Published in final edited form as:

Spat Spatiotemporal Epidemiol. 2010 July 1; 1(2-3): 133–141. doi:10.1016/j.sste.2010.03.005.

JOINT DISEASE MAPPING OF CERVICAL AND MALE OROPHARYNGEAL CANCER INCIDENCE IN BLACKS AND WHITES IN SOUTH CAROLINA

Georgiana Onicescu¹, Elizabeth G. Hill¹, Andrew B. Lawson¹, Jeffrey E. Korte¹, and M. Boyd Gillespie²

¹Division of Biostatistics and Epidemiology, Department of Medicine, Medical University of South Carolina

²Department of Otolaryngology, Medical University of South Carolina

Abstract

Human papillomavirus (HPV) infection is an established causal agent for cervical cancer and a subset of oropharyngeal cancers. It is hypothesized that orogenital transmission results in oral cavity infection. In this paper we explore the geographical association between cervical and male oropharyngeal cancer incidence in blacks and whites in South Carolina using Bayesian joint disease mapping models fit to publicly available data. Our results suggest weak evidence for county-level association between the diseases, and different patterns of joint disease behavior for blacks and whites.

Keywords

Bayesian hierarchical models; Ecological study; Disease mapping; Markov chain Monte Carlo; Shared component model; Cervical cancer; Oropharyngeal cancer; Human papillomavirus

1. Introduction

High-risk human papillomavirus (HPV) types have been shown to be a necessary cause of cervical cancer (Franco et al., 2001; Bosch and de Sanjose, 2003; Schiffman et al., 2007) and growing evidence shows these same high-risk HPV types, especially HPV16, are causal agents in the etiology of some head and neck cancers, particularly in the oropharynx and tonsillar regions (Adelstein et al., 2009; Vidal and Gillison, 2008; Gillison et al., 2008). HPV-related oropharyngeal cancer is associated with increased numbers of lifetime sexual partners and increased numbers of oral sex partners; it is therefore hypothesized that oral HPV infection occurs via sexual transmission during orogenital contact (Anaya-Saavedra et al., 2008; Gillison et al., 2008; D'Souza et al., 2009).

In this study we explored the putative etiological link between cervical and oropharyngeal (tongue, tonsil and oropharynx) cancer in an ecological study of cancer incidence in blacks and whites in South Carolina. Using a version of the shared-component model of Knorr-

Held and Best (2001), we examined the joint spatial distribution of these diseases using publicly available incidence data from 1996 to 2005. Our objective was to explore the county-level geographic association of the two diseases supporting a hypothesis of a common (or shared) behavioral etiology. Because of established differences in the racial distribution of HPV-associated oropharyngeal cancers (Gillespie et al., 2009), we performed separate analyses for blacks and whites, with the additional goal of identifying racial disparities in the shared risk of the two diseases.

We provide an overview of the relationship between HPV infection, and cervical and oropharyngeal cancer in section 2. We describe our analytic methods in section 3 and summarize results in section 4. We conclude with a discussion of findings and future directions in section 5.

2. Clinical background

2.1 Cervical cancer and HPV

The human papillomavirus is a double-stranded DNA virus classically associated with development of cutaneous and genital warts (Gillespie et al., 2009). Based on epidemiologic and natural history studies of cervical HPV infection and cancer, HPV infection is known to be sexually acquired (Adelstein et al, 2009) and is a necessary agent for the development of cervical cancer. While over 1000 HPV strains have been isolated, two in particular are highly associated with cervical cancer development – HPV types 16 and 18. HPV16 accounts for 50% to 60% of cervical cancer cases globally, while HPV18 accounts for 10% to 12% (Bosch and de Sanjose, 2003). HPV-DNA is present in 90% to 100% of cervical cancer specimens, compared to rates of 5% to 20% in specimens from women without cervical cancer (Bosch and de Sanjose, 2003). Approximately 45% of 20 to 24 year-old women are infected with HPV, with the prevalence of infection declining steadily with increasing age to 20% for women ages 50 to 59 years (Dunne et al., 2007). Vaccines targeting HPV16 and HPV18 have been developed for cervical cancer prevention, and current US recommendations are that girls of ages 11 to 12 years be vaccinated (Schiffman et al., 2007).

Unraveling the biological mechanisms by which HPV infection progresses to cervical cancer is an area of active research. Although most HPV infections resolve, a major component of progression is infection persistence (for years or even decades), with the greatest development of cancer occurring among persistent high-risk type infections (Bosch et al., 2002). In a large population-based study of Costa Rican women, Schiffman and colleagues (2005) estimated nearly 20% of HPV16-infected women developed severe dysplasia or cancer.

Cervical cancer is graded on a scale of 1 to 3 based on the histology (or differentiation) of the cancer cells, a measure of how closely cancer cells resemble normal cells. Grade 1 cancer cells are well-differentiated, grade 2 are moderately well-differentiated, and grade 3 are poorly differentiated or undifferentiated (Canadian Cancer Society). Cervical cancer is further characterized by stage, an indication of disease severity, according to guidelines established by the American Joint Commission of Cancer (AJCC) Tumor, Node, Metastasis (TNM) system. Stage 0 cervical cancer (or cervical carcinoma in situ) indicates disease found only on the surface of the cervix; stage 1 disease has spread into the cervix and possibly the uterus; stage 2 disease has grown beyond the cervix and uterus, but has not invaded the pelvic wall or lower vaginal region; stage 3 disease has spread to the lower vaginal region or pelvic wall; and stage 4 disease has spread to the bladder or rectum or is growing out of the pelvis (American Cancer Society).

Although all cervical cancers arise from HPV infection, additional risk factors for the disease are: high parity; history of smoking; long-term use of oral contraceptives; use of exogenous hormones; and sexual characteristics of the male partner, including number of lifetime sexual partners and lack of circumcision (Bosch et al., 2002; Bosch, 2003). In a small case control study in Honduras, Ferrera and colleagues identified additional risk factors including: age at first intercourse, age at first pregnancy, education, and number of previous cervical cancer screens (Ferrera et al., 2000). In addition, some sexually transmitted diseases, such as HIV, have been found to be associated with cervical cancer (Bosch et al., 2002).

It is estimated that 11,270 women were newly diagnosed with cervical cancer in 2009 and 4,070 women died from their disease (National Cancer Institute). US cervical cancer incidence and mortality rates in 2005 were 8.1 and 2.4 per 100,000 women, respectively. Black women have higher rates of both incidence and mortality compared to white women. Specifically, incidence estimates for 2005 were 7.7 versus 10.3 per 100,000 white and black women, respectively. Race-specific mortality estimates were 2.2 and 4.4 per 100,000 white and black women, respectively (U.S. Cancer Statistics Working Group, 2009).

2.2 Oropharyngeal cancer and HPV

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common solid tumor worldwide (Gillespie et al., 2009). Oral cancer subtypes include cancer of the lip, tongue, salivary glands, floor of the mouth, gums, nasopharynx, tonsil, oropharynx, hypopharynx, pharynx, and other locations in the oral cavity. It is estimated that in 2009 there were over 30,000 new oral cancer cases and more than 6,000 deaths (National Cancer Institute). US HNSCC prevalence estimates in 2004 were 157,000 (0.11%) men and 87,000 (0.07%) women (National Institute for Dental and Craniofacial Research). Oral cancer incidence estimates based on Surveillance, Epidemiology, and End Results (SEER) data indicate comparable rates for whites and blacks (10.6 and 10.7 per 100,000, respectively). Male incidence rates are higher for both races, with gender-specific rates of 15.7 and 6.1 per 100,000 white men and women, respectively. Gender-specific rates for blacks are 17.2 and 5.7 per 100,000 men and women, respectively (National Institute for Dental and Craniofacial Research).

Head and neck squamous cell carcinoma grading is similar to that of cervical cancer. Likewise, disease severity is established according to AJCC TNM staging system. Stage 1 includes tumors that are less than 2 centimeters in greatest diameter without nodal disease; stage 2 includes tumors that are between 2 and 4 cm in greatest diameter without nodal disease; stage 3 includes tumors that are 6 cm or less with or without a single ipsilateral lymph node metastasis less than 3 cm in greatest diameter; and stage 4 tumors are those involving deep neck structures (deep tongue muscles, bone, etc.) or that have multiple lymph nodes or lymph nodes greater than 3 cm in diameter.

While the majority of HNSCC cases are attributable to alcohol and tobacco abuse (the latter being a common risk factor for cervical cancer), up to 20% of HNSCC patients present with little to no tobacco and alcohol exposure, suggesting alternative risk factors for the disease. Recent evidence links specific HPV strains (especially HPV 16) to HNSCC, especially those arising from oropharyngeal subsites including the base of the tongue, the tonsils, and the oropharynx (Gillespie et al., 2009; Herrero et al, 2003). HPV-positive HNSCC cases typically are white males under the age of 45 with no history of tobacco use and mild to moderate alcohol use (Gillespie et al., 2009). Additionally, the sexual behaviors of HPV-positive cases differ significantly from those of HPV-negative cases: positive cases have greater numbers of lifetime sexual partners, oral sex partners, and history of oral-anal contact (Gillison et al., 2008; Herrero et al., 2003). It is therefore hypothesized that oral-

genital contact is the primary vector by which HPV is transmitted to the upper aerodigestive tract (Gillespie et al., 2009).

Among oropharyngeal cancer subtypes linked to HPV-infection, 2005 age-adjusted incidence rates per 100,000 men for cancers of the tongue, tonsil and oropharynx, respectively, are 4.1, 2.7, and 0.7. Corresponding race-specific incidence estimates per 100,000 white and black men, respectively, are: 4.3 and 3.5 (tongue); 2.8 and 2.6 (tonsil); and 0.7 and 1.2 (oropharynx). Age-adjusted mortality rates per 100,000 men for these cancer subtypes are 0.9, 0.3 and 0.3. Race-specific mortality rates per 100,000 white and black men, respectively, are: 0.9 and 1.4 (tongue); 0.3 and 0.5 (tonsil); and 0.3 and 0.7 (oropharynx).

Dunne and colleagues conducted a systematic literature review of 40 studies reporting prevalence of HPV infection in men (Dunne et al., 2006). For studies in which multiple anogenital anatomic sites or specimens (semen or urine) were evaluated, HPV prevalence varied from 1.3% to 72.9%. Furthermore, HPV has been detected in 20% of HNSCC and in 50% of oropharyngeal or tonsillar carcinomas (Gillison, 2004). Analysis of data from nine SEER program registries from 1973 to 2004 indicate HPV-related HNSCC cancer (that is, cancers of the oropharynx, base of tongue and tonsils) is more common among men than women, with 72.6% and 27.4% of such cancers attributable to men and women, respectively (Chaturvedi et al., 2008).

3. Methods and analysis

3.1 Study design

It is accepted that HPV infection is causally associated with both cervical and oropharyngeal cancer. HPV is associated with oral sexual behaviors, such as the lifetime number of oral sex partners (D'Souza G et al., 2007; D'Souza et al., 2009). Assuming oral cavity HPV infection is transmitted via orogenital contact, we hypothesized that elevated cervical cancer rates would align geographically with higher male oropharyngeal cancer rates. In addition, we hypothesized that the geographical association between cervical and male oropharyngeal cancer rates would differ by race since HPV-positive HNSCC patients tend to be white. We investigated our hypotheses based on an ecological analysis of race-specific cervical and male oropharyngeal incidence in South Carolina. We selected a modeling approach that facilitates discovery of a common risk between the diseases, the shared component model of Knorr-Held and Best (2001).

3.2 Data sources and description

Observed frequencies of county-level incident cancer cases were downloaded from the South Carolina Community Assessment Network (SCAN) website (<http://scangis.dhec.sc.gov/scan/>), which allows the generation of user-specified tables of cancer incidence based on data from the South Carolina Central Cancer Registry (SCCCR). The SCCCR data are derived from case reports from SC hospitals with and without cancer registries, pathology laboratories and treatment centers. Additionally, the registry includes data from cases who are SC residents treated in 16 other states (including bordering states) with which SCCCR has an agreement. The data used in the present analysis were last updated on 31 January 2008 and include incident cases from 1996 to 2005. The SCCCR conducts data quality audits and has an estimated data case completeness of 99.7% (South Carolina Department of Health and Environmental Control).

We restricted HNSCC frequencies to male cases with cancer subsites of the tongue, tonsil and oropharynx. For both diseases, we obtained case frequencies for the 46 SC counties separately for blacks and whites ages 20 to 59 years for all cancer grades and stages, except

carcinoma in situ of the cervix which is not reported by SCCCR. Other races represent a small proportion of cases in the database for our selected age group (2.2% for cervical cancer and less than 0.5% for oropharyngeal cancer); we therefore limited our investigation to black and white cases.

To facilitate age-adjustment of expected county-level case frequencies, we obtained sex-by-age population counts for blacks and whites in each SC county using 2000 US census data (<http://factfinder.census.gov/>). We computed state-level estimates of cervical and male oropharyngeal cancer rates separately for blacks and whites by dividing the total number of incident cases by the total population in age groups of 20 to 39 years and 40 to 59 years. The expected number of cancer cases in each county was then calculated for each age group and race category by taking the product of the appropriate census count data and the estimated state cancer rate. Age-group specific counts were then summed to obtain age-adjusted county-level expected cervical and male oropharyngeal counts for blacks and whites.

We obtained county-level race specific counts of the number of individuals living below poverty level from the South Carolina Community Profiles website (<http://www.sccommunityprofiles.org/>). For each county, we then calculated race-specific estimates of the proportion of the population living below poverty level by dividing the appropriate count by the total county population. These values were used in our poverty-adjusted models.

3.3 Bayesian modeling approach

For the spatial analysis we adopted a Bayesian approach modeling the two diseases jointly using the shared component model of Knorr-Held and Best (2001). In our analysis we examined cervical cancer incidence among females ages 20 to 59 years and its relation to oropharyngeal cancer incidence among males in the same age group. We were interested in the hypothesis that sexual behavior influences the incidence of HPV-related oral cancers. In particular we were interested in evidence supporting the hypothesis that a common behavioral risk exists between the incidence of cervical and male oropharyngeal cancer. Because of differences by race in the incidence of HPV-positive HNSCC, we performed separate analyses for blacks and whites. In addition, we sought to establish whether there was a racial disparity in any linkage found. To assess this linkage we examined the geographical distribution of case counts of the two diseases to assess whether there is any commonality. Spatial similarity or correlation between disease rates may hold evidence for common confounding or explanatory factors that are unobserved in the study. One such unobserved factor could be sexual behavior. Before considering modeling approaches, we computed and mapped the standardized incidence ratios (SIR) as the ratio of the observed to expected counts (Fig. 1). We notice racial differences in the geographic distribution of SIRs. For white cases, counties with high SIRs for both diseases are located primarily in the south, while for black cases counties with elevated SIRs are located mostly in the north-east.

Different approaches can be taken to the joint spatial analysis of diseases. One approach is to consider a cross-correlation between diseases (Gelfand and Vounatsou, 2003). This yields a single measure of co-variation but does not have a spatial expression. Instead we adopted the shared component approach of Knorr-Held and Best (2001). In this approach it is assumed each disease has a shared component in the relative risk and this component is allowed to vary spatially. We believe this approach yields more information about putative association than the cross-correlation approach. In our example we hypothesized that HPV infection transmitted via orogenital contact was a major common risk for both diseases.

The model adopted is implemented easily within the Bayesian modeling software WinBUGS (Lawson, 2009, ch. 9; Lunn et al., 2009). For each county, observed cervical and

male oropharyngeal cancer counts are considered to have independent Poisson distributions conditional on the modeled relative risk. Specifically,

$$y_{ik} \sim \text{Poisson}(E_{ik}\theta_{ik})$$

where y_{ik} is the observed number of cases in the i^{th} SC county ($i = 1, \dots, 46$) for the k^{th} disease ($k = 1$ for cervical cancer and $k = 2$ for oropharyngeal cancer). In this formulation the Poisson mean consists of a product of the expected count (E_{ik}) and relative risk (θ_{ik}). The expected count is assumed fixed and obtained from age by gender stratified rates for a standard population (here, the population of South Carolina for the same time period). We then model the log relative risk as a linear combination of effects, both covariate and random. We assume a hierarchical form for the model and so the relative risk is defined to be a function of a set of parameters. Specifically, we define the following relative risk model for the two diseases:

$$\begin{aligned}\log(\theta_{i1}) &= \alpha_1 + \Phi_{i1} + \phi_i \xi \\ \log(\theta_{i2}) &= \alpha_2 + \Phi_{i2} + \phi_i / \xi\end{aligned}$$

where $\Phi_{ik} = u_{ik} + v_{ik}$ is a random convolution component consisting of, respectively, a spatial correlation effect (u_{ik}) and an uncorrelated term (v_{ik}). In addition each disease has an intercept α_k and a shared component ϕ_i which is scaled by a parameter ξ . We fit separate models for black and white cases.

3.4 Imputing case counts

Due to confidentiality concerns, cell counts of 10 or fewer are suppressed in SCAN data tables. Specifically, any cell with 1 to 4 observations is reported as “<5” and those with 5 to 10 observations are reported as 10. For cervical cancer, the proportion of values reported as <5 was 17% for whites and 26% for blacks, while 20% and 37% were rounded to 10 for whites and blacks respectively. For tongue, tonsils and oropharynx cancers combined, the proportion of values reported as <5 was 39% for whites and 50% for blacks, while 22% and 37% of values were rounded to 10 for whites and blacks respectively.

WinBUGS treats missing variables as unknown model parameters, thereby facilitating imputation via posterior estimates. Specifically, we specified prior distributions for missing case counts via appropriate truncated Poisson distributions,

$$y_{ik} \sim \begin{cases} \text{Poisson}(E_{ik}\theta_{ik}) \cdot 1(1 \leq y_{ik} < 5), & y_{ik} = 1, 2, 3, 4 \\ \text{Poisson}(E_{ik}\theta_{ik}) \cdot 1(5 \leq y_{ik} \leq 10), & y_{ik} = 5, 6, 7, 8, 9, 10 \end{cases}$$

and imputed values were obtained by sampling from their posterior distributions after model convergence.

3.5 Prior distribution specifications for model components

In our Bayesian formulation we specified a variety of prior distributions for our parameters and random effects. In general, we assumed non-informative prior distributions where appropriate. We assigned non-informative uniform priors to the intercept terms, α_k . The disease specific random effects, Φ_{i1} and Φ_{i2} , were assumed to have a convolution prior, which is the sum of unstructured (v_{ik}) and spatial (u_{ik}) effects, the latter accounting for the spatial correlation between the random effects. We assumed the unstructured component had a normal prior distribution with mean zero and a hyperparameter precision having a large

variance gamma distribution. In our model we used a gamma distribution with mean of 1000 and very large variance of 2×10^6 . The spatial components of the disease specific random effects were assumed to have an intrinsic normal conditional autoregressive (CAR) distribution with unit weight for neighboring areas. For the k th disease and i th area with neighborhood δ_i , this conditional specification is:

$$u_{ik} \sim N\left(\sum_{j \in \delta_i} u_{jk} / n_{\delta_i}, \tau_k^{-1} / n_{\delta_i}\right)$$

where τ_k is the precision corresponding to the k th disease, and n_{δ_i} is the number of neighbors for region i . The precision of the spatial effects was assigned a gamma distribution with the same mean and variance as the uncorrelated effect precision hyperprior distribution.

The shared random effects ϕ_i were similarly assigned a convolution prior (sum of unstructured and spatial random effects). The unstructured shared component was assigned a normal distribution with mean zero and a precision with large variance gamma hyperprior distribution. The spatial shared component was assigned an intrinsic normal CAR distribution, with unit neighborhood weight and the precision also assumed to have a large variance gamma prior distribution.

The log scaling factor ($\log(\xi)$) was assigned a normal distribution with mean zero and precision 5.9, thus leading to a 95% probability that ξ^2 is between 1/5 and 5.

3.6 Model implementation

Statistical analyses were performed using WinBUGS version 1.4.3 with the GEOBUGS version 1.2 add on (Lunn et al., 2009). Each model was run to convergence based on multiple chain diagnostics (Brooks-Gelman-Rubin statistics). In the converged sample we monitored DIC and pD, and also all relevant parameters of interest. Convergence usually took place within 10000 iterations and we based inference on a chain length of 10000 after convergence. Our criteria for model selection was a relative goodness-of-fit measure based on the well know deviance information criterion (DIC) (Spiegelhalter et al., 2002) and also a secondary criterion, pD reflecting the effective number of model parameters. The DIC is a criterion adjusted for the number of parameters in the model. The lower the DIC value, the better the model fit. For models with DICs within approximately 2 units, a secondary measure, pD, can be considered. Here, a model with a lower pD and therefore greater parsimony is preferred. We used these criteria to guide model selection.

4. Results

Table 1 contains the DIC and parameter degrees of freedom (pD) for various models considered. We explored models adjusted and unadjusted for poverty level, with only unstructured random effects (UH) for the shared and specific relative risks, only spatial random effects (CAR) and both unstructured and spatial random effects (UH+CAR). Additionally, for the UH+CAR models, we explored competing models assuming no scaling for the shared relative risk. We considered a model adjusted for poverty since we believed poverty could influence a county's risk profile. For whites, although not the most parsimonious, the model with the lowest DIC was the unadjusted UH+CAR (pD=16.87 and DIC=397.56). For blacks, the unadjusted UH+CAR model had a DIC value of 271.71, smaller than the other considered models except for the unadjusted UH, which had a lower DIC. The difference was smaller than 2, suggesting a not significant difference in the fit of the two models. However, this model did not perform well in whites, having a DIC with more than 2 units higher compared to the unadjusted UH+CAR model. Based on all the

above consideration, we considered the unadjusted UH+CAR the best overall model for both whites and blacks that we explored in our analyses.

From our model we obtained posterior estimates for the parameters and measures of interest, including means, standard deviations, Monte Carlo error estimates, medians and 95% credible intervals (CIs) (Table 2). The Monte Carlo error for the parameters of interest ranged between 3.5% and 6.9% of the standard deviation (SD) for whites and between 2.59% and 6.80% of the SD for blacks.

The log shared relative risk, ϕ_i , measures the commonality between the two diseases in each county. Knorr-Held and Best (2001) weight the shared relative risk by a scaling factor, ξ , to accommodate different risk gradients for each disease. Thus, $\phi_i\xi$ and ϕ_i/ξ reflect the influence of the shared component on county-level cervical and male oropharyngeal cancer log relative risks, respectively. For white cases, the median value of the scaling factor was 1.001 (95% CI = 0.45 to 2.19) indicating equivalent contribution of the shared component to the two diseases. For black cases, the median was 1.036 (95% CI = 0.49 to 2.20) indicating a slightly higher contribution of the shared component to cervical cancer than to male oropharyngeal cancer.

The relative risk (RR) ratio was computed as the ratio of the scaled shared log relative risk for cervical cancer to the scaled shared log relative risk for oral cancer, thus being equal to the squared scaling factor (RR ratio = $(\phi_i\xi)/(\phi_i/\xi) = \xi^2$). A relative risk ratio greater than one indicates greater contribution of the shared component to cervical cancer risk; a value between 0 and 1 indicates greater contribution of the shared component to male oropharyngeal cancer risk. The estimated posterior median RR ratio was 1.003 (95% CI = 0.20 to 4.82) and 1.073 (95% CI = 0.24 to 4.86) for white and black cases, respectively, supporting the conclusion that unobserved common risk factors for the diseases contribute more to cervical cancer risk, especially for black cases.

We computed the variability of the scaled log shared relative risk, $Var(\phi_i\xi)$ and $Var(\phi_i/\xi)$ for CC and OC respectively (called *shared variance*) and the variability of the disease specific log relative risks, $Var(\Phi_{ik})$ (called *specific variance*). We also measured the fraction of the total variation in log relative risk for each disease explained by the shared component (called *shared fraction*), computed as the ratio of the shared variance divided by the sum of the shared and specific variance. Specifically, for cervical cancer the proportion of variability attributable to the shared component is

$$Var(\phi_i\xi)/(Var(\phi_i\xi)+Var(\Phi_{i1}))$$

and for male oropharyngeal cancer, the proportion is obtained as the ratio

$$Var(\phi_i/\xi)/(Var(\phi_i/\xi)+Var(\Phi_{i2})).$$

The fraction shared can take values between 0 and 1, with higher values indicating the shared component explains a greater proportion of the variability in relative risk for the disease. For white cases, the estimated posterior medians of the fraction shared were 0.30 (95% CI = 0.016 to 0.93) and 0.25 (95% CI = 0.013 to 0.93) for cervical and male oropharyngeal cancer, respectively. For black cases, estimated medians of the fraction shared were 0.43 (95% CI = 0.016 to 0.96) and 0.50 (95% CI = 0.031 to 0.97) for cervical and male oropharyngeal cancer, respectively. Although the credible intervals are wide, these results suggest differences by race of the two diseases and warrants further studies at a finer

resolution such as zip codes or individual level in order to be able to estimate these measures with a higher precision.

The median shared relative risk ranged from 0.62 to 1.83 in whites and from 0.44 to 2.15 in blacks. The posterior medians for the specific relative risks in whites were between 0.31 and 2.54 for cervical cancer and between 0.33 and 2.59 for oral cancer. In blacks the minimum and maximum values of the disease specific relative risk were 0.31 to 2.47 for cervical cancer and from 0.38 to 2.36 for oral cancer. Figure 2 displays South Carolina maps of the posterior means of the unscaled shared relative risk. For white cases, regions with slightly higher shared risk were observed in the southeast while for black cases regions with higher shared risk were in the northeast of the state.

Figure 3 shows the South Carolina maps of the disease specific components for black and white cases for cervical ($e^{\Phi_{i1}}$) and oropharyngeal cancer ($e^{\Phi_{i2}}$). Racial differences in the distribution of regions with higher specific risk are noticed for both oropharyngeal cancer and cervical cancer. Regions with higher cervical cancer specific risk are in the central part of the state for whites and in the east for blacks. Regions with higher oral cancer specific risk are more scattered throughout the state, while for blacks the estimated oropharyngeal cancer specific risk is similar for all counties and there are no high risk regions.

5. DISCUSSION

In this paper, we fit the shared component model of Knorr-Held and Best (2001) to examine the joint spatial distributions of cervical and oropharyngeal cancer incidence in South Carolina. Our interest in this joint analysis stems from the hypothesis that the common causal agent (HPV infection) is transmitted from the genitalia to the oral cavity via sexual behavior. We provided a detailed description of the data sources, calculation of expected rates, and model structure and assumptions that can be used for the planning of similar analyses.

Because HPV-associated HNSCC is more common among white as opposed to black males, we expected the shared component to contribute more to the log relative risk of oropharyngeal cancer among white rather than black males. However, the estimated total between-area variation in risk captured by the shared component was higher for blacks for both cervical and oropharyngeal cancers. The reasons for this finding are unclear and can be due to residual confounding. A possible explanation for these findings could be the racial differences in HPV knowledge, beliefs, access to care and medical insurance, educational programs and screening. More whites have heard of HPV compared to blacks and whites have higher HPV knowledge than blacks (Cates et al., 2009). Finally, although we performed our analysis on only those HNSCCs most commonly associated with HPV infection, the dominant risk factors for the disease are alcohol and tobacco abuse. Thus, the subset of oropharyngeal cancers truly associated with HPV infection, even within this enriched population, is still likely to be relatively small and even more difficult to detect at the county level of geographic aggregation. Analyses including person-level behavioral data are needed to gain insight into the common role of HPV infection in cervical and oropharyngeal cancer.

In general, we have used non-subjective choices for the model parameters by assuming non-informative priors. However, there are several limitations to our study that should be acknowledged. First, we performed county-specific analysis, thus assuming same county sexual partners. In addition, we could not track residential mobility that may have affected the sexual partner network. Another study limitation is the fact that the SCCCR data did not include earlier cervical cancer stages (carcinoma in situ) that may also have been associated

with oral cancer in males. This may have decreased the detected association between cervical cancer and oral cancer. Furthermore, our county-level analysis did not allow for adjustment with potential individual-level confounders (e.g. education, income, marital status). Nevertheless, the shared component model was able to group the known and unknown spatially structured risk factors common to both diseases into the shared component and the risk factors specific to each disease into the disease specific component. However, a constraint of this model is the fact that it assumes independence between the shared and specific components, which may not be true. Also, as we analyzed aggregate count data we can only estimate gross variation in the joint distribution of the two diseases. This did not allow us to obtain estimates for some parameters of interest with precision, thus obtaining wide confidence intervals. Analysis of data at a finer resolution (such as zip codes or census tracts) may provide more precise estimates.

Having missing case counts is another limitation of our study. However, we imputed the values via posterior estimates and, since we had lower and upper bounds for the unknown values, we believe this approach does not induce major biases in our results.

Future studies implementing an extension of the shared component model to other HPV-related cancers such as cancer of the vulva, anus or esophagus, or conjunctival squamous cell carcinoma (Gillison et al., 2003), could provide additional insight regarding transmission and potential interventions.

Our findings suggest regions in SC where the shared risk for cervical and male oropharyngeal cancer for whites and blacks is elevated, and provide additional avenues of research for studying the associations between the HPV-associated cancers.

Abbreviations

CC	Cervical Cancer
OC	Oropharyngeal Cancer
HNSCC	Head and Neck Squamous Cell Carcinoma
SC	South Carolina
SCAN	South Carolina Community Assessment Network
SCCCR	South Carolina Central Cancer Registry
SIR	Standardized incidence ratio

Acknowledgments

GO was partially supported by NIH grants K25DE016863 and 1P30 CA138313-01. EGH was partially supported by NIH grant K25DE016863. AL was partially supported by NIH grant 5R21HL088654-02

References

- Adelstein, DJ.; Ridge, JA.; Gillison, ML.; Chaturvedi, AK.; D'Souza, G.; Gravitt, PE., et al. Head Neck. Vol. 31. Washington, D.C.: 2009. Head and neck squamous cell cancer and the human papillomavirus: summary of a national cancer institute state of the science meeting, November 9–10, 2008; p. 1393-1422.doi:10.1002/hed.21269
- American Cancer Society. [accessed 02/25/10]. Available from <http://documents.cancer.org/115.00/115.00.pdf>
- Anaya-Saavedra G, Ramirez-Amador V, Irigoyen-Camacho E, Garcia-Cuellar C, Guido-Jimenez M, Mendez-Martinez R, et al. High association of human papillomavirus infection with oral cancer: a

- case-control study. *Arch Med Res*. 2008; 39(2):189–197. doi:10.1016/j.arcmed.2007.08.003. [PubMed: 18164962]
- Bosch FX, Lorincz A, Munoz N. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol*. 2002; 55(4):244–265. [PubMed: 11919208]
- Bosch FX, de Sanjose S. Chapter 1: Human papillomavirus and cervical cancer – burden and assessment of causality. *J Natl Cancer Inst Monogr*. 2003; 2003(31):3–13. [PubMed: 12807939]
- Canadian Cancer Society. [accessed 02/25/10]. Available from <http://www.cancer.ca>
- Cates JR, Brewer NT, Fazekas KI, Mitchell CE, Smith J, Smith JS. Racial differences in HPV knowledge, HPV vaccine acceptability, and related beliefs among rural, southern women. *J Rural Health*. 2009; 25(1):93–97. [PubMed: 19166567]
- Chaturvedi AK, Engels EA, Anderson WF, Gillison ML. Incidence trends for human papillomavirus–related and –unrelated oral squamous cell carcinomas in the United States. *J Clin Oncol*. 2008; 26(4):612–619. doi:10.1200/JCO.2007.14.1713. [PubMed: 18235120]
- D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med*. 2007; 356(19):1944–1956. doi:10.1056/NEJMoa065497. [PubMed: 17494927]
- D'Souza G, Agrawal Y, Halpern J, Bodison S, Gillison M. Oral sexual behaviors associated with prevalent oral human papillomavirus infection. *J Infect Dis*. 2009; 199(9):1263–1269. doi:10.1086/597755. [PubMed: 19320589]
- Dunne EF, Nielson CM, Stone KM, Markowitz LE, Giuliano AR. Prevalence of HPV infection among men: a systematic review of the literature. *J Infect Dis*. 2006; 194(8):1044–1057. doi:10.1086/507432. [PubMed: 16991079]
- Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, Patel SS, et al. Prevalence of HPV infection among females in the United States. *JAMA*. 2007; 297(8):813–819. doi:10.1001/jama.297.8.813. [PubMed: 17327523]
- Ferrera A, Velema J, Figueroa M, Bulnes M, Toro L, Claros J, et al. Co-factors related to the causal relationship between human papillomavirus and invasive cervical cancer in Honduras. *Int J Epidemiol*. 2000; 29(5):817–825. [PubMed: 11034963]
- Franco EL, Duarte-Franco E, Ferenczy A. Cervical cancer: epidemiology, prevention and the role of human papillomavirus infection. *CMAJ*. 2001; 164(7):1017–1025. [PubMed: 11314432]
- Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*. 2003; 4(1):11–25. [PubMed: 12925327]
- Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992; 7(4):457–472.
- Gillespie MB, Rubinchik S, Hoel B, Sutkowski N. Human papillomavirus and oropharyngeal cancer: what you need to know in 2009. *Curr Treat Options Oncol*. 2009 Sep 19. Epub. doi:10.1007/s11864-009-0113-5.
- Gillison ML, Shah KV. Chapter 9: Role of mucosal human papillomavirus in nongenital cancers. *J Natl Cancer Inst Monogr*. 2003; 2003(31):57–65. [PubMed: 12807947]
- Gillison ML. Human papillomavirus-associated head and neck cancer is a distinct epidemiologic, clinical, and molecular entity. *Semin Oncol*. 2004; 31(6):744–754. doi:10.1053/j.sonc.2004.09.011. [PubMed: 15599852]
- Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst*. 2008; 100(6):407–420. doi:10.1093/jnci/djn025. [PubMed: 18334711]
- Herrero R, Castellsague X, Pawlita M, Lissowska J, Kee F, Balaram P, et al. Human papilloma virus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst*. 2003; 95(23):1772–1783. doi:10.1093/jnci/djg107. [PubMed: 14652239]
- Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. *J R Stat Soc Ser A*. 2001; 164(1):73–85.
- Lawson, AB. *Bayesian Disease Mapping: hierarchical modeling in spatial epidemiology*. New York: CRC press; 2009.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modeling framework: concepts, structure, and extensibility. *Stat Comput*. 2000; 10(4):325–337. doi:10.1023/A:1008929526011.

- National Cancer Institute. [accessed 02/25/10]. Available from <http://www.cancer.gov>
- National Institute for Dental and Craniofacial Research. [accessed 02/25/10]. Available from <http://www.nidcr.nih.gov>
- Peltecu G, Bari M, Lancu G, Popa F. Human papilloma virus and cervical preinvasive disease. *J Med Life*. 2009; 2(4):373–377. [PubMed: 20108750]
- Plummer M, Herrero R, Franceschi S, Meijer CJ, Snijders P, Bosch FX, et al. Smoking and cervical cancer: pooled analysis of the IARC multi-centric case-control study. *Cancer Causes Control*. 2003; 14(9):805–814. doi:10.1023/B:CACO.0000003811.98261.3e. [PubMed: 14682438]
- Schiffman M, Castle PE, Jerronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007; 370(9590):890–907. doi:10.1016/S0140-6736(07)61416-0. [PubMed: 17826171]
- South Carolina Community Assessment Network. [accessed 02/25/10]. Available from <http://scangis.dhec.sc.gov/scan>
- South Carolina Community Profiles. [accessed 02/25/10]. Available from <http://www.sccommunityprofiles.org>
- South Carolina Department of Health and Environmental Control. [accessed 02.25.10]. Available from <http://www.scdhec.gov>
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol*. 2002; 64(4):583–639.
- Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. GeoBUGS User Manual, Version 1.2. 2004 September.
- US census 2000. [accessed: 02/25/10]. Available from <http://factfinder.census.gov>
- U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2005 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2009 [accessed 02.25.10]. Available from www.cdc.gov/uscs
- Vidal L, Gillison ML. Human papillomavirus in HNSCC: recognition of a distinct disease type. *Hematol Oncol Clin North Am*. 2008; 22(6):1125–1142. doi:10.1016/j.hoc.2008.08.006. [PubMed: 19010263]

Fig. 1.

Map of the standardized incidence ratios for cervical and oropharyngeal cancer.

Fig. 2.
Average unscaled shared relative risk based on 10000 draws from posterior distributions.

Fig. 3.
Average disease specific relative risks based on 10000 draws from posterior distributions.

Table 1

Goodness of fit of the models based on DIC. Parameters degrees of freedom (pD) are also presented and should be positive. Models with negative pD were not considered appropriate. UH=unstructured heterogeneity; CAR=conditional autoregressive model; UH+CAR=both unstructured heterogeneity and conditional autoregressive components.

	Whites		Blacks	
	pD	DIC	pD	DIC
Unadjusted UH	11.78	400.54	5.45	270.10
Unadjusted CAR	13.71	398.14	7.50	272.73
Unadjusted UH+CAR	16.87	397.56	9.87	271.71
Unadjusted UH+CAR, no scaling for the shared RR ($\xi=1$)	18.13	399.69	10.59	272.53
Adjusted for poverty UH	<0 (model not appropriate)	-	12.26	275.47
Adjusted for poverty CAR	11.82	400.43	9.94	276.23
Adjusted for poverty UH+CAR	16.12	398.60	12.54	275.66
Adjusted for poverty UH+CAR, no scaling for the sharedRR ($\xi=1$)	18.19	400.68	13.57	276.49

Table 2

Posterior results for the parameter estimates (unadjusted UH+CAR model):

	Whites		Blacks	
	Mean (std)	Median (2.5%, 97.5%)	Mean (std)	Median (2.5%, 7.5%)
RR ratio (ξ^2)	1.40 (1.32)	1.003 (0.20, 4.82)	1.46 (1.31)	1.073 (0.24, 4.86)
Scaling factor (ξ)	1.08 (0.47)	1.001 (0.45, 2.19)	1.12 (0.45)	1.036 (0.49, 2.20)
Shared fraction CC $\text{Var}(\phi_i \xi) / (\text{Var}(\phi_i \xi) + \text{Var}(\Phi_{i1}))$	0.38 (0.29)	0.30 (0.016, 0.93)	0.45 (0.30)	0.43 (0.016, 0.96)
Shared fraction OC $\text{Var}(\phi_i / \xi) / (\text{Var}(\phi_i / \xi) + \text{Var}(\Phi_{i2}))$	0.34 (0.29)	0.25 (0.013, 0.93)	0.50 (0.30)	0.50 (0.031, 0.97)
Shared variance CC $\text{Var}(\phi_i \xi)$	0.0047 (0.0060)	0.0026 (0.00023, 0.022)	0.0077 (0.011)	0.0035 (0.00021, 0.040)
Shared variance OC $\text{Var}(\phi_i / \xi)$	0.0050 (0.0069)	0.0026 (0.00023, 0.025)	0.0069 (0.012)	0.0028 (0.00021, 0.038)
Specific variance CC $\text{Var}(\Phi_{i1})$	0.0097 (0.011)	0.0059 (0.00056, 0.039)	0.0090 (0.011)	0.0049 (0.00052, 0.041)
Specific variance OC $\text{Var}(\Phi_{i2})$	0.013 (0.015)	0.0076 (0.00066, 0.055)	0.0053 (0.0079)	0.0026 (0.00041, 0.027)