentally accessible targets such as fibroblasts, animal models, and skin erythema have been investigated [12,30,31]. Such a threshold may be estimated from the data, but estimation is difficult using standard statistical approaches, since it involves estimating a component of the design matrix, which in standard methods is considered fixed and known.

The models we employ have natural hierarchical structure, resulting from both spatial and temporal indexing in our data. As a result, we adopt a hierarchical Bayesian analytic strategy, in order to more easily borrow estimative power across the large number of random effects we employ. Longitudinal structure is handled using random slope and intercept effects [28], while spatial structure is accommodated using the conditionally autoregressive (CAR) models of Besag [8]. While computationally intensive due to use of Markov chain Monte Carlo (MCMC) methods (see e.g. [10, Sec. 3.4]), all of the models we employ are easily fit using the WinBUGS

software package [44], freely available from the website www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.

In this paper, we explore models to estimate the dose-response relationship between UV exposure and skin cancer outcome, allowing for threshold estimation. We do this in the context of a logistic model relating spatially referenced UV measurements to self-reports of time spent in the sun and number of blistering sunburns in childhood. We use a sample of data from the United States Radiologic Technologists (USRT) cohort (<http://www.radtechstudy.org/>). The USRT is a nationwide cohort that includes all radiologic technologists registered with the American Registry of Radiologic Technologists for at least two years prior to 1982 [9]. Beginning in 1983, members of this cohort have been surveyed to collect information on cancer and other health outcomes, occupational histories, medical histories, and lifestyle habits. The most recent survey (2004) also ascertained a brief residential history and risk factors for skin cancer, including, hair and eye color, history of blistering sunburns, and self-reported time outdoors in the summer. Our UV dose measures come from NASA's Total Ozone Mapping Spectrometer (TOMS) project over a regular latitude/longitude-based grid [22]. These data are freely available (http://toms.gsfc.nasa.gov/ftpdata_v8.html). In order to avoid awkward spatial misalignment in our data [5, Ch. 6], we aggregate the USRT data to the same geographic grid used by the TOMS study.

A preliminary goal of our analysis is to explore the pattern of UV exposure over time. Statistically, our interest here lies both in the pattern itself (which need not be linear) and in rates of change in this pattern, especially when such change is sudden. Substantively, the question is important since, if UV exposure is increasing over time, identical self-reports of exposure time from widely differing years (say those exposed in the 1940s versus those exposed today) may actually correspond to different cancer risks.

The remainder of our paper proceeds as follows. We begin in Section 2 with an analysis of the spatiotemporal TOMS exposure data alone, thus addressing our preliminary analytic goal above. We find evidence that exposure is increasing over time and isolate geographic regions of higher and lower exposure. Next, Section 3 incorporates the NMSC outcome data from the USRT cohort, and fits various standard models for the impact of UV dose. Section 4 also models dose-response relationships, but this time using peak exposure rather than average. Our results indicate that UV dose is increasing over time, but that these increases are not varying geographically across the study region. Further, using a constant, spatially referenced estimate of UV dose improves prediction of NMSC outcomes. Time spent outdoors is best added as a separate covariate, and such models provide additional predictive value above that provided by models using only skin sensitivity markers and reports of severe sunburns. Finally, Section 5 summarizes and offers direction for future substantive and statistical research in this important area. WinBUGS code to fit the various models is provided in the Appendix.

2 Spatiotemporal Modeling of UV Erythral Dose

Exposure to biologically damaging UV radiation varies according to latitude, meteorological conditions, and time of year. We utilize an erythral exposure dose product Y , measured in milli-watts/ m^2 , that weights UV radiation reaching the surface according to a model of Caucasian sensitivity to sunburn [22,31],

$$Y = \frac{1}{d_{es}^2} \int_{280nm}^{400nm} S(\lambda) A(\lambda) \int_{t_{sr}}^{t_{ss}} C(\lambda, \vartheta, \tau_{cl}) F(\lambda, \vartheta, \omega) dt d\lambda. \quad (1)$$

In this complex equation, d_{es} is the distance (in A.U.) from the earth to the sun, S is the solar irradiance incident on the top of the atmosphere when the earth is 1 A.U. from the sun [47], A is the McKinley-Diffey action spectrum,

$$A(\lambda) = \begin{cases} 1 & , \lambda < 298 \\ 10^{-0.094(\lambda-298)} & , 298 \leq \lambda \leq 328 \\ 10^{-0.015(\lambda-139)} & , 328 \leq \lambda \end{cases}$$

t_{sr} and t_{ss} are the times of sunrise and sunset, C is a cloud attenuation factor, τ_{cl} is cloud optical thickness, v is the solar zenith angle (function of time t), F is the spectral irradiance (direct plus diffuse) at the surface under clear skies, normalized (calculated using a table of solutions of the radiative transfer equation, which are used in the TOMS ozone retrievals), and Ω is the total column ozone, measured by the TOMS instrument.

We remark that τ_{cl} and ω are modeled values based on measured inputs, while v is a function of time, and so gets integrated out in (1). Estimation of τ_{cl} and ω is beyond the scope of our study; the interested reader is referred to the online notes of Herman and Celarier (<http://macuv.gsfc.nasa.gov/doc/erynotes.pdf>) for full explanation of model (1) and its components $S(\lambda)$, $C(\lambda, v, \tau_{cl})$, and $F(\lambda, v, \omega)$. After the integrations in (1), Y emerges as a function of total column ozone (ω) and a model of cloud attenuation ($C(\tau_{cl})$) that depends on physical constants and measured backscattered radiance. Each daily value may be interpreted as an index of the potential for skin damage, given the (measured) ozone and cloud conditions and the (modeled) biological action spectrum. We do not partition the variability in these measurements due to random versus modeling error; both are represented in the overall σ^2 variance terms in the models below.

We consider the analysis of approximately 150 days of UV dose data from May 1 to September 30 during the years 1979 to 2003 (excluding 1993-1996 when the satellite was offline). The geographic region (left panel of Figure 1) includes 81 grid locations arranged in a 9×9 grid spanning -78.75° to -90° longitude and 34° to 43° latitude (each grid box being 1° by 1.25°). We selected this region since it was among the largest square regions that had no subregions containing unacceptably few participants, yet still included a wide range of summer UV exposures. Plots of the raw data (right panel) reveal clear spatial and temporal patterns, with increasing UV exposure over time and in more southern latitudes.

2.1 Model Comparison

Let Y_{ij} be the average UV exposure in the i^{th} grid location during the j^{th} year. We assume $Y_{ij} \sim N(\mu_{ij}, \sigma^2)$, where the variance component σ^2 soaks up excess variability arising from uncertainty in the input parameters in the aforementioned erythral exposure model for Y , as well as ordinary statistical observation error. The mean structure μ_{ij} follows one of four different models: a simple linear regression model (Model 2), random effects longitudinal model (Model 3), the same adjusted for latitude (Model 4), and the same with a spatially-structured prior on random intercepts (Model 5). We also use two covariates: year (W_1 , centered around 1991 so that $W_1 \in \{-12, \dots, 12\}$), and latitude (W_2 , centered around 38.5° so that $W_2 \in (-4.5, 4.5)$).

In Model 2, we model each grid location separately, without a latitude covariate, as

$$\mu_{ij} = \beta_{0i} + \beta_{1i} W_{1j}. \quad (2)$$

We then place a noninformative $U(0.01, 100)$ prior on σ , and flat priors on the β_{0i} and β_{1i} so that these parameters are treated as fixed effects in the model.

In Model 3 we reparameterize using a grand mean and slope, μ_0 and μ_1 , and random intercept and slope effects, γ_{0i} and γ_{1i} , obtaining

$$\mu_{ij} = \mu_0 + \gamma_{0i} + (\mu_1 + \gamma_{1i})W_{1j}. \quad (3)$$

We place vague normal priors on the grand intercept and time slope, μ_0 and μ_1 , but potentially informative normal mixing distributions, $\gamma_{0i} \stackrel{iid}{\sim} N(0, \sigma_0^2)$ and $\gamma_{1i} \stackrel{iid}{\sim} N(0, \sigma_1^2)$, on the random effects. The new standard deviation hyperparameters, σ_0 and σ_1 , are now assigned $U(0.01, 100)$ priors, thus allowing the data to determine the amount of similarity across grid squares.

In Model 4, we “detrend” the observed north-south pattern in intercepts (γ_{0i}) by adding the latitude variable (W_2), and place an improper flat prior on its coefficient, α . The mean structure thus changes again to

$$\mu_{ij} = \mu_0 + \gamma_{0i} + (\mu_1 + \gamma_{1i})W_{1j} + \alpha W_{2i}. \quad (4)$$

Finally, in Model 5, we allow *spatial* shrinkage in the random intercepts by placing a spatially structured prior on $\gamma_0 = (\gamma_{01}, \dots, \gamma_{0,81})^T$. The simplest such model for our areal data is the *conditionally autoregressive* (CAR) model [8], which has full conditional distributions

$$\gamma_{0i} | \gamma_{0,j \neq i} \sim \text{Normal}\left(\bar{\gamma}_{0i}, \frac{1}{\tau m_i}\right),$$

where $\bar{\gamma}_{0i} = \frac{1}{m_i} \sum_{j \neq i} \gamma_{0j}$ is the average of the γ_{0j} from regions j that neighbor region i , and m_i is the number of these neighbors.

The random slopes from Models 3 and 4 emerge as statistically insignificant, suggesting we may drop the γ_{1i} parameters. This then implies that the temporal gradient in UV does not change with location i . Retaining the priors on μ_0 , μ_1 , and α described above, we obtain the model

$$\mu_{ij} = \mu_0 + \gamma_{0i} + \mu_1 W_{1j} + \alpha W_{2i} \text{ where } \gamma \sim \text{CAR}(\tau). \quad (5)$$

Here we set $\tau = 1/\sigma_0^2$ and use a vague $\text{Unif}(0.01, 100)$ hyperprior for σ_0 .

We use the Deviance Information Criterion (DIC; [43]) and mean squared error of prediction (MSEP) criteria to select the best model from among those just described. A hierarchical model generalization of the Akaike information criterion (AIC), DIC is a sum of the posterior mean deviance score ($\bar{D}$) that is small for good-fitting models, and an “effective number of parameters” (p_D) that is small when a model’s random effects shrink more completely back toward their prior mean (in this case, zero). Our simultaneous desire for good fit and parsimony implies models with small DIC are preferred.

The predictive approach leaves out either one year's data across all grids, Y_{ij} , $i = 1, \dots, 81$ for $j = 1$ (1979) or $j = 14$ (1992), or three grids' data across all years, Y_{ij} , $j = 1, \dots, 21$ for $i = 21$, 36, or 68; again see the left panel of Figure 1. Though prediction of the last year of data (2003) may seem natural, there is some indication that this point is an outlier (or an early sign of an sharp upward trend in subsequent years), so since the goal is overall model choice we elect to base our predictive checks on more "typical" years. In symbols, MSEP is computed as

$$MSEP = \frac{1}{81} \sum_{i=1}^{81} (Y_{ij} - \hat{Y}_{ij})^2 \text{ or } \frac{1}{21 * 3} \sum_{i \in \{21, 36, 68\}} \sum_{j=1}^{21} (Y_{ij} - \hat{Y}_{ij})^2,$$

where $j = 1$ or 14 in the first (temporal prediction) expression, and $\hat{Y}_{ij}$ is the posterior mean for the $(i, j)^{th}$ omitted point, given the data from the other 20 years (or 78 grids, in the case of the second, spatial prediction expression).

Table 1 shows DIC values and MSEP posterior means for each of the models above, with the exception of Model 2, where prediction is not possible when omitting all years of data for any grid. Each progressive model enhancement produces a drop in DIC. Detrending with latitude (Model 3 to Model 4) did not improve fit ($\bar{D}$) but did decrease p_D , apparently since adding this important covariate enabled more shrinkage in the random effects. Finally, removing random slopes and placing a CAR prior on the random intercepts (Model 5) lowered DIC further, by both improving fit ($\bar{D}$) and decreasing effective size (p_D). The MSEP results do not display the same stepwise progression, but in each case, the "best" model is 5. The results are most dramatic for grid prediction, where the addition of latitude dramatically improves prediction and the CAR prior provides another large drop in MSEP.

2.2 Parameter Estimation

We present posterior parameter estimates for the best model, Model 5, in Table 2. In this and subsequent results, we report posterior density summary statistics based on 5000 Gibbs samples from two chains (after a suitable burn-in period). The estimated gradient over time (μ_1) is significantly positive, as expected based on examination of Figure 1(b). Average UV exposure is estimated to have increased by 0.59 units per year over our observation period. On the familiar 1-to-11+ "UV Index" scale, this translates into an annual change of $.59/25 = .0236$ units, a fairly small amount (see Section 4 below for further discussion of the UV Index). The negative estimate for the latitude coefficient (α) agrees with our visual assessment that higher latitude (i.e., more northerly location) is associated with lower UV exposure.

Regarding the random effects, Figure 2(a) maps the posterior means of the γ_{0i} . There appears to be an interesting decrease in baseline exposure in western Pennsylvania and northern West Virginia, and some higher baseline exposures toward the western edge of the map. Panel (b) of the figure shows that these parameters are significantly different from zero for several grid squares. Moreover, the scale of these residuals is large enough to motivate a search for missing spatially-varying covariates. Here, we speculate that the spatial pattern in Figure 2(a) may be the result of the cloud cover typical across the region.

3 Modeling Cancer Outcomes using UV Exposure Measures

We now consider UV exposure as a predictor of self-reported nonmalignant skin cancer diagnoses in the US Radiological Technologists (USRT) cohort [9]. The present subsample comprises 19,081 people who resided primarily in one of the 81 grid locations while under 13 years of age.

The outcome of interest is self-reported, cumulative, incident diagnoses of non-melanoma skin cancer (NMSC). We thus define a new variable, Z_k , to be 1 for those participants k who reported a diagnosis of squamous cell carcinoma (SCC) or basal cell carcinoma (BCC), and 0 otherwise. Total number of cases in the sample were 383 (2.0%) for SCC, 1283 (6.7%) for BCC, and 1495 (7.8%) for the combined NMSC variable (i.e., 171 people reported both cancers). Using the participants' grid squares of residence when they were under age 13, Figure 3 gives choropleth maps of the numbers of people living in each grid square, the number of NMSC cases, and their proportion (i.e., cumulative incidence). The major metropolitan areas in the region (Chicago, Cincinnati, Detroit, Cleveland, and Pittsburgh) emerge readily in the first two maps, but the incidence map is reasonably flat over the spatial domain.

3.1 Model Comparison

To incorporate the cancer outcomes Z_k into our Section 2 spatiotemporal modeling of the UV doses Y_{ij} , we now develop a two-stage hierarchical model. Here, random intercepts estimated in the first stage (which models UV dose by grid location over time) enter as exposure variables at the second stage (which models cancer outcomes by individual). We use Model 5 as the first stage of our subsequent models, with its parameters entering the second stage as one of several predictors of the binary cancer outcome.

Note that the exposure trajectories were seen in the previous section not to cross over time ($\gamma_{1i} = 0$); the ordering of the grid locations does not depend on year of exposure. As such, assuming that the age distribution in the USRT cohort is reasonably stable across the spatial domain, we are free to take fitted exposure at any point in time for each grid square. A convenient choice is the fitted grid-specific intercept, $\gamma_{0s_k} + \alpha W_{2s_k}$, where s_k is the grid location in which the k^{th} participant lived while under age 13, for $s_k \in \{1, \dots, 81\}$. These fitted values fall in the range $(-22.4, 24.2)$. Note that this definition of exposure accounts for both latitude and any necessary adjustment provided by the spatial residuals $\{\gamma_{0i}\}$. Our full model now takes the form:

Stage 1:

$$Y_{ij} \sim \text{Normal}(\mu_{ij}, \sigma^2), \quad i=1, \dots, 81, \quad j=1, \dots, 21$$

where

$$\mu_{ij} = \mu_0 + \gamma_{0i} + \mu_1 W_{1j} + \alpha W_{2i}, \quad \gamma_{0i} \stackrel{iid}{\sim} \text{CAR}(\tau_0);$$

Stage 2:

$$Z_k \sim \text{Bernoulli}(p_k), \quad k=1, \dots, 19,081$$

where

$$\begin{aligned} \text{logit}(p_k) &= \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \beta_4 X_{4k} \\ &\quad + \beta_5 (\gamma_{0s_k} + \alpha W_{2s_k}) I_{(\text{Model 6 or 7})} + \beta_6 X_{6k} I_{(\text{Model 7})} \\ &\quad + \beta_5 (\gamma_{0s_k} + \alpha W_{2s_k}) X_{6k} I_{(\text{Model 8})}. \end{aligned}$$

Here, X_1 = age (centered around its mean of 55.8, and taking values in $(-12.8, 38.2)$), X_2 = sex (1 = female, -1 = male), X_3 = number of blistering sunburns (centered around its mean of 4.3, and taking values in $(-4.3, 194)$), X_4 = complexion (1 = light, -1 = otherwise), and X_6 = standardized cumulative hours spent outside during summers under age 13. We complete the specification with vague normal priors on the β and μ parameters, and $U(0.01, 100)$ priors on the σ parameters.

The structure of our two-stage model does allow learning about the p_k from the exposure data Y_{ij} as well as the outcomes Z_k , but because the Y s are much farther from the p parameters in the graphical model, this learning should be minimal and mostly confined to increasing the posterior variance of the β s. Our empirical examinations support this: we compared the posterior distributions of the NMSC risk parameters (the β s) for the one- and two-stage models. In the former (see Table 3), we simply imputed modeled doses as fixed and known exposures in the NMSC logistic model. In the two-stage model (see Table 5 below), the posterior variance of β_5 is properly inflated to account for our having modeled its covariate values. Analogous comparisons of UV parameters estimated individually (as in Section 2 above), or as the first stage of a two-stage model (as in this section) also yielded very similar posteriors (compare e.g. Tables 5 and 2).

Figure 4 illustrates similar robustness for the γ_{0i} random intercepts. Horizontal grey lines mark posterior mean, 2.5, and 97.5 percentiles of γ_{0i} for six selected grid squares in the one-stage model, while the black bars show these same statistics for posteriors in the first stage of three two-stage models (described in detail below). It is clear that the UV random effects are not significantly affected by the NMSC portion of the model. Together, these findings indicate that Bayesian learning about the UV parameters from the NMSC data, and about the outcome parameters from the exposure data, is as minimal as expected.

Notice the two-stage model expression above contains several indicator functions I to provide various exposure modeling options. For instance, Model 6 does not include X_6 , each individual's self-reported time spent outdoors under age 13. Thus the UV exposure term is simply

$$\beta_5(\gamma_{0s_k} + \alpha W_{2s_k}). \quad (6)$$

By contrast, Model 7 does include a separate time spent outdoors covariate. Each individual's value on this variable is the cumulative number of hours spent outside during summers under age 13, standardized so that $X_{6k} \in (-2.0, 1.3)$. Hence the exposure terms are now

$$\beta_5(\gamma_{0s_k} + \alpha W_{2s_k}) + \beta_6 X_{6k}. \quad (7)$$

Finally, Model 8 is an extension of Model 6, in which we weight the spatially shrunk UV erythral dose estimate for the k^{th} person's childhood grid of residence by his or her time outdoors value (X_{6k} , scaled to be between 0 and 1). Thus, the coefficient β_5 indicates the effect of *time-weighted* erythral UV dose, and the exposure term is therefore

$$\beta_5(\gamma_{0s_k} + \alpha W_{2s_k}) X_{6k}. \quad (8)$$

To compare the models for fit, complexity, and predictive ability, Table 4 presents $\bar{D}$, p_D , and DIC values, as well as a log-predictive marginal likelihood (LPML) scoring rule for a random 10% hold-out sample. For the latter predictive check, we randomly delete Z_k for $k = 1, \dots, 1908$ (10% of the participants) and check overall predictive performance via the log-Bernoulli likelihood-based score,

$$LPML = \frac{1}{1908} \sum_{k=1}^{1908} [Z_k \log(\widehat{p}_k) + (1 - Z_k) \log(1 - \widehat{p}_k)],$$

where the hat denotes the posterior mean given the 90% subset of the data used for fitting. Higher scores indicate better predictive performance. For the DIC and related statistics, we include only the deviance contributions from Stage 2 (Z) of the models, since Stage 1 (Y) was unchanged across models. DIC was best when time outdoors was either not included (Model 6) or added separately (Model 7); time-weighting (Model 8) worsened fit and DIC score, though it did save one effective parameter over Model 7, as expected. The LPML values are very similar but also support this ordering of the models, with Model 7 predicting best, followed by 6 and then 8.

3.2 Parameter Estimation

Table 5 presents posterior parameter estimates for Model 7. Significant covariates include age ($\beta_1 = 0.054$), blistering sunburns ($\beta_3 = 0.022$), and light complexion ($\beta_4 = 0.328$), all of which are positively associated with risk of NMSC. The coefficient for the fitted UV exposure term (β_5) is also significant and positive. The posterior 95% credible interval around the time-outdoors parameter (β_6) does not quite exclude zero. However, the posterior mean ($\beta_6 = 0.046$) indicates a trend toward association with higher risk of NMSC. Moreover, β_5 and β_6 have posterior correlation of just 0.034, indicating these two exposure variables work not only separately, but essentially *independently* to determine NMSC risk.

4 Peak Exposure by UV Index

As mentioned in Section 1, previous researchers have argued that only UV exposures over a certain threshold are important in predicting eventual skin cancer outcomes. As such, we next wish to investigate a measure of “peak” UV exposure, in contrast to the average values used above. To facilitate this, we first transform our erythral dose values to the scale of the *UV Index*, a commonly used measure of UV irradiation reaching the Earth’s surface that is weighted according to the McKinlay-Diffey Erythema action spectrum [31]. To calculate the UV Index, our daily noontime UV erythral dose values were simply divided by 25 and the results rounded up to the nearest integer values; we used the aforementioned integer categories, namely 1, 2, ..., 10, 11+ [34]. Then, for each grid square i , we computed the proportion of summer days above a given UV Index threshold C in each summer. Finally, we averaged these proportions over all years (1979 to 2003, excluding 1993-1996 when the satellite was offline) and standardized to obtain our new peak exposure covariate $X_5(i, C)$. The values of C that yielded a reasonable amount of variation in X_5 among the grids squares were 4 to 11, inclusive. Thus, we limit the possible values of C to these.

4.1 Model Comparison

With just one exposure value per grid square, we do not attempt (spatio)temporal modeling, thus eliminating the first stage of our Section 3 models. Instead, in the models of this section, the historical proportion of summer days at or above the given UV Index threshold C enters the cancer outcome model directly. The resulting model is as follows:

$$Z_k \sim \text{Bernoulli}(p_k), k=1, \dots, 19,081$$

where

$$\begin{aligned} \text{logit}(p_k) = & \beta_0 + \beta_1 X_{1k} + \beta_2 X_{2k} + \beta_3 X_{3k} + \beta_4 X_{4k} \\ & + \beta_5 X_5(s_k, C) I_{(\text{Model } 9 \text{ or } 10)} \\ & + \beta_6 X_{6k} I_{(\text{Model } 10)} + \beta_5 X_5(s_k, C) X_{6k} I_{(\text{Model } 11)}, \end{aligned}$$

and $C \sim \text{Discrete Uniform}(4, 6, \dots, 11)$. That is, we allow the UV Index threshold C to range over integer values from 4 to 11, but do not favor any of these values *a priori*. Thus the exposure covariate X_5 represents the standardized historical proportion of summer days with UV Index greater than or equal to C in s_k , the k^{th} participant's grid location of residence under age 13.

As in the previous section, the above basic expression uses indicator functions to specify whether UV exposure, time-outdoors, or both are included in the model. Model 9 includes no covariate for individual hours spent in the sun, so that the exposure variable is simply

$$\beta_5 X_5(s_k, C). \quad (9)$$

Next, Model 10 adds the standardized cumulative summer hours spent outside under age 13, X_6 . Here we now have two separate exposure terms,

$$\beta_5 X_5(s_k, C) + \beta_6 X_{6k}. \quad (10)$$

Finally, Model 11 weights the exposure term according to time spent in the sun. The β_5 exposure coefficient therefore estimates the effect of historical proportion of summer days with UV Index above some threshold, weighted by that person's childhood cumulative hours spent in the sun (X_6 , now scaled to be between 0 and 1). The exposure term is now

$$\beta_5 X_5(s_k, C) X_{6k}. \quad (11)$$

Strictly speaking, the assumption of asymptotic posterior normality that forms the theoretical support for DIC is not met in the case of a model containing a discrete finite parameter, such as C in these threshold models. However, given the unimodal appearance of the posterior distribution of C (Figure 5), DIC should still be sensible to use (if nothing else, as a score statistic similar to LPML). As such, we again compute DIC, using the posterior mean of C (rounded to the nearest integer) as the plug-in for the calculation of $\hat{D}$. These "quasi-DIC" scores and related statistics are shown in Table 6.

Similar to the previous section, the models that either did not contain the time-outdoors variable (Model 9) or added it as a separate covariate (Model 10) had the best DIC scores. Time-weighting the exposure measure (Model 11) worsened fit and DIC, also as in the two-stage models of average exposure. So again, we see that this idea, while intuitively appealing, is not well-supported by our data.

4.2 Parameter Estimation

Table 7 presents posterior parameter estimates for Model 10. The coefficients for age, blistering sunburns, and light complexion are significant and positively associated with NMSC risk, as is that of our peak exposure measure ($\beta_5 = 0.161$). The new time-outdoors parameter is not significant, but the trend is toward an association with increased risk of NMSC, with posterior mean ($\beta_6 = 0.046$) very similar to that of Model 7.

As seen in Figure 5, the posterior distribution of the threshold parameter C is unimodal and slightly skewed to the right. For public health communication, the International Commission on Non-Ionizing Radiation Protection (ICNRP) and the World Health Organization (WHO) have divided the UV Index scale into color-coded risk categories labeled “Low” (0–2), “Moderate” (3–5), “High” (6–7), “Very High” (8–10), and “Extreme” (11 and up). Above an index of 7, the international guidelines recommend avoiding midday sun, seeking shade, and using all protective measures (clothing, hat, sunscreen). The threshold density estimate in Figure 5 clearly favors UV Index values in the “High to Very High” categories. Thus, our results indicate that persons living in regions having larger proportions of summer days with noontime UV Index of “High” (6–7) or “Very High” (8–10) or greater have significantly higher NMSC risk.

As further validation of these threshold models, we performed sixteen simulations, four each with C fixed at 5, 7, 9 or 11. In each, we fixed the β s at their posterior means from Model 9 and fit this model to data Z_k^* simulated using the observed covariates of a random 50% ($19,081/2 \approx 9541$) subsample of the original dataset. The small number of replications (4) at each value of C was dictated by the lengthy run times of these models in WinBUGS. Nevertheless, in all 16 simulations, the posterior credible interval for C covered the true value, and the β s were very close to their true values; these results are shown graphically in Figure 6. Thus we are able to replicate the true C even when it differs from that (7 or 8) suggested by the full dataset.

5 Discussion and Future Work

In this paper, we have shown that modeling NMSC outcomes can be improved with additional information about UV exposure. In Section 3, the DIC-best model was one that incorporated several known covariates of NMSC risk (blistering sunburns, light complexion, age) along with a spatially modeled estimate of average UV erythemal dose during childhood, plus a term capturing time spent in the summer sun during childhood. A similar model form was selected as best in Section 4, using a deterministically-determined “peak” UV dose measure, instead of a statistically modeled average dose. We estimated a skin damage threshold corresponding to a “High to Very High” UV index (7 or 8). Thus, the UV dose covariate represented the historical proportion of summer days with noontime UV Index above this threshold in each geographical location. Again, adding time spent in the summer sun, rather than weighting the UV dose by this value, provided the best fit.

The methods described in this paper provide a useful model by which spatially oriented UV exposure information obtained from the TOMS satellites can characterize health effects of sunlight exposure. The potential to segregate exposure by smaller geographic area has clear advantages over methods which rely on single measures for entire states. The ability to identify an objective UV index threshold, of above 7 or 8 in this analysis, associated with disease risk will guide the development of appropriate causal models in more comprehensive analyses. In this paper we used only a subset of a much larger cohort to develop and test the methods. The application of the method to the entire cohort, which includes more sparsely populated grids, will require additional model development, but should ultimately improve the analysis.

Strengths of our analysis include the large and comprehensive nature of the USRT cohort, the geographical information available on each participant, and the full use of all geographical and individual-level information by our hierarchical Bayesian model. Limitations include the lack of temporal alignment in our Section 3 model between the years considered relevant for exposure in cohort participants (ages < 13 correspond to 1911-1974) and the years during which UV doses were measured (1979-2003). Fortunately, since our data did not show evidence of differential UV temporal changes over space, we were able to use the fitted temporal intercepts as a surrogate for exposure. However, if the age distribution of our cohort were shown to vary substantially over space (say, with most of the older members of the cohort having grown up in the north, and the younger members having grown up in the south), this could introduce systematic bias into our exposure estimates, since in this case the older members' exposure could be somewhat overstated due to their being less recent (when the sun's rays were less intense).

Our work thus far has also ignored spatial misalignment [50] in our data: the UV observations are the result of a weather-informed model that collects data at a point level, but produces grid square level exposure estimates, which we assume apply to every resident of the grid square. Fuentes et al. [16] give a generalized Poisson model for areal mortality counts $Y_{jk}(t)$ over strata (combinations of gender, age, and ethnicity) k and counties j by month t . Covariates in the log link include county-level confounders (e.g., ozone), strata-level confounders, and I different particulate matter (PM) exposure species. The spatial misalignment is resolved by integrating an underlying "total" PM surface over the counties, where this is in turn modeled linearly as a function of weather variables. A multivariate CAR (MCAR) model for spatial random effects that impact the PM exposure species allows a spatial covariance structure that evolves over time. Future work with our data should account for within-grid variability in the UV values, possibly using some of these ideas from Fuentes et al. [16].

On a related point, the form used in Section 2 was linear, forcing an assumption of a constant rate of change over time. But the raw data plots in Figure 1 suggest a nonlinear increase, with a sharp uptick in the last few years of observation. As such, future work looks to fitting more complex *Gaussian process* models, where we replace model (5) by

$$Y_i(t) = \mu_0 + \gamma_{0i} + \alpha W_{2i} + Z_i(t) + \epsilon_i(t),$$

where $Z_i(t)$ is a Gaussian process capturing temporal evolution, which can also be thought of as a nonparametric surrogate for our ignorance about the precise form of the (nonlinear) growth model. Specifically, writing $\mathbf{Z} = (\mathbf{Z}'(t_1), \dots, \mathbf{Z}'(t_T))'$ where $\mathbf{Z}(t) = (Z_1(t), \dots, Z_m(t))'$ for any t , we have $\mathbf{Z}$ as multivariate normal with mean vector 0 and covariance matrix $\Sigma = R(\phi) \otimes \tau^2(D - \alpha W)^{-1}$, the Kronecker product of a $T \times T$ temporal correlation matrix $R(\phi)$ (say, having (t_1, t_2) element $\exp(-\phi|t_1 - t_2|)$) and an $M \times M$ spatial correlation matrix of the usual CAR type.

We may now evaluate the posterior temporal gradient $Z_i'(t)$ at any time t for any grid box i as described in [6]; note this will no longer be constant across t , as our linear models assume. In particular, we are interested in the t where the gradient becomes significantly greater than zero. We might also investigate individual-specific temporal processes, $Z_{ij}(t)$, along with other specific assumptions such as additive spatial and non-spatial effects. Note that while the derivative process is never observed, inference may still proceed from the predictive distribution, $P(Z_i'(t_0)|\mathbf{Z})$. This is possible to compute because the associated covariance functions yield mean square differentiable paths, and because sampling from the posterior predictive is simple via composition (of a Gaussian kernel with the Gibbs posterior draws). Positive gradients are potentially important in our IR/UV/NMSC setting, since if UV exposure is increasing over time, identical self-reports of exposure from widely different years may

correspond to quite different exposures. Ideally, all of this structure would be added to our logistic model for the NMSC responses, now indexed by both individual k and exposure epoch t . While these temporal gradient models are likely too challenging to be fit using the standard spatial software tools available in WinBUGS, they may offer more accurate results that are worth the investment of analytical power and computing time.

In such continuous time model extensions as these, refinements of the UV dose variable are also possible. For example, we may redefine UV exposure as cumulative over ages 3 to 13

$$UV_k = \frac{1}{150} \sum_{j=1}^{11} \int_0^{150} \mu_{g_k, a_{k+j}}(t) dt,$$

where g_k indicates subject k 's grid of residence at age 10, a_k indexes the year in which k turned 2 years old, and 150 days of summer UV exposure are averaged in each year. An analogous measure for peak UV might be

$$UV_k^* = \frac{1}{150} \sum_{j=1}^{11} \int_0^{150} I(\mu_{g_k, a_{k+j}}(t) > C) dt.$$

Several additional quantities observed on the cohort members also inform future work. Yearly ionizing radiation dose estimates have been calculated [42], which may be combined additively or multiplicatively with UV doses, either contemporaneously or lagged by L years. Finally, we have data on the year of diagnosis, facilitating possibly more sensitive survival models.

Appendix: WinBUGS Code

Spatiotemporal UV Model (Section 2)

```

model
{
  for (i in 1:grids) {
    for (j in 1:T) {
      Y[i, j] ~ dnorm(mu[i, j], tau)
      mu[i, j] <- mu0 + gamma[i] + mu1 * (year[j]-1990)
      + alpha * (lat[i]-38.5)
    }
  }
  mu1 ~ dnorm(0.0, 1.0E-6)
  mu0 ~ dnorm(0.0, 1.0E-6)

  sigma0 ~ dunif(0.01, 100.0)
  tau0 <- 1/(sigma0*sigma0)

  for(j in 1:sumNumNeigh) {weights[j] <- 1}
  gamma[1:grids] ~ car.normal(adj[], weights[], num[], tau0)

  sigma ~ dunif(0.01, 100.0)
  tau <- 1/(sigma*sigma)

  alpha ~ dflat()
}

```


Hierarchical UV and Outcome Models (Section 3)

```

model
{
  for (i in 1:grids) {
    for (j in 1:T) {
      Y[i,j] ~ dnorm(mu[i,j], tau)
      mu[i,j] <- mu0 + gamma[i] + mu1*(year[j]-1991)
        + alpha*(lat[i]-38.5)
    }
  }

  for(k in 1:N ) {
    Z[k]~dbern(p[k])
    logit(p[k]) <-
      beta[1]
      + beta[2]*(age[k]-mean(age[]))
      + beta[3]*gender[k]
      + beta[4]*(blister[k]-mean(blister[]))
      + beta[5]*comp[k]
      #+ beta[6]*(gamma[grid[k]] + alpha*(lat[grid[k]]-38.5));
      #+ beta[6]*(gamma[grid[k]] + alpha*(lat[grid[k]]-38.5))
      #+ beta[7]*suntime[k];
      #+ beta[6]*(gamma[grid[k]] + alpha*(lat[grid[k]]-38.5))*suntime[k];
  }
  mu1 ~ dnorm(0.0,1.0E-3)
  mu0 ~ dnorm(0.0,1.0E-3)
  alpha ~ dflat()

  sigma0 ~ dunif(0.01,100.0)
  tau0 <- 1/(sigma0*sigma0)
  sigma ~ dunif(0.01, 100.0)
  tau <- 1/(sigma*sigma)

  for(j in 1:sumNumNeigh) {weights[j] <- 1}
  gamma[1:grids] ~ car.normal(adj[], weights[], num[], tau0)

  for(b in 1:6) {beta[b]~dnorm(0.0,1.0E-3)}
}

```

UV Peak Exposure and Outcome Models (Section 4)

```

model
{
  for(j in 1 : N ) {
    Z[j]~dbern(p[j])
    logit(p[j]) <-
      beta[1]
      + beta[2]*(age[j]-mean(age[]))
      + beta[3]*gender[j]
      + beta[4]*(blister[j]-mean(blister[]))
      + beta[5]*comp[j]
      # + beta[6]*X.5[grid[j],C]
      # + beta[6]*X.5[grid[j],C] + beta[7]*suntime[j]
      # + beta[6]*X.5[grid[j],C]*suntime[j]
  }

  for (k in 1:K) {p.c[k] <- 1/K}
  C ~ dcat(p.c[])

  for (b in 1:6) {beta[b]~dnorm(0.0,.001)}}

```

Acknowledgments

The work of the first and last authors was supported in part by NIH grant 1-R01-CA95955-01, while that of the second and third authors was supported by a grant from the University of Minnesota Cancer Center. The USRT Cohort Study is supported by the intramural research program of the National Cancer Institute, Contract N01-CP-31018. The authors are grateful to Ms. Sang Mee Lee for crucial initial programming and data analytic assistance, and to Dr. Sudipto Banerjee for helpful discussions.

References

1. Ambach W, Blumthaler M. Biological effectiveness of solar UV radiation in humans. *Experientia* 1993;49(9):747–753. [PubMed: 8405296]
2. American Cancer Society. Skin cancer: basal and squamous cell. 2007. <http://documents.cancer.org/118.00/118.00.pdf>
3. Armstrong BK. Epidemiology of malignant melanoma: intermittent or total accumulated exposure to the sun? *Journal of Dermatologic Surgery & Oncology* 1988;14(8):835–849. [PubMed: 3397443]
4. Armstrong BK, Krickler A. The epidemiology of UV induced skin cancer. *Journal of Photochemistry & Photobiology.B - Biology* 2001;63(13):8–18.
5. Banerjee, S.; Carlin, BP.; Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC Press; Boca Raton, FL: 2004.

6. Banerjee S, Gelfand AE. Bayesian wombling: curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association* 2006;101:1487–1501.
7. Bataille V, Sasieni P, Grulich A, Swerdlow A, McCarthy W, Hersey P, Bishop JAN, Cuzick J. Solar keratoses: a risk factor for melanoma but negative association with melanocytic naevi. *International Journal of Cancer* 1998;78(1):8–12.
8. Besag J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* 1974;36:192–236.
9. Boice JD, Mandel JS, Doody MM, Yoder RC, McGowan R. A health survey of radiologic technologists. *Cancer* 1992;69(2):586–598. [PubMed: 1728391]
10. Carlin, BP.; Louis, TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC Press; Boca Raton,FL: 2000.
11. Clouser MC, Harris RB, Roe DJ, Saboda K, Ranger-Moore J, Duckett L, Alberts DS. Risk group, skin lesion history, and sun sensitivity reliability in squamous cell skin cancer progression. *Cancer Epidemiology, Biomarkers & Prevention* 2006;15(11):2292–2297.
12. de Gruijl FR, van Kranen HJ, van Schanke A. UV exposure, genetic targets in melanocytic tumors and transgenic mouse models. *Photochemistry & Photobiology* 2005;81(1):52–64. [PubMed: 15535734]
13. Diffey BL, Gies HP. The confounding influence of sun exposure in melanoma. *Lancet* 1998;351(9109):1101–1102. [PubMed: 9660585]
14. Elwood JM, Gallagher RP, Hill GB, Spinelli JJ, Pearson JC, Threlfall W. Pigmentation and skin reaction to sun as risk factors for cutaneous melanoma: Western Canada Mlanoma Study. *British Medical Journal Clinical Research Ed* 1984;288(6411):99–102. [PubMed: 6419839]
15. Elwood JM, Jopson J. Melanoma and sun exposure: an overview of published studies. *International Journal of Cancer* 1997;73(2):198–203.
16. Fuentes M, Song HR, Ghosh SK, Holland DM, Davis JM. Spatial association between speciated fine particles and mortality. *Biometrics* 2006;62(3):855–863. [PubMed: 16984329]
17. Gallagher RP, Hill GB, Bajdik CD, Coldman AJ, Fincham S, McLean D, Threlfall WJ. Sunlight exposure, pigmentary factors, and risk of nonmelanocytic skin cancer: II squamous cell carcinoma. *Archives of Dermatology* 1995;131(2):164–169. [PubMed: 7857112]
18. Gallagher RP, Hill GB, Bajdik CD, Fincham S, Coldman AJ, McLean D, Threlfall WJ. Sunlight exposure, pigmentary factors, and risk of nonmelanocytic skin cancer: I basal cell carcinoma. *Archives of Dermatology* 1995;131(2):157–163. [PubMed: 7857111]
19. Godar DE, Wengraitis SP, Shreffler J, Sliney DH. UV doses of Americans. *Photochemistry & Photobiology* 2001;73(6):621–629. [PubMed: 11421067]
20. Grossman D, Leffell DJ. The molecular basis of nonmelanoma skin cancer: new understanding. *Archives of Dermatology* 1997;133(10):1263–1270. [PubMed: 9382565]
21. Han J, Colditz GA, Hunter DJ. Risk factors for skin cancers: a nested case-control study within the Nurses' Health Study. *International Journal of Epidemiology* 2006;35(6):1514–1521. [PubMed: 16943234]
22. Herman JR, Krotkov N, Celarier E, Larko D, Labow G. Distribution of UV radiation at the earths surface from TOMS-measured UV-backscattered radiances. *Journal of Geophysical Research* 1999;104(D10):12059–12076.
23. Kollias N, Baqer AH, Ou-Yang H. Diurnal and seasonal variations of the UV cut-off wavelength and most erythemally effective wavelength of solar spectra. *Photodermatology, Photoimmunology & Photomedicine* 2003;19(2):89–92.
24. Kricker A, Armstrong BK, English DR. Sun exposure and non-melanocytic skin cancer. *Cancer Causes & Control* 1994;5(4):367–392. [PubMed: 8080949]
25. Kricker A, Armstrong BK, English DR, Heenan PJ. Pigmentary and cutaneous risk factors for non-melanocytic skin cancer-a case-control study. *International Journal of Cancer* 1991;48(5):650–662.
26. Kricker A, Armstrong BK, English DR, Heenan PJ. Does intermittent sun exposure cause basal cell carcinoma? A case-control study in Western Australia. *International Journal of Cancer* 1995;60(4):489–494.
27. Kricker A, Armstrong BK, English DR, Heenan PJ. A dose-response curve for sun exposure and basal cell carcinoma. *International Journal of Cancer* 1995;60(4):482–488.

28. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38(4):963–974. [PubMed: 7168798]
29. MacKie RM. Long-term health risk to the skin of ultraviolet radiation. *Progress in Biophysics & Molecular Biology* 2006;92(1):92–96. [PubMed: 16616325]
30. Matta J, Ramos J, Armstrong R, D'Antoni H. Environmental UV-A and UV-B threshold doses for apoptosis and necrosis in human fibroblasts. *Photochemistry and Photobiology* 2005;81:563–568.
31. McKinlay AF, Diffey BL. A reference action spectrum for ultraviolet induced erythema in human skin. *Commission Internationale de l'Eclairage* 1987;6:17–22.
32. Miller AJ, Melanoma MCM Jr. *New England Journal of Medicine* 2006;355(1):51–65. [PubMed: 16822996]
33. Neale RE, Davis M, Pandeya N, Whiteman DC, Green AC. Basal cell carcinoma on the trunk is associated with excessive sun exposure. *Journal of the American Academy of Dermatology* 2007;56(3):380–386. [PubMed: 17097387]
34. I. C. on Non-Ionizing Radiation Protection. Guidelines on limits of exposure to ultraviolet radiation of wavelengths between 180 nm and 400 nm (incoherent optical radiation). *Health Physics* 2004;87(2):171–186. [PubMed: 15257218]
35. Ortonne JP. From actinic keratosis to squamous cell carcinoma. *British Journal of Dermatology* 2002;146(Suppl 61):20–23. [PubMed: 11966728]
36. Ramirez CC, Federman DG, Kirsner RS. Skin cancer as an occupational disease: the effect of ultraviolet and other forms of radiation. *International Journal of Dermatology* 2005;44(2):95–100. [PubMed: 15689204]
37. Ramos J, Villa J, Ruiz A, Armstrong R, Matta J. UV dose determines key characteristics of nonmelanoma skin cancer. *Cancer Epidemiology, Biomarkers & Prevention* 2004;13(12):2006–2011.
38. Rosso S, Zanetti R, Pippione M, Sancho-Garnier H. Parallel risk assessment of melanoma and basal cell carcinoma: skin characteristics and sun exposure. *Melanoma Research* 1998;8(6):573–583. [PubMed: 9918420]
39. Sabburg J, Rives JE, Meltzer RS, Taylor T, Schmalzle G, Zheng S, Huang N, Wilson A. Comparisons of corrected daily integrated erythemal uvr data from the US EPA/UGA network of Brewer spectroradiometers with model and TOMS-inferred data. *Journal of Geophysical Research* 2002;107(D23):5–1.
40. Scotto J, Cotton G, Urbach F, Berger D, Fears T. Biologically effective ultraviolet radiation: surface measurements in the United States, 1974 to 1985. *Science* 1988;239(4841 Pt 1):762–764. [PubMed: 3340857]
41. Scotto J, Fears TR. The association of solar ultraviolet and skin melanoma incidence among Caucasians in the United States. *Cancer Investigation* 1987;5(4):275–283. [PubMed: 3664331]
42. Simon, SL.; Weinstock, RM.; Doody, MM.; Neton, J.; Wenzl, T.; Stewart, P.; Mohan, AK.; Yoder, C.; Hauptmann, M.; Freedman, M.; Cardarelli, J.; Feng, HA.; Bouville, A.; Linet, M. Status report on estimating historical radiation doses to a cohort of U.S. radiologic technologists. *Proceedings of the 11th International Congress of International Radiation Protection Association*; Madrid.
43. Spiegelhalter DJ, Best N, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002;64:583–639.
44. Spiegelhalter DJ, Thomas A, Best N, Lunn D. WinBUGS User Manual, Version 1.4.1. 2004
45. Tucker MA. Sun exposure measurements in populations. *Nutrition Reviews* 2007;65(8 Pt 2):S84–6. [PubMed: 17867377]
46. van Dam RM, Huang Z, Rimm EB, Weinstock MA, Spiegelman D, Colditz GA, Willett WC, Giovannucci E. Risk factors for basal cell carcinoma of the skin in men: results from the health professionals follow-up study. *American Journal of Epidemiology* 1999;150(5):459–468. [PubMed: 10472945]
47. Woods TN, Prinz DK, Rottman GJ, London J, Crane PC, Cebula RP, Hilsenrath E, Brueckner GE, Andrews MD, White OR, VanHoosier ME, Floyd LE, Herring LC, Knapp BG, Pankratz CK, Reiser PA. Validation of the UARS solar ultraviolet irradiances: Comparison with the ATLAS 1 and 2 measurements. *Journal of Geophysical Research* 1996;101(D6):9541–9569.

48. Yoshinaga S, Hauptmann M, Sigurdson AJ, Doody MM, Freedman DM, Alexander BH, Linet MS, Ron E, Mabuchi K. Nonmelanoma skin cancer in relation to ionizing radiation exposure among US radiologic technologists. *International Journal of Cancer* 2005;115(5):828–834.
49. Zanetti R, Rosso S, Martinez C, Navarro C, Schraub S, Sancho-Garnier H, Franceschi S, Gafa L, Perea E, Tormo MJ, Laurent R, Schrameck C, Cristofolini M, Tumino R, Wechsler J. The multicentre south European study 'Helios' I: Skin characteristics and sunburns in basal cell and squamous cell carcinomas of the skin. *British Journal of Cancer* 1996;73(11):1440–1446. [PubMed: 8645595]
50. Zhu L, Carlin BP, Gelfand AE. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* 2003;14(5):537–557.

Figure 1.
Spatiotemporal visualization of TOMS UV erythemal dose.

Figure 2. Posterior mean map and 95% equal-tail credible intervals for the γ_{0i} in Model 5.

Figure 3.
Spatial distribution of cohort participants, cancer cases, and ratios (cancer incidence).

Figure 4. Posterior means and 95% credible intervals for γ_{0i} in selected grids. One-stage results are shown as grey line segments, two-stage models by black error bars (Models 6, 7, & 8, left-to-right in each cluster).

Figure 5.
Posterior density of the UV Index threshold parameter C .

Figure 6. Posterior density estimates for C in 50% subsample simulated data threshold models. Dotted lines indicate true C values, while solid error bars give 95% credible intervals.

Table 1

Model comparison, spatiotemporal UV models.

Model	D^*	D^*	p_D	DIC	MSEP (1979)	MSEP (1992)	MSEP (3 grids)
Model 2	715.50	552.50	163.02	878.50	159.51	83.70	NA
Model 3	650.40	567.30	83.06	733.50	140.86	80.65	400.00
Model 4	650.20	582.40	67.81	718.00	145.06	81.27	144.27
Model 5	621.50	579.40	42.04	663.50	140.12	79.49	121.51

Table 2

Posterior estimates for fixed effects in Model 5.

Parameter	Mean	S.D.	2.5%	Median	97.5%
α (latitude)	-4.76	0.30	-5.40	-4.76	-4.19
μ_0 (grand mean)	166.8	0.18	166.4	166.8	167.1
μ_1 (year)	0.59	0.02	0.55	0.59	0.64
σ (error standard deviation)	7.37	0.13	7.13	7.37	7.62
σ_0 (spatial standard deviation)	3.82	0.43	3.04	3.79	4.75

Table 3
Posterior summaries of NMSC parameters from one-stage (UV plug-in) version of Model 7.

Parameter	Mean	S.D.	2.5%	Median	97.5%
β_0 (intercept)	-2.474	0.040	-2.554	-2.474	-2.396
β_1 (age)	0.054	0.003	0.049	0.054	0.060
β_2 (sex)	0.013	0.035	-0.058	0.013	0.080
β_3 (# of blistering sunburns)	0.022	0.002	0.018	0.022	0.027
β_4 (complexion)	0.328	0.029	0.272	0.328	0.384
β_5 (UV exposure)	0.012	0.002	0.008	0.012	0.017
β_6 (summer hrs outside < age 13)	0.046	0.028	-0.010	0.045	0.101

Table 4

Model comparison, two-stage (UV-cancer outcome) models.

Model	D^*	D^*	p_D	DIC	LPML (10% hold-out)
Model 6	926.70	920.48	6.22	932.92	-0.2771
Model 7	925.02	917.79	7.24	932.26	-0.2769
Model 8	934.04	927.92	6.12	940.16	-0.2773

Table 5

Posterior estimates for fixed effects in Model 7

Parameter	Mean	S.D.	2.5%	Median	97.5%
β_0 (intercept)	-2.473	0.040	-2.552	-2.472	-2.396
β_1 (age)	0.054	0.003	0.049	0.054	0.060
β_2 (sex)	0.012	0.034	-0.056	0.012	0.079
β_3 (# of blistering sunburns)	0.022	0.002	0.017	0.022	0.027
β_4 (complexion)	0.328	0.028	0.273	0.328	0.384
β_5 (UV exposure)	0.013	0.002	0.008	0.013	0.017
β_6 (summer hrs outside < age 13)	0.046	0.028	-0.010	0.046	0.101
μ_0 (grand mean)	166.8	0.312	166.1	166.8	167.4
μ_1 (year)	0.590	0.024	0.544	0.590	0.636
α (latitude)	-4.71	0.285	-5.30	-4.71	-4.16
σ (error standard deviation)	7.373	0.126	7.13	7.37	7.63
σ_0 (spatial standard deviation)	3.83	0.435	3.03	3.80	4.74

Table 6

DIC model comparison across peak exposure models.

Model	$D^{\sim}$	$D^{\hat{}}$	p_D	DIC
Model 9	925.80	919.15	6.66	932.46
Model 10	924.13	916.48	7.65	931.77
Model 11	933.56	927.14	6.43	939.99

Table 7

Posterior estimates for fixed effects in Model 10.

Parameter	Mean	S.D.	2.5%	Median	97.5%
C (exposure threshold)	7.923	0.971	7	8	10
β_0 (intercept)	-2.474	0.040	-2.554	-2.474	-2.397
β_1 (age)	0.054	0.003	0.049	0.054	0.060
β_2 (sex)	0.013	0.035	-0.058	0.012	0.082
β_3 (# of blistering sunburns)	0.022	0.002	0.017	0.022	0.027
β_4 (complexion)	0.328	0.028	0.273	0.328	0.383
β_5 (UV exposure)	0.161	0.031	0.100	0.161	0.223
β_6 (summer hrs outside < age 13)	0.046	0.028	-0.009	0.046	0.101