Infection with MCPyV, KIV, WUV, and HPV as potential risk factors for lung cancer

Danny V. Colombara

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Lisa E. Manhart, Chair

Noel S. Weiss

Stephen E. Hawes

Program Authorized to Offer Degree:

Public Health - Epidemiology

University of Washington

Abstract

Infection with MCPyV, KIV, WUV, and HPV as potential risk factors for lung cancer

Danny V. Colombara

Chair of the Supervisory Committee:
Associate Professor Lisa E. Manhart
Department of Epidemiology

Lung cancer is the most commonly diagnosed cancer worldwide and the only cancer among the top ten leading causes of death. Due to the lung's propensity for infection, and higher lung cancer incidence among immunocompromised individuals, some pathogens have been assessed as potential causes. As oncogenic viruses with the ability to infect the respiratory track, human polyomaviruses (HPyV) and human papillomaviruses (HPV) are candidate pathogens. However, cross-sectional studies of the former are limited and studies of the latter have generated conflicting results. Prospective sero-studies, utilizing liquid bead microarray antibody (LBMA) assays to assess the relationship between prior infection and incident lung cancer could potentially resolve the conflict. However, standard LBMA data analysis methods are lacking. We compared six LBMA analytic methods in Monte Carlo-type simulated datasets. Logistic regression trend tests and unpaired t-tests had superior statistical efficiency and less bias than logistic regression with dichotomized predictors. We also performed a nested-case control study assessing the association of antibodies to Merkel cell (MCPyV), KI (KIV), and WU (WUV) HPyVs as well as six high-risk and two low-risk HPV types with lung cancer. We tested serum from the Carotene and Retinol Efficacy Trial (CARET), conducted 1985-2005, using LBMA.

Cases (n=200) had incident lung cancer and controls (n=200) were frequency matched on age at enrollment, year of enrollment, date of serum collection, intervention arm assignment, and the number of serum freeze-thaw cycles. Based on results of our simulation study, we performed trend tests for all antibody measures. To assure comparability with previous studies, we also assessed the association of HPyV antibodies (divided into quartiles) and HPV antibodies (seropositive versus seronegative) with lung cancer using logistic regression. There was no evidence of an association between HPyV (P>0.10 for all trend tests; odds ratio (OR) range 0.59 to 1.22, P>0.10 for all) or HPV antibodies (P≥0.10 for all trend tests; OR range 0.25 to 2.54, P>0.10 for all), and incident lung cancer. These results suggest that MCPyV, KIV, and WUV are not associated with lung cancer, and are consistent with prior studies that found no evidence for an association between HPV infection and lung cancer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation was generously supported by the Fogarty International Clinical Research Scholars and Fellows program (R24 TW007988), the Cancer Prevention Training in Nutrition, Exercise & Genetics program at the University of Washington (R25 CA094880), the Mustard Seed Foundation, and numerous individuals through Experiment.com.

I would like to thank Lisa Manhart for not only serving as my academic advisor and dissertation committee chair, but also as an exemplary mentor. Every grant and fellowship application that I pursued seemed possible because of her enduring faith and optimism. I am also indebted to Noel Weiss for patiently listening to my many dissertation concepts and gently redirecting me until I found a suitable topic. I would like to thank Stephen Hawes for modeling excellence in teaching, mentoring, and research – with a healthy dose of humor and good cheer. I am grateful for Denise Galloway, for this dissertation would not have been possible without gracious invitation to work in her lab and learn from her and her staff. I also owe my deepest gratitude to James Hughes, who introduced me to statistics and helped me move from theory to practice with this dissertation.

I would also like to thank Joseph "Jody" Carter, for his patience in training me to perform the Luminex assay, troubleshooting experiments, and answering countless questions. I also extend my thanks to Greg Wipf for getting me up to speed in the lab, Gary Goodman for providing access to the CARET biorepository, and Matt Barnett for consultation regarding specimen selection and study design. I cannot thank You-Lin Qiao and Jennifer S. Smith enough for extending the opportunity to train in China and helping conceive this dissertation.

I am forever indebted to my wife, Anita, who steadfastly supports my every endeavor, including this dissertation process. I am thankful for my son, Silas, whose humor and music remind me that there is more to life than research. I am grateful for my daughter, Cassia, whose beauty and joy remind me to live life, not plow through it. Above all, I thank Jesus in whom I live and move and have my being.

Chapter 1: Analysis of Liquid Bead Microarray Antibody Assay Data for Epidemiologic Studies of Pathogen-Cancer Associations

Abstract

Background: Liquid bead microarray antibody (LBMA) assays are used to assess pathogen-cancer associations, yet analytic methods differ between studies, limiting comparability.

Methods: To assess methods for analyzing LBMA data, we generated 10,000 Monte Carlo-type simulations of log-normal antibody distributions (exposure) with 200 cases and 200 controls (outcome). We estimated type I error rates, statistical power, and bias associated with three types of analytic techniques: (a) t-tests; (b) logistic regression with a linear predictor; and (c) logistic regression with predictors dichotomized according to four methods of defining cut-points: 200 or 400 MFI determined *a priori*; the mean MFI among controls plus two standard deviations; and the optimal value based upon receiver operating characteristic (ROC) curve analysis. We also applied these models, and data visualizations (kernel density plots, ROC curves, predicted probability plots, Q-Q plots), to empirical data evaluating the association between HPV16 L1 antibody response and colorectal polyps to assess the consistency of the exposure-outcome relationship.

Results: All strategies had acceptable type I error rates ($0.030 \leq P \leq 0.048$), except for the dichotomization according to optimal sensitivity and specificity (type I error rate = 0.27). Among the remaining methods, logistic regression with a linear predictor and t-tests had the highest power (*Power*=1.00 for both) to detect a mean difference of 1.0 MFI (median fluorescence intensity) on the log scale and were unbiased. Dichotomization methods upwardly biased the risk estimates.

Conclusion: Logistic regression with linear predictors and unpaired t-tests were superior to logistic regression with dichotomized predictors for assessing disease associations with LBMA data.

1.1 Introduction

Serologic antibody assays are important tools for investigating associations of infectious diseases with acute and chronic conditions in epidemiologic studies. Since the start of the third millennium (1), liquid bead microarray antibody (LBMA) assays have been used to screen for antibodies to pathogens that may be associated with human cancers (2–8). One such pathogen, human papillomavirus (HPV), is an established cause of cervical (9), other anogenital (10), and oropharyngeal (11) cancers. High-risk oncogenic HPV types are sexually transmitted and have been the subject of numerous studies utilizing LBMA.

LBMA assays test for serum antibodies by incubating sera with fluorescently labeled microspheres that are bound to antigens of interest. Bound antibodies are then labeled with secondary antibodies, and flow cytometry is used to generate the median fluorescence intensity (MFI), a continuous measure of the strength of an antibody response to a particular antigen. The core advantage of LBMA is its multiplexing capability, which allows for screening of hundreds of distinct antibodies at once, far more than with enzyme-linked immunosorbent assays (ELISA). LBMA multiplexing saves time, conserves sera for future studies, and is high-throughput, with the ability to analyze up to 1,000 specimens daily. Yet, the lack of a standard method for analyzing LBMA data limits the comparability of results across studies of the same agents.

As with ELISA (12), the absence of available standards and researchers' desire for binary immune status indicators (i.e., seropositive vs. seronegative) have led to defining MFI cut-points in a variety of ways in studies utilizing LMBA. For example, previous publications on human papillomavirus (HPV) defined HPV-related MFI cut-points in at least four different ways: the mean value in healthy blood donors plus three standard deviations (SD) (2), five SD above the mean of the sampled distribution (after removing outliers) (3), the mean values among virgins plus two SDs (7), and the mean values among healthy blood donors with only one lifetime sex

partner plus two SDs (8). Studies that used standard deviations to define cut-points likely assumed a normal distribution of MFI among unexposed individuals. This would result in approximately 95.4% or 99.7% of unexposed persons falling within two or three SD, respectively. However, the distributions are not necessarily normal and these methods would misclassify a percentage of uninfected participants as infected. In studies of sexually active individuals that lack samples from persons with low exposure to sexually transmitted HPV (i.e., virgins), MFI cut-points have been determined using receiver operating characteristic (ROC) curves (4), arbitrary *a priori* thresholds (5), and the upper quartile of MFI among control participants (6). This broad variety of methods used hinders comparisons of results across studies.

Even if a standard method for dichotomizing LBMA data existed, doing so may be problematic. Dichotomizing continuous predictors can reduce statistical power by up to a third or half (13–15), inflating the needed sample sizes and associated costs. In addition, dichotomization may conceal more complex exposure-outcome relationships (16) and may influence the estimated measure of association by introducing bias (17). In response, some have sought to avoid the dichotomization dilemma by testing for differences between mean MFI using one-way analysis of variance (ANOVA) or t-tests (4). Others suggested modeling continuous data, while freeing it from the constraints of linear assumptions. For example, though it does not provide a concise summary test statistic, cubic spline regression allows for close approximation of almost any smooth curve and can provide risk estimates over the full range of the continuous predictor (18). Similarly, quantile-quantile (Q-Q) plots, which depict the cumulative distribution of cases and controls as well as the odds ratio (OR) for all levels of a continuous predictor, have been suggested as an exploratory tool (17).

Our primary objective was to identify the most statistically efficient and unbiased method(s) for detecting associations between LBMA antibody measures (MFI) and disease status in epidemiologic studies of virus-cancer associations. Our secondary objective was to compare the utility of different data visualizations of MFI and disease status as exploratory tools. We used simulated and empirical datasets of natural HPV infection as our motivating examples. We pursued both objectives with the goal of encouraging researchers working in this field to adopt standards that will improve comparability across studies.

1.2 Methods

1.2.1 Data Sources

To estimate the statistical power (1 minus the false-negative proportion), type I error (the false-positive proportion), and bias resulting from different analytic methods, we generated 10,000 Monte Carlo-type simulated datasets per analysis. Each dataset contained 200 controls and 200 cases randomly selected from natural log (ln) normal distributions of MFI. We set the parent distribution of the controls to have a mean of 3 and a standard deviation (SD) of 1.6. The parent distributions of the cases had set means of 3, 4, 5, 6, 7, 8, or 9 and a SD of 1.6. The selected distributions represent positive associations over the linear range of the assay and within the range of MFI data that are often observed in studies of natural HPV infection using LBMA. For example, lnMFI of 3 corresponds with MFI=20, which is low, and lnMFI of 9 corresponds with MFI=8,103, which is extremely high.

We also investigated potential differences in measures of the association between MFI and disease status in empirical data. We selected a dataset with a published null association to assess whether using other analytic methods would have identified a positive association. Specifically, we utilized a dataset from a case-control study of HPV antibodies in association with colorectal hyperplastic polyps in men from the Minnesota Cancer Prevention Research Unit Polyp Study (19). Using LBMA to test plasma from 97 cases and 184 controls and a cut-point of greater than 400 MFI for seropositivity, the authors found no significant association between any HPV antibodies and hyperplastic polyps. The crude OR for HPV-16 L1 antibodies was 0.62 (95% confidence interval (CI) 0.16-2.35), and no rationale for the pre-specified cut-point was provided in the publication. We reanalyzed the HPV-16 antibody data from that study because it is the HPV type most commonly associated with cancer (20,21) and is commonly assessed in LBMA-based serologic studies of HPV.

1.2.2 Analytic Methods

We evaluated the simulated datasets with case-control status as the outcome and MFI as the predictor. We tested associations using three types of analytic techniques: (a) t-tests; (b) logistic regression with a linear predictor; and (c) logistic regression with predictors dichotomized according to four methods described in the literature. Those methods defined cut-points as: 200 or 400 MFI (5,19) determined *a priori*; the mean MFI among controls plus two standard deviations (8); and the optimal value based upon ROC curve analysis (4). Because the last method was not described in detail in previous publications, we used the MFI cut-point corresponding to the maximum J-statistic (sensitivity + specificity - 1) (22). For each analytic method, we calculated statistical power as the ratio of the number of tests with *P*-values <0.05 compared to the total number of tests (10,000). Type I error calculations were the same as those for power, but were limited to situations where the true mean difference was zero.

We also estimated bias using the simulated datasets with a true mean difference of zero. The bias for the logistic regression models was estimated by calculating the mean OR minus 1. Similarly, we estimated the bias for the t-tests by calculating the mean of the mean differences.

We then analyzed the Minnesota Cancer Prevention Unit Polyp Study dataset using: logistic regression with a linear predictor; logistic regressions with each of the four types of dichotomous predictors; and an unpaired t-test. We compared the *P*-value, magnitude of the association, and the direction of the association for each type of analysis.

We also compared four different methods of plotting the relationship between continuous MFI and disease status. First, we generated kernel density plots of the distribution of lnMFI among the cases and controls. Second, following an example in the literature (4), we performed nonparametric ROC regression and plotted the results as an ROC curve, with the corresponding

area under the curve (AUC). We overlaid lines on the ROC curve highlighting the optimal cut-point based upon the maximum J-statistic described above. Third, we created a graph of the predicted probability of being a case (y-axis) in relation to lnMFI (x-axis), based upon a logistic regression model using a restricted cubic spline of MFI with three knots as the predictor (23). Fourth, we generated Q-Q plots with the empirical cumulative distribution function for controls and cases on the x- and y-axes, respectively (17). We superimposed the Q-Q plot data on constant OR curves of 100, 20, 10, 5, 4, 3, 2, 1, .5, .33, .25, .20, .10, .05, and .01 to facilitate visual estimation of the OR for any lnMFI value.

We considered *P*-values less than 0.05 to be statistically significant and conducted all analyses using Stata/SE 13.1 (StataCorp, College Station, TX). The Institutional Review Board of the Fred Hutchinson Cancer Research Center approved the secondary analysis of the Minnesota Cancer Prevention Research Unit Polyp Study data.

1.3 Results

1.3.1 Simulation assessment of type I error, power, and bias

In analyses of the simulated data, the type I error using the J-statistic was inflated (P=0.27), but was reasonable for all other methods (0.03≤P≤0.048) (Table 1.1). For a true mean difference of 1 lnMFI, logistic regression with continuous MFI, logistic regression with dichotomization using the maximum J-statistic, and the unpaired t-test had the highest power (*Power*=1.00). Power was lower for logistic regression with dichotomization at 200 MFI (*Power*=0.97), 400 MFI (*Power* =0.86), and the mean of the controls plus two SD (*Power*=0.76).

Logistic regression with a linear predictor and unpaired t-tests generated no bias (data not shown). However, dichotomization at 200 MFI biased the OR upward by 8%; dichotomization at 400 MFI biased the OR upward by 24%; dichotomization at the mean of the controls plus two SD biased the OR upward by 29%; and dichotomization based upon the J-statistic biased the OR upward by 62%.

Kernel density, ROC, predicted probability, and Q-Q plots were created for mean differences of 0, 1, 2, and 3 and are available as Figures 1.1, 1.2, 1.3, and 1.4, respectively. Figure 1.1 shows completely overlapping kernel density plots, an ROC curve with an AUC of 0.50, a predicted probability graph centered on 0.5 over the full range of lnMFI, and a Q-Q plot with the data plotted along the null referent (OR=1.0). Figure 1.2 shows a right shift of the kernel density plot toward higher MFI, an increased AUC (0.67), a slightly sigmoidal diagonal on the predicted probability graph, and a slight downward curve on the Q-Q plot. Figure 1.3 continues the trends seen in Figure 1.2, with a greater shifting of the kernel density plot, an increase in the AUC (0.81), a steeper sigmoidal predicted probability graph, and a deeper bend in the Q-Q plot. Figure 1.4 is similar to Figure 1.3, but with more pronounced shifting of the kernel density plots,

an increased AUC (0.908), an even steeper sigmoidal predicted probability plot, and an even deeper bend in the Q-Q plot.

1.3.2 Empirical assessment of association with the outcome

In logistic regression analyses of the Minnesota Cancer Prevention Research Unit Polyp Study data, the ORs were not statistically significant ($P>0.05$), regardless of the MFI cut-point used (Table 1.2). However, there was variation in the estimated direction of the association. Modeling MFI as a continuous linear predictor yielded no association (OR=1.00, 95 CI: 0.87-1.15). In contrast, cut-points at 200 MFI (OR=0.76, 95% CI: 0.31-1.91), 400 MFI (OR=0.62, 95% CI: 0.16-2.35), and the mean among controls plus two SD (OR=0.75, 95% CI: 0.14-3.96) generated ORs less than one, whereas a cut-point based on the maximum J-statistic yielded an OR greater than one (OR=1.59, 95% CI: 0.93-2.72). The t-test produced no evidence of an association ($P=0.99$) and a mean difference of 0.

The kernel density plot of the Minnesota data shows overlapping log-normal curves for cases and controls (Figure 1.5). The area under the ROC curve is 0.509, with an optimal cut-point of 2.9 lnMFI, based upon the maximum J-statistic. The predicted probability graph is essentially flat, with no evidence of an association. The entire lnMFI line on the Q-Q plot is close to the null referent curve (OR=1), with no evidence for an association with increasing MFI.

1.4 Discussion

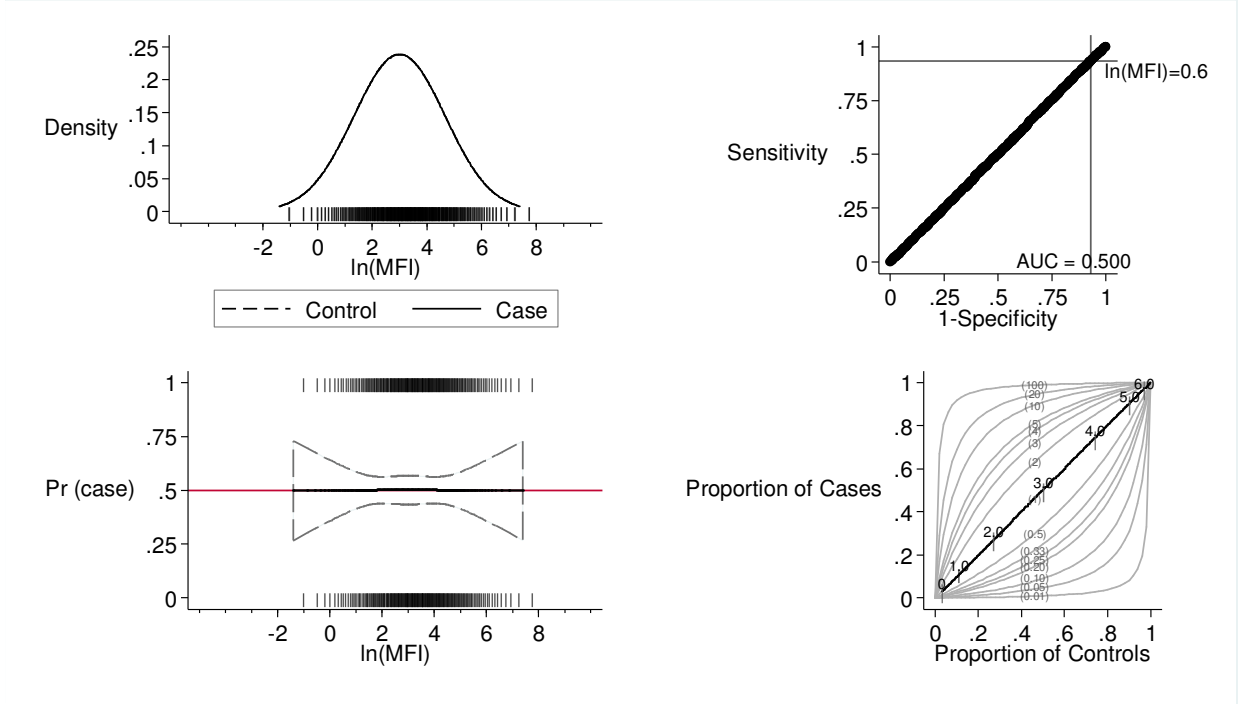
Based upon our simulated data, logistic regression with continuous linear MFI and unpaired t-tests provided the best, and nearly identical, combinations of high statistical power, reasonable type I error, and unbiased estimates. Application of these methods to empirical data also demonstrated the similarity between these two analytic methods compared to the other four that we assessed.

The four data visualizations clearly depicted the lack of association between lnMFI and disease status in the empirical dataset. However, unlike the other three graphs, the ROC curve lacks information regarding the data density informing the structure of the curve. Of the remaining three, a benefit of the predicted probability (restricted cubic spline) plot is the facility with which researchers can adjust for potential confounding in the underlying logistic regression model. Regardless of the visualization selected, if researchers are interested in formal hypothesis testing, rather than simply describing the data, considerations should be given to the problem of multiple comparisons (17). Specifically, when researchers use graphs to inform the selection of a cut-point, they have in fact already visually assessed and discarded a series of potential cut-points to select the one they believe is most promising in light of their hypothesis. This is a violation of core assumptions underlying *a priori* hypothesis testing (17) and leads to inflation of the type I error rate. Therefore, risk estimates based on cut-points selected in this manner are best reported as exploratory findings and should be presented alongside estimates based on a range of alternative cut-points. *P*-values should not be reported. Those desiring to use visualizations to inform selection of cut-points for causal inference may consider data splitting (24). Partitioning the data randomly into two parts would allow the investigator to develop a hypothesis based on visual exploratory analyses of the first portion of the data while preserving the second portion of the data for hypothesis testing.

The literature demonstrates a strong preference among researchers for dichotomizing LBMA-based HPV antibody data. In light of our findings, we recommend that researchers consider other analytic methods. Analyses using unpaired t-tests and logistic regression with linear continuous MFI are two simpler and yet statistically more powerful alternatives. However, there are valid concerns regarding the use of linear models. A linear model may obscure threshold effects or other non-linear relationships due to its assumption that a 1-unit change at low MFI values has the same association with disease risk as a 1-unit change at high MFI values. The validity of these assumptions could be assessed using predicted probability plots in an exploratory sub-dataset as mentioned above. Alternatively, analyzing the data by quartiles would allow the assessment of both threshold effects and dose response relationships at the same time. The use of non-dichotomized predictors also makes interpretation of causal inference challenging. For example, the meaning of "HPV seropositivity increases the odds of developing a cancer two-fold" is more straightforward than "a 1 unit increase in HPV lnMFI increases the odds of developing a cancer two-fold." For this reason, linear models are probably best suited for research attempting to establish, rather than quantify, a correlation between an infection and cancer status. In addition, use of an empirical positive study may have better informed the utility of the different analysis and data visualizations methods than the null study to which we had access. Unfortunately, despite requests to three independent researchers, we were unable to obtain a dataset from a published positive study of an HPV associated cancer.

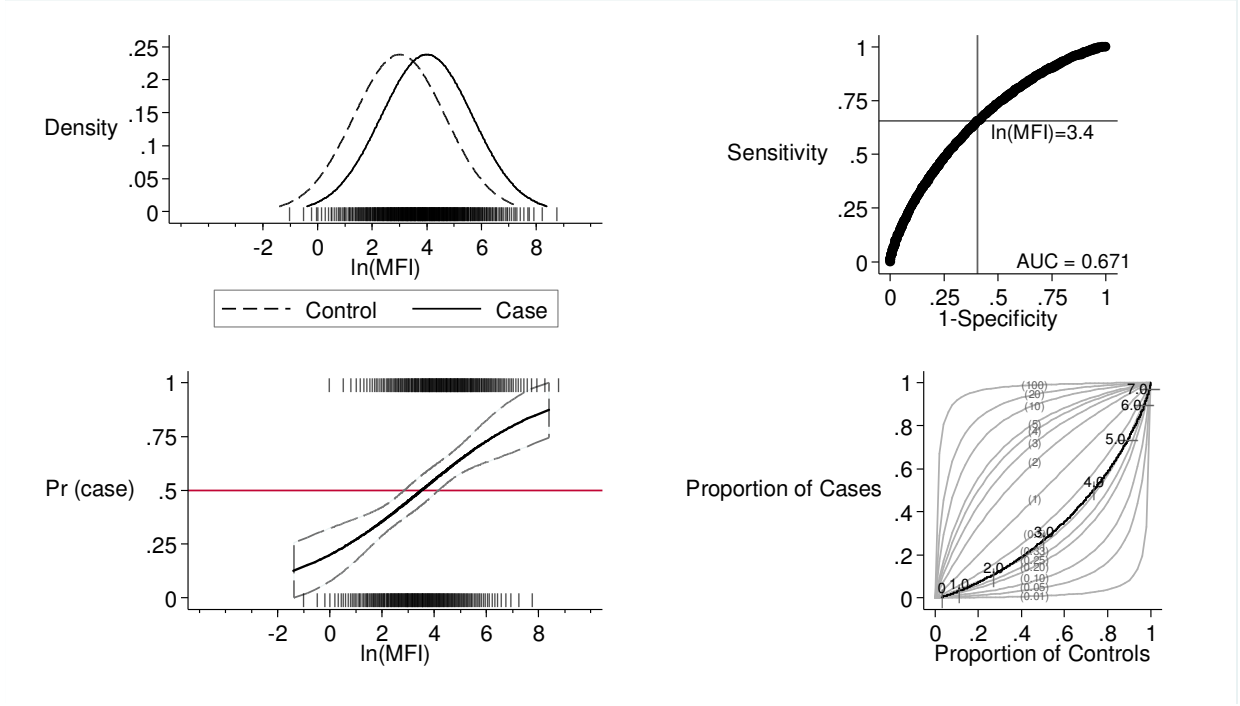
In conclusion, based upon type I error, statistical power, and bias, both logistic regression with continuous linear MFI and unpaired t-tests were superior to logistic regression with dichotomized MFI. Data splitting should be considered if visualizations are to inform selection of cut-points for causal inference.

Figure 1.1. Graphical representations of the association between MFI and disease status in simulated data with a mean difference of 0.



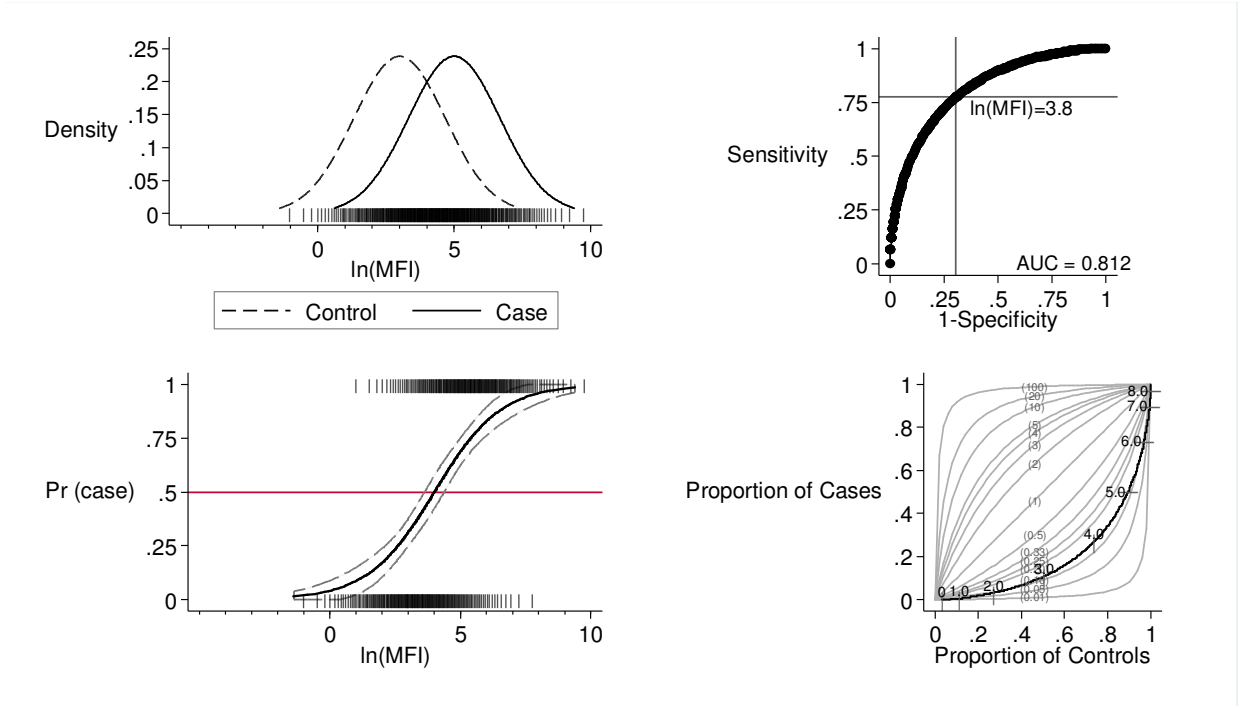
Top left. Overlapping kernel density plots. *Top right*. An ROC curve with reference lines noting the optimal cut-point based on the maximum J-Statistic. *Bottom left*. A predicted probability plot based on a restricted cubic spline of MFI, with three knots. The plot perfectly aligns with the horizontal reference line for no association and the dashed lines outline the 95% confidence interval. *Bottom right*. A Q-Q plot of the cumulative distribution function for cases (y-axis) and controls (x-axis) in association with lnMFI follows the line representing an OR=1.

Figure 1.2. Graphical representations of the association between MFI and disease status in simulated data with a mean difference of 1.



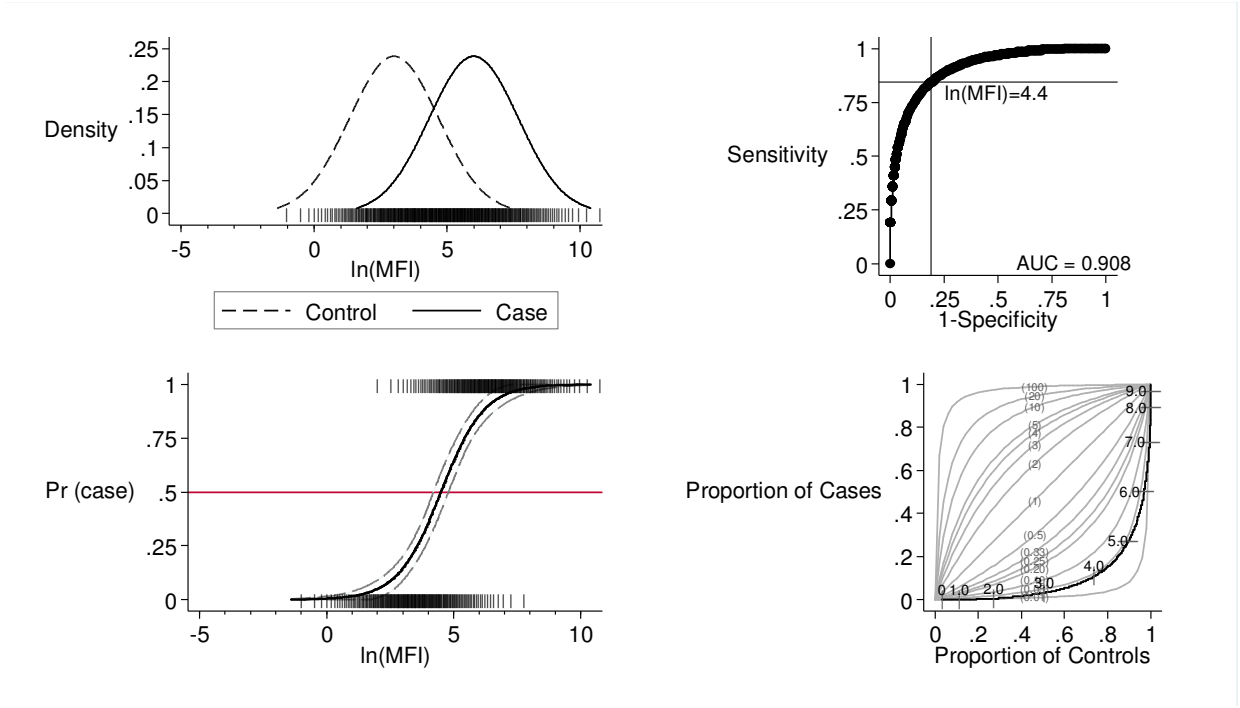
Top left. Overlapping kernel density plots. *Top right*. An ROC curve with reference lines noting the optimal cut-point based on the maximum J-Statistic. *Bottom left*. A predicted probability plot based on a restricted cubic spline of MFI, with three knots. The dashed lines outline the 95% confidence interval and the horizontal line is the reference for no association. *Bottom right*. A Q-Q plot of the cumulative distribution function for cases and controls.

Figure 1.3. Graphical representations of the association between MFI and disease status in simulated data with a mean difference of 2.



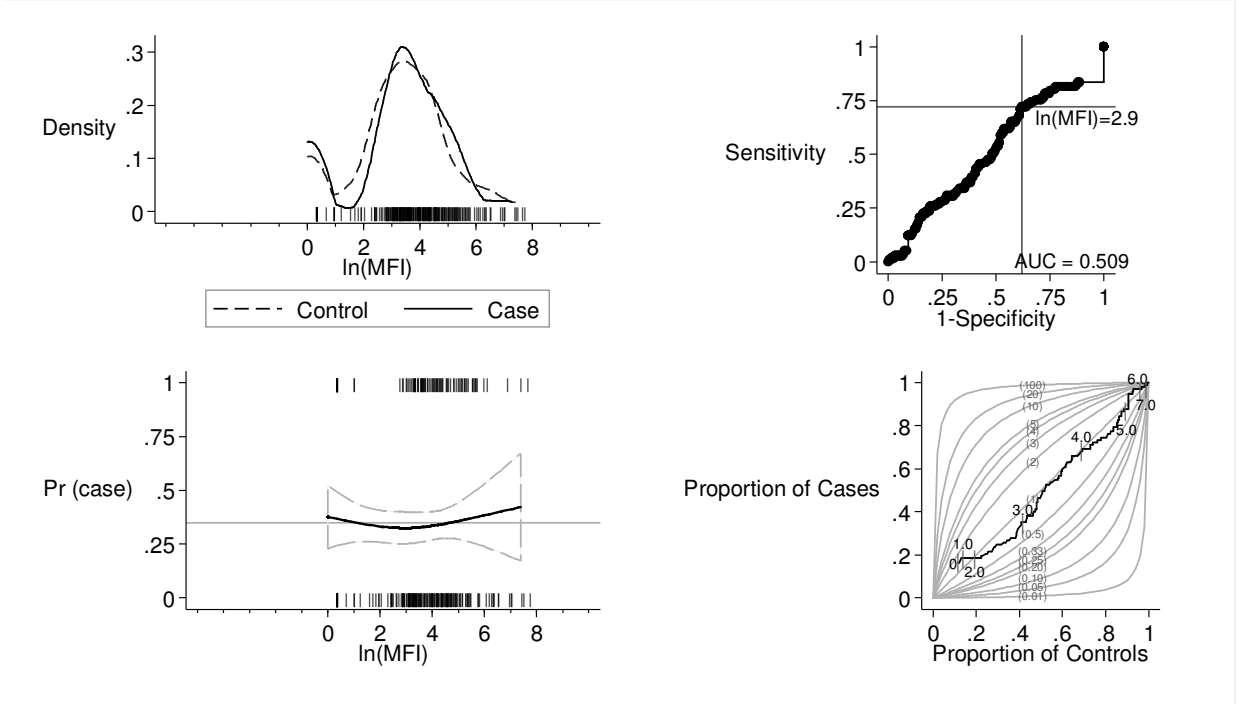
Top left. Overlapping kernel density plots. *Top right*. An ROC curve with reference lines noting the optimal cut-point based on the maximum J-Statistic. *Bottom left*. A predicted probability plot based on a restricted cubic spline of MFI, with three knots. The dashed lines outline the 95% confidence interval and the horizontal line is the reference for no association. *Bottom right*. A Q-Q plot of the cumulative distribution function for cases and controls.

Figure 1.4. Graphical representations of the association between MFI and disease status in simulated data with a mean difference of 3.



Top left. Overlapping kernel density plots. *Top right*. An ROC curve with reference lines noting the optimal cut-point based on the maximum J-Statistic. *Bottom left*. A predicted probability plot based on a restricted cubic spline of MFI, with three knots. The dashed lines outline the 95% confidence interval and the horizontal line is the reference for no association. *Bottom right*. A Q-Q plot of the cumulative distribution function for cases and controls.

Figure 1.5. Graphical representations of the association between HPV-16 L1 MFI and hyperplastic polyps in the Minnesota Cancer Prevention Research Unit Polyp Study.



Top left. The kernel density plot illustrates the probability density function of log-normal MFI curves for cases and controls. The tick marks represent the density of the data informing the curves. *Top right*. The ROC curve has reference lines noting the sensitivity and 1-specificity corresponding to lnMFI 2.9, the optimal cut-point based on the maximum J-Statistic. *Bottom left*. A predicted probability plot based on a restricted cubic spline of MFI, with three knots. The dashed lines outline the 95% confidence interval. Tick marks depict the density of the data informing the curve. The horizontal line is the reference for no association. (#cases/(#cases+#controls)). *Bottom right*. The Q-Q plot depicts the cumulative distribution function for cases (y-axis) and controls (x-axis) in association with varying lnMFI. The heavy curve represents lnMFI values, with tick marks noting one-unit increments. The distance between tick marks is proportional to the amount of MFI data within that range and the light curves are constant odds ratios.

Table 1.1. Estimated type 1 error rate and power to detect differences in MFI between cases and controls using logistic regression and an unpaired t-test in simulated datasets.[a]

	Logistic regression of dichotomized MFI[b]					t-test
Mean difference	None[c]	200 MFI[d]	400 MFI[e]	mean + 2 SD[f]	J-Statistic[g]	
			Type 1 Error Rate			
0[h]	0.047	0.041	0.030	0.039	0.271	0.048
			Statistical Power			
1	1.000	0.973	0.858	0.759	1.000	1.000
2	1.000	1.000	1.000	1.000	1.000	1.000
3	1.000	1.000	1.000	1.000	1.000	1.000
4	1.000	1.000	1.000	1.000	1.000	1.000
5	1.000	1.000	1.000	1.000	1.000	1.000
6	1.000	1.000	1.000	1.000	1.000	1.000

[a] 10,000 simulations of 200 controls (mean=3, SD=1.6) and 200 cases (mean=3 + mean difference, SD=1.6).
[b] MFI is median fluorescence intensity, a measure of the strength of an antibody response.
[c] Linear continuous form of lnMFI.
[d] Less than or equal to 200 MFI was considered unexposed.
[e] Less than or equal to 400 MFI was considered unexposed.
[f] Less than or equal to the mean lnMFI of the controls plus two standard deviations was considered unexposed.
[g] Less than the lnMFI corresponding to the maximum J-statistic (Sensitivity + Specificity – 1) was considered unexposed.
[h] This row, where the null hypothesis is true, estimates the type I error rather than power.

Table 1.2. The association of HPV 16 L1 MFI[a] with hyperplastic polyps in the Minnesota Cancer Prevention Research Unit Polyp Study[b].

MFI[a] cut-point	OR	Mean Difference	95% CI	P
None (continuous)[c]	1.00		0.87-1.15	0.99
200 MFI[d]	0.76		0.31-1.91	0.57
400 MFI[e]	0.62		0.16-2.35	0.48
Mean among controls + 2 SD[f]	0.75		0.14-3.96	0.74
J-Statistic[g]	1.59		0.93-2.72	0.09
t-test		0.00	-0.44-0.45	0.99

[a] MFI is median fluorescence intensity, a measure of the strength of an antibody response.
[b] Cancer Epidemiol Biomarkers Prev 2012;21:1599–601.
[c] OR is per 1 unit change in lnMFI.
[d] Less than or equal to 200 MFI was considered unexposed.
[e] Less than or equal to 400 MFI was considered unexposed.
[f] Less than or equal to the mean lnMFI of the controls plus two standard deviations was considered unexposed.
[g] Less than the lnMFI corresponding to the maximum J-statistic (Sensitivity + Specificity - 1) was considered unexposed.

Chapter 2: Prior Human Polyomavirus and Papillomavirus Infection and Incident Lung Cancer: A Nested Case-Control Study

Abstract

Background: Some lung cancers may have an infectious etiology. As oncogenic viruses with the ability to infect the respiratory track, human polyomaviruses (HPyV) and human papillomaviruses (HVP) are candidate pathogens.

Methods: We performed a nested case-control study, testing serum from the Carotene and Retinol Efficacy Trial (CARET), conducted 1985-2005, for antibodies to Merkel cell (MCPyV), KI (KIV), and WU (WUV) HPyVs as well as to six high-risk and two low-risk HPV types. Cases (n=200) were participants with incident lung cancer. Controls (n=200) were frequency matched on age at enrollment, year of enrollment, date of serum collection, intervention arm assignment (β-carotene plus retinyl palmitate supplementation vs. placebo), and the number of serum freeze-thaw cycles. Sera were tested using multiplex liquid bead microarray antibody assays. We used logistic regression to assess the association between HPyV and HPV antibodies and lung cancer. We also performed linear trend tests for all antibody measures.

Results: There was no evidence of an association between levels of MCPyV, KIV, or WUV antibodies and incident lung cancer (P>0.10 for all trend tests; odds ratio (OR) range 0.59 to 1.22, P>0.10 for all). There was also no clear suggestion of an association between infection with HPV 16 or 18 and the risk of lung cancer (P≥0.10 for all trend tests; OR range 0.25 to 2.54, P>0.10 for all), but the number of persons with serologic evidence of these infections was small.

Conclusion: Prior infection with any of several types of HPyV or HPV was not associated with subsequent diagnosis of lung cancer. Infection with these viruses likely does not influence a person's risk of lung cancer.

2.1 Introduction

Carcinoma of the lung is the most commonly diagnosed cancer worldwide and the only cancer among the top ten leading causes of death globally (25). Seven viruses are strongly associated with the incidence of human cancers (26) and most of these cancers occur at increased rates in immune deficient populations (27). The lung's propensity for infection and increased lung cancer rates in immunocompromised patients (28) suggest that viral infections may contribute to lung cancer risk.

Merkel cell (MCPyV), KI (KIV), and WU (WUV) polyomaviruses have been examined in association with lung cancer due to their membership in a carcinogenic viral family (6,29) and their ability to infect the lower respiratory tract (30–32). The existing literature is informative and yet limited in important ways. Prior DNA-based studies of MCPyV have been small, but have reported prevalences of MCPyV DNA in tumors between 5% and 39% (33–36). The evidence for an association between KIV or WUV and lung cancer is not consistent. Though an Italian study reported finding KIV DNA in 45% (9/20) of lung tumors compared to 5% (1/20) of adjacent normal tissues (37), other studies reported finding no KIV or WUV DNA in lung tumors (38,39). To our knowledge, there have been no seroepidemiologic studies of the association between polyomavirus infection and lung cancer. This is a limitation because, unlike nucleic acid amplification test (NAAT) based studies, prospective seroepidemiologic studies may be able to detect the association of viruses that initiate carcinogenesis, but whose viral DNA is no longer detectable in the tumors. Such "hit-and-run" mechanisms have been proposed for the role of MCPyV in the development some Merkel Cell Carcinomas (40) and may be present in other human polyomavirus (HPyV) associated cancers.

Human papillomavirus (HPV) infections are associated with up to 35% of oropharyngeal cancers (41). In addition, HPV 6 and 11 are involved in the formation of respiratory papillomas

(42), with occasional malignant transformation of infected cells (43). Based on this carcinogenic potential in the respiratory tract, previous studies have sought to evaluate the association between lung cancer and HPV infection. A 2009 meta-analysis (44) and systematic review (45) evaluated the accumulated evidence and independently concluded that HPV may be a risk factor for some histologies of lung cancer. However, both manuscripts noted the substantial heterogeneity in the reported data and argued that further studies were needed.

The vast majority of previous studies of the association of these viruses with lung cancer utilized a cross-sectional design, and so were unable to establish the temporal sequence of viral infection and incident cancer. We therefore conducted a nested case-control study of lung cancer within a longitudinal study to assess antibodies to MCPyV, KIV, WUV, and eight HPV types, using liquid bead microarray antibody (LBMA) assays.

2.2 Methods

2.2.1 Study population

The Carotene and Retinol Efficacy Trial (CARET) was a randomized double-blind, multicenter chemoprevention trial, which tested whether supplementation with retinyl palmitate (25,000 ID/day) in combination with β-carotene (30 mg/day) could reduce lung cancer incidence among asbestos exposed participants (n=4,060) and smokers with 20 or more pack-years of exposure (n=14,254) (46). Participants were recruited from study centers in Seattle, Washington; Portland, Oregon; Irvine, California; San Francisco, California; Baltimore, Maryland; and Groton, Connecticut. The last three locations primarily enrolled asbestos exposed participants. The trial began in May 1985 and was stopped on January 18, 1996 due to evidence of increased risk among those receiving supplementation. Follow-up activities continued through June 30, 2005.

Details of the CARET study have been previously described (47). Briefly, original CARET staff collected health histories, demographic data, and anthropomorphic measurements. Annual in person and semi-annual telephone-based interviews elicited information on relevant signs, symptoms, and new medical diagnoses. Participants also received brief physical exams during the yearly study center visit. Participants provided baseline (pre-randomization) serum, and additional sera samples were collected at 2-year intervals thereafter. All sera were stored at -70°C at the CARET Coordinating Center at the Fred Hutchinson Cancer Research Center in Seattle, WA. We excluded all asbestos exposed participants from this analysis.

All participants provided signed informed consent, and the institutional review boards at each trial center reviewed and approved CARET activities annually (46). The institutional review boards of the University of Washington and the Fred Hutchinson Cancer Research Center approved this current analysis.

2.2.2 Case definition

Through February 28, 1998, the CARET end-points review committee obtained clinical records and pathology or cytology specimens for independent review by the CARET pathologist (47). Three independent physician adjudicators determined the origin, location, histology and date of lung cancer diagnosis (48). Beginning March 1, 1998, pathology reports from diagnosing institutions, without independent specimen review by the CARET pathologist, were reviewed by the adjudicators (48). After October 1, 1998, CARET endpoint specialists reviewed the pathology reports, with adjudication by a single independent physician (48). Searches of local cancer registries and the National Death Index were used to identify cases among those lost to follow-up. Self-report was not considered adequate evidence of case status.

Cases were defined as individuals with incident lung cancer of any histology (small cell carcinoma (SCLC), adenocarcinoma (ADC), squamous cell carcinoma (SCC), large cell carcinoma (LCLC), non-small cell lung cancer not otherwise specified (NSCLC, NOS), and unknown). We randomly selected case specimens (n=200) among participants who were free of all cancers prior to lung cancer diagnosis and who had serum available, with no more than two freeze-thaw cycles, from a blood draw 366 to 1095 days prior to diagnosis.

Controls were defined as participants free of cancer at the time of lung cancer diagnosis of the matched case. We selected control specimens (n=200), using incidence density sampling with replacement, among participants who had available serum with no more than two freeze-thaw cycles. Controls were frequency matched on age at enrollment (five year age groups), intervention arm assignment, year of enrollment, date of blood draw (six month intervals), and the number of serum freeze / thaw cycles.

2.2.3 Exposure measurement

We performed a liquid bead microarray antibody assay following an established protocol (2,49) with previously described modifications (6). We used a Bio-Plex 200 instrument (Bio-rad Laboratories, Hercules, CA) to obtain the median fluorescence intensity (MFI), a measure of the strength of antibody response. We tested sera for antibodies against the primary structural protein (VP1) and the small T antigen (ST-Ag), an oncoprotein, of MCPyV, KIV, and WUV. We also tested for antibodies against the major structural protein (L1) of six high-risk (16, 18, 31, 33, 52, and 58) and two low-risk (6 and 11) HPV types. In addition, we tested for antibodies to the E6 and E7 oncoproteins of HPV types 16 and 18. Because of the expected high prevalence of BK seropositivity (50), antibodies to BK polyomavirus VP1 antigens served as a positive control and glutathione S-transferase (GST)/ "Tag" (51) fusion proteins served as a negative control. According to previously described criteria (MFI>400) (50), 91.3% of our sera were seropositive for BK VP1. Existing fusion proteins were used for all antigens except for KIV ST-Ag (NCBI Reference Sequence: NC_009238.1) and WUV ST-Ag (NCBI Reference Sequence: NC_009539.1). We designed novel fusion proteins with the "Tag" 11-amino acid sequence on the C-terminus and expressed them in pEX-N-GST vectors (Blue Heron Bio, Bothell, WA) so that GST was fused to the N-terminus.

2.2.4 Statistical methods

Because prior infection with these viruses is nearly ubiquitous (6,52), we evaluated quartiles of HPyV MFI levels. The lowest quartile served as the referent in logistic regression analyses. In order to maximize our study's comparability with previous studies of HPV using LBMA, we defined HPV seropositivity as >400 MFI in our primary logistic regression analysis (2,19,53), and as >200 MFI in a sensitivity analysis (5,53). We also performed logistic regression linear trend tests for the association between lung cancer and the MFI for each viral antibody. In our HPyV analysis, we assessed MCPyV (VP1 & ST-Ag), KIV (VP1 & ST-Ag), and WUV (VP1 & ST-Ag) antibodies as six distinct exposures. HPV infection was grouped into four categories: HPV-

16 (E6, E7, and L1), HPV-18 (E6, E7, and L1), other high-risk HPV (31, 33, 52, 58 L1), and low-risk HPV (6, 11 L1). MFI were natural log transformed (lnMFI) to improve normality and we adjusted all logistic regression analyses for matched variables. As an exploratory exercise, we used boxplots to assess the association of individual histology types (SCLC, ADC, SCC, LCLC, NSCLC, NOS, unknown) with antigen-specific MFI.

We used permutation tests with 10,000 permutations to correct *P*-values for multiple comparisons. We assessed effect modification by smoking history (pack-years) and sex using likelihood ratio tests. Family history of lung cancer (yes, no), smoking history (pack-years), and sex were considered potential confounders. Confounders were retained in the final model only if inclusion in the model changed the odds ratios (OR) of interest by ≥10% and if they were not found to be effect modifiers. Analyses used two-sided statistical tests and were performed with Stata/IC 13.1 (StataCorp LP, College Station, TX).

2.3 Results

There were almost no differences between cases and controls with respect to year of enrollment, year of blood draw, and the number of serum freeze / thaw cycles (data not shown) as well as age at enrollment and intervention arm assignment (Table 2.1), suggesting frequency matching was successful. The median age at enrollment was 61.6 (interquartile range 57.2 – 65) and 58.5% of sera were from those assigned to the intervention arm. The sampled population was approximately 41% female and 95% White. More than 50% of cases and controls had at least a college education and approximately 71% were married. The distribution of BMI was different between cases and controls (P=0.034), with fewer cases (65.3%) being overweight or obese than controls (73.5%). Though not statistically different, more than 75% of cases were current smokers at enrollment compared to 67% of controls. Though all participants had at least 20 pack-years of smoking exposure, on average cases had higher-level exposures (P<0.001). A family history of lung cancer was reported by almost twice as many cases (16%) as controls (8.5%) (P=0.032). Asthma was reported by nearly 9% of participants and tuberculosis by 1% and did not differ between cases and controls. Chronic bronchitis or emphysema was reported by 22% of cases and 15.5% of controls (P=0.095). A history of pneumonia was reported by 30% of cases and 21.5% of controls (P=0.051).

With regard to histology, 22% (n=44) of cases were diagnosed as SCLC, 73% (n=146) as NSCLC, and 5% (n=10) as unknown. Among NSCLC cases, 45% (n=66) were diagnosed as ADC, 25% (n=37) as LCLC, 21% (n=31) as SCC, and 8% (n=12) as NSCLC, NOS (data not shown).

The distributions of case and control lnMFI were similar for all assessed antibodies (Table 2.2).

Analysis of HPyV MFI quartiles provided no evidence of an association with lung cancer (Table 2.3). With the lowest quartile as the referent, odds ratios ranged from 0.72 (95% CI: 0.41-1.26,

P-corrected=0.33) for the highest quartile of MCPyV ST-Ag to 1.22 (95% CI: 0.70-2.13, *P*-corrected=0.45) for the third quartile of WUV ST-Ag. HPyV trend tests were not statistically significant and family history of lung cancer, smoking, and sex were not effect modifiers or confounders (data not shown).

Based upon a cut-point of 400 MFI, there was no evidence of an association between HPV seropositivity and incident lung cancer (Table 2.4). Odds ratios ranged from 0.25 (95% CI: 0.08-0.77, *P*-corrected=0.14) for HPV 16 L1 to 2.54 (95% CI: 0.49-13.3, *P*-corrected=0.16) for HPV 16 E6. Sensitivity analyses, with a seropositivity cut-point of 200 MFI, also showed no evidence of an association (Table 2.5). Trend tests for the four HPV categories, and individual HPV antibodies, were not statistically significant. Family history of lung cancer, smoking, and sex were not effect modifiers or confounders (data not shown).

We found no evidence of an association between specific HPyV or HPV antibodies and any individual histologic type of lung cancer (Figures 2.1a, 2.1b, and 2.1c).

2.4 Discussion

In this primarily Caucasian population of heavy smokers, we found no evidence of an association between HPyV antibodies or HPV seropositivity and incident lung cancer, whether considered as a whole or as individual histologic types.

Although to our knowledge there have been no previous seroepidemiologic studies of the association between HPyV infection and lung cancer, some prior NAAT based studies have reported associations of MCPyV and KIV DNA with lung tumors. An American study reported a prevalence of 16.7% (5/30) for MCPyV DNA in NSCLC compared to 9.5% (2/21) in benign adjacent tissue (34), a difference that was not statistically significant (*P*=0.466). In addition, a German study of MCPyV in SCLC reported that 39% (7/18) of lung tumors had MCPyV DNA compared with 0% (0/18) of controls (33). Though this was statistically significant (*P*=0.003), controls were blood samples rather than lung tissue, so a true referent was missing. A Chilean study that reported 4.7% (4/86) prevalence among ADC and SCC lacked controls entirely (35), as did a Japanese study that reported a prevalence of 17.9% (20/112) (36). Therefore, our null results suggest that MCPyV DNA may be present in healthy as well as cancerous lung tissue. It is also possible that many of the DNA positive lung specimens represent transient infections, which are unrelated to lung cancer initiation. As in the US study mentioned above, an Italian study that reported a positive association for KIV DNA in lung tumors used surrounding normal tissue as the controls (37). However, that study was small (n=40) and other studies of KIV and WUV in lung tumors found no evidence of infection with these viruses in lung tumors (38,39).

Our type-specific HPV results are consistent with the null results of several studies: 1) a Finnish nested case-control study of HPV 16 and 18 infections and female lung cancer (54); 2) the nested case-control portion of a recently published large European study of HPV 6, 11, 16, 18, or

31 antibodies and lung cancer (55); and 3) robust NAAT-based studies of a variety of HPV types and lung cancer in Western populations (55,56).

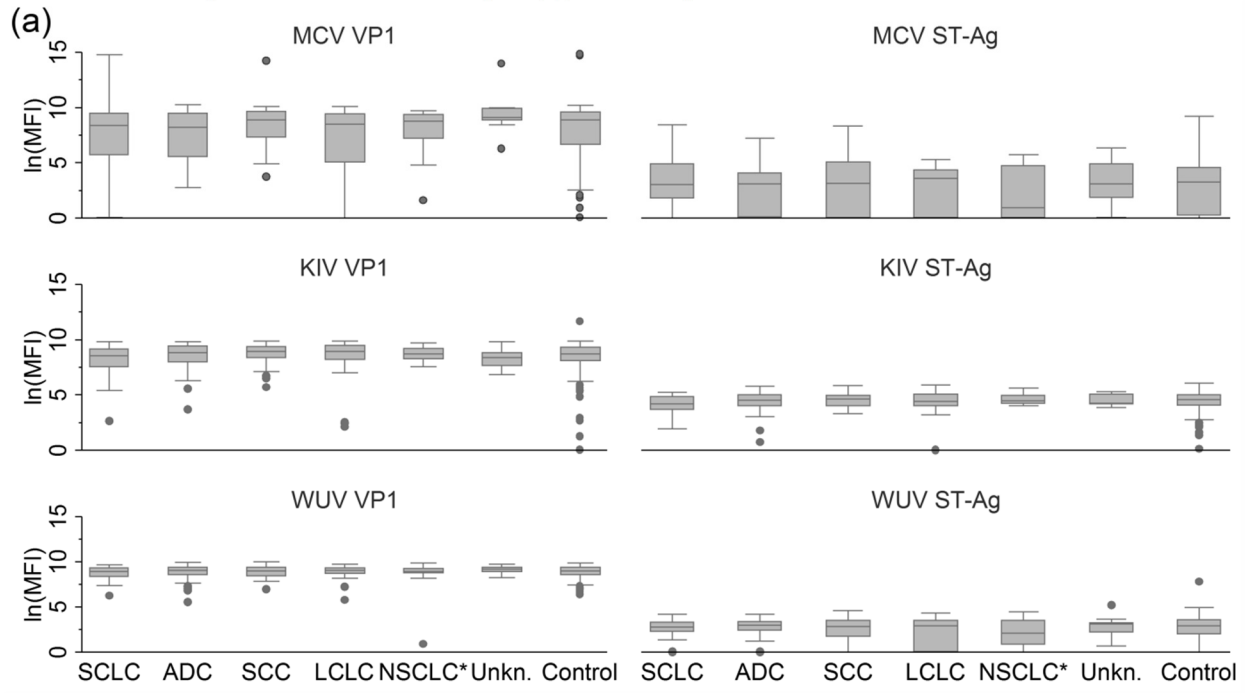
In 2012, a new systematic review and meta-analysis reported substantial heterogeneity in the literature assessing HPV's association with lung cancer (57). The author reported that the majority of this variation could be explained by geography, with stronger associations reported in Asia than in Australia, Europe, and North America (57). Similarly, a previous meta-analysis found only 50% (7/14) of European studies reported a positive association, compared to 78% (14/18) of Asian studies (44). It is possible that there are true regional differences in HPV's association with lung tumors due to variation in sexual practices (e.g., engagement in oral sex (56)) or other exposures. However, the prior meta-analysis also noted considerable intraregional variation (44), with the reported prevalence of HPV in lung tumors ranging from 11.8% to 55% in China and 0% to 78.3% in Japan (58–61). Laboratory, regional, and temporal differences in efforts to reduce specimen contamination may also have contributed to the heterogeneity (44). For example, both of the large null European NAAT-based studies mentioned above took extensive precautions to avoid DNA contamination (55,56). Therefore, prior reported associations of HPV and lung cancer should be interpreted cautiously.

This study has several limitations. First, all participants were current or former heavy smokers and the carcinogenic effect of this tobacco exposure may have overwhelmed our ability to detect important, but weaker associations with viral infections. Because there is evidence that HPyVs may induce cancers through interaction with known carcinogens (62) and cigarette smoking can induce squamous metaplasia in the lower respiratory tract, giving rise to squamocolumnar junctions that are suitable for HPV infection and development of associated lesions (63), we hypothesized that an association between prior viral infection and lung cancer incidence in a population of current and former cigarette smokers might be particularly strong. However, the

relationship between prior infection with these viruses and lung cancer may differ in a non-smoking population. Second, while we tested for L1 antibodies to all included HPV types, antibodies to viral oncoproteins E6 and E7 were only assessed for HPV 16 and 18. Since L1 antibodies are markers of HPV infection, rather than specific to HPV-induced cancers, our ability to determine an association with the remaining six HPV types was comparatively weaker. Yet, in the event that one of those HPV types was associated with lung cancer, some signal would still be expected, albeit potentially muted. Third, serology may be less sensitive than nucleic acid amplification-based tests. However, based upon a reported concordance of 93% for HPV 16 DNA and HPV 16 seropositivity in oropharyngeal tumors (64), this may not have substantially affected the observed outcome. Fourth, our assay is unable to determine the site of infection. If possible, future similar studies should be conducted using biorepositories with available tumor specimens to confirm positive serologic associations. Fifth, although HPyV analyses were adequately powered, our study lacked sufficient statistical power to examine a potential role of infection with HPV 16 and 18. Sixth, multiple comparisons inflated the type 1 error rate. For example, based upon 19 seropositive participants, HPV 16 L1 had a statistically significant association with lung cancer. However, when permutation tests were used to correct *P*-values for multiple comparisons, the OR was no longer significant (P=0.14), suggesting this finding was likely due to chance.

In our population of heavy smokers from the U.S., there was no evidence of an association between HPyV antibody levels or prior HPV infection and the development of lung cancer. These findings, in conjunction with broadly-similar findings in other studies (55,56), suggest that neither HPV nor HPyV infections are associated with lung cancer in Western populations.

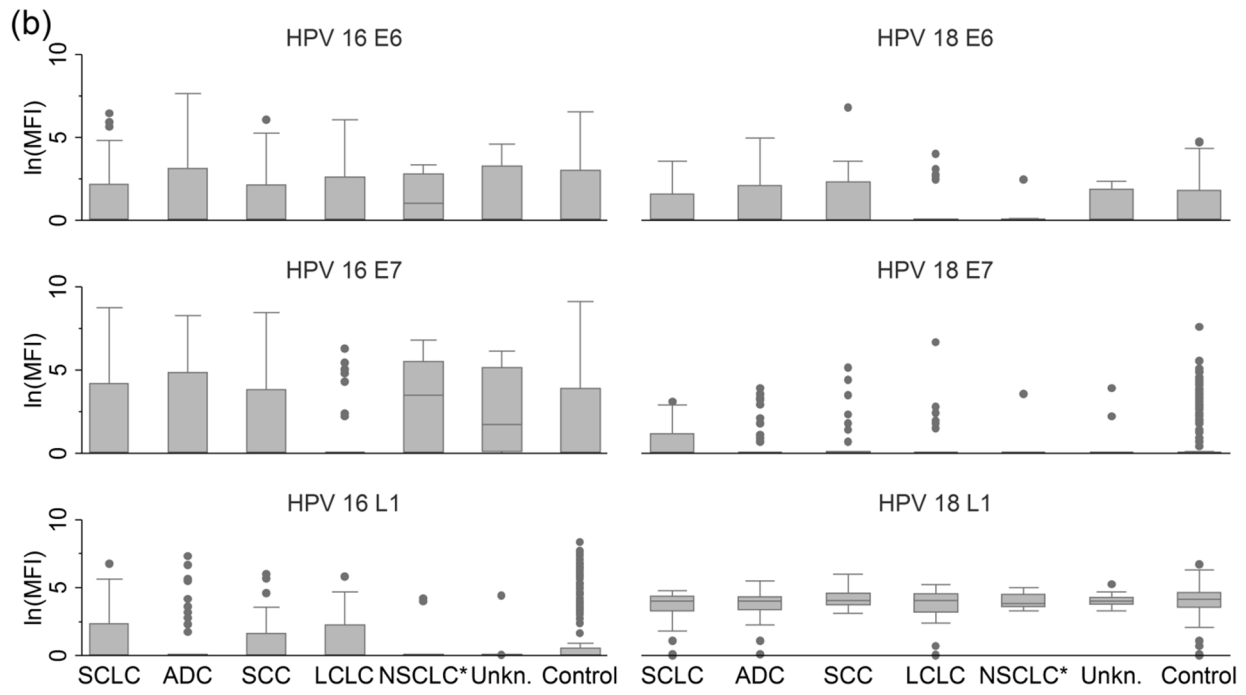
Figure 2.1a. Boxplots[a] of human polyomavirus (HPyV) antigen specific antibody[b] distributions by individual histologic type of lung cancer.



[a] The top and bottom of each shaded box represents the inter-quartile range (IQR), that is, the 25th to 75th percentiles. The horizontal line within the box represents the median and the vertical lines coming out of the box extend to 1.5 times the IQR, with dots representing outliers.
[b] Measured in units of median fluorescence intensity (MFI), a measure of the strength of an antibody response. MFI were natural log transformed to improve normality.

Abbreviations: SCLC, small cell lung cancer; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LCLC, large cell lung cancer; NSCLC*= NSCLC, not otherwise specified; Unkn.=unknown

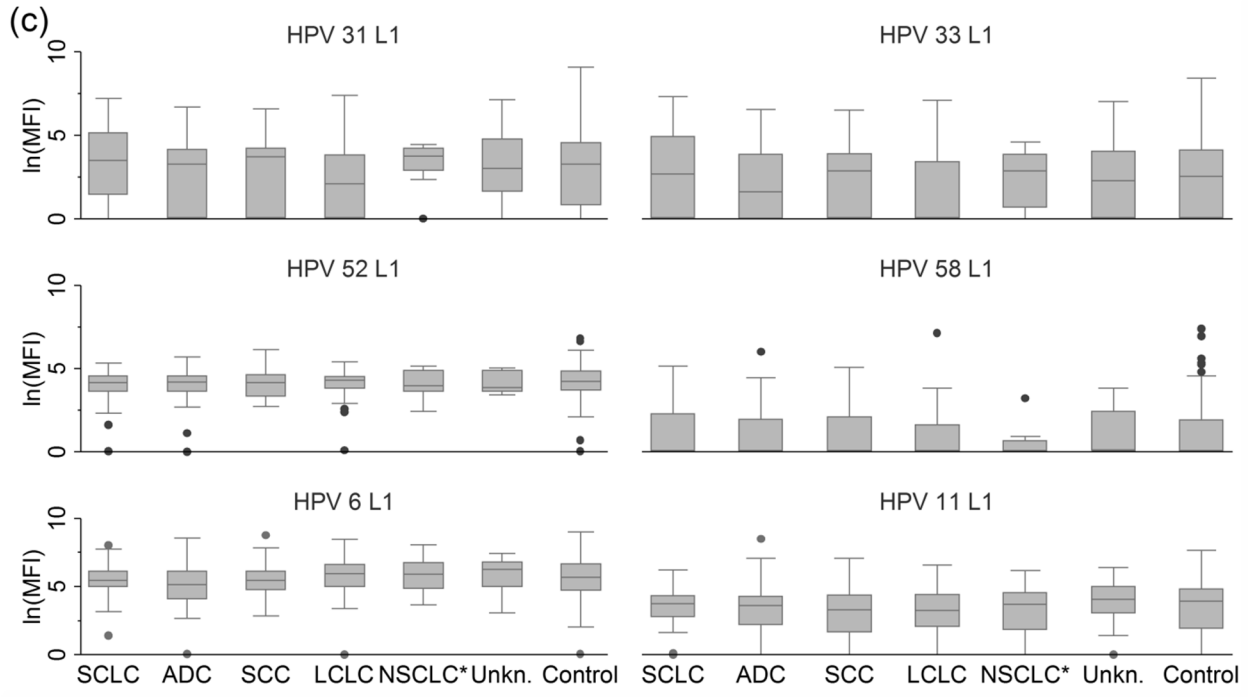
Figure 2.1b. Boxplots[a] of human papillomavirs (HPV) 16 and 18 antigen specific antibody[b] distributions by individual histologic type of lung cancer.



[a] The top and bottom of each shaded box represents the inter-quartile range (IQR), that is, the 25th to 75th percentiles. The horizontal line within the box represents the median and the vertical lines coming out of the box extend to 1.5 times the IQR, with dots representing outliers.
[b] Measured in units of median fluorescence intensity (MFI), a measure of the strength of an antibody response. MFI were natural log transformed to improve normality.

Abbreviations: SCLC, small cell lung cancer; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LCLC, large cell lung cancer; NSCLC*= NSCLC, not otherwise specified; Unkn.=unknown

Figure 2.1c. Boxplots[a] of high-risk (31, 33, 52 and 58) and low-risk (6 and 11) human papillomavirs (HPV) antigen specific antibody[b] distributions by individual histologic type of lung cancer.



[a] The top and bottom of each shaded box represents the inter-quartile range (IQR), that is, the 25th to 75th percentiles. The horizontal line within the box represents the median and the vertical lines coming out of the box extend to 1.5 times the IQR, with dots representing outliers.
[b] Measured in units of median fluorescence intensity (MFI), a measure of the strength of an antibody response. MFI were natural log transformed to improve normality.

Abbreviations: SCLC, small cell lung cancer; ADC, adenocarcinoma; SCC, squamous cell carcinoma; LCLC, large cell lung cancer; NSCLC*= NSCLC, not otherwise specified; Unkn.=unknown

Table 2.1. Characteristics of selected incident lung cancer cases and frequency matched controls from the Beta-Carotene and Retinol Efficacy Trial (CARET) for lung cancer chemoprevention.

	Cases (n=200)		Controls (n=200)	
	N	**(%)**	**N**	**(%)**
Age at enrollment (years)				
50-54	34	(17.0)	35	(17.5)
55-59	42	(21.0)	41	(20.5)
60-64	75	(37.5)	74	(37.0)
65-70	49	(24.5)	50	(25.0)
Intervention arm[a]	117	(58.5)	117	(58.5)
Female sex	78	(39.0)	85	(42.5)
White race	189	(94.5)	191	(95.5)
Education level[b]				
Grade school	3	(1.8)	6	(3.5)
High school	65	(38.0)	52	(30.6)
College	88	(51.5)	85	(50.0)
Graduate school	15	(8.8)	27	(15.9)
Married	144	(72.4)	140	(70.4)
Body Mass Index (BMI) (kg/m²)				
Underweight (<18.5)	4	(2.0)	1	(0.5)
Normal (18.5-24.99)	65	(32.7)	52	(26.0)
Overweight (25-29.99)	93	(46.7)	88	(44.0)
Obese (≥30)	37	(18.6)	59	(29.5)
Current smoker	151	(75.5)	134	(67.0)
Pack-years of smoking				
20-35	21	(10.5)	38	(19.0)
35-50	69	(34.5)	91	(45.5)
50-65	44	(22.0)	37	(18.5)
65+	66	(33.0)	34	(17.0)
Years since quitting smoking				
<1	4	(8.2)	7	(10.6)
1-2	24	(49.0)	22	(33.3)
3-4	12	(24.5)	21	(31.8)
5-6	9	(18.4)	16	(24.2)
Family history of lung cancer	32	(16.0)	17	(8.5)
History of respiratory illness				
Asthma	20	(10.0)	14	(7.0)
Tuberculosis	2	(1.0)	2	(1.0)
Chronic bronchitis or emphysema	44	(22.0)	31	(15.5)
Pneumonia	60	(30.0)	43	(21.5)

[a] Supplementation with retinyl palmitate in combination with β-carotene vs. placebo
[b] Highest educational level started, missing 15% of the data

Table 2.2. The distribution of antigen specific antibodies[a] among cases and controls.

	Cases (n=200)		Controls (n=200)		
Antibody	**Mean**	**SD**	**Mean**	**SD**	**Difference**
HPyV					
MCPyV VP1	7.7	2.5	7.9	2.4	-0.2
MCPyV ST-Ag	2.8	2.2	3.0	2.2	-0.2
KIV VP1	8.5	1.3	8.4	1.4	0.1
KIV ST-Ag	4.4	0.9	4.5	0.8	-0.1
WUV VP1	8.8	0.9	8.9	0.7	-0.1
WUV ST-Ag	2.6	1.2	2.6	1.4	0.0
HPV16					
E6	1.4	1.9	1.4	1.8	0.0
E7	2.0	2.5	2.1	2.6	-0.1
L1	1.0	1.8	1.2	2.2	-0.2
HPV18					
E6	0.8	1.3	0.9	1.3	-0.1
E7	0.6	1.1	0.7	1.4	-0.1
L1	3.9	1.0	4.0	1.0	-0.1
Other high-risk HPV					
31 L1	2.8	2.1	3.1	2.3	-0.3
33 L1	2.3	2.2	2.4	2.4	-0.1
52 L1	4.0	1.0	4.2	1.0	-0.2
58 L1	1.0	1.5	1.0	1.6	0.0
Low-risk HPV					
6 L1	5.5	1.5	5.6	1.5	-0.1
11 L1	3.3	1.8	3.4	2.1	-0.1

[a] Measured in units of median fluorescence intensity (MFI), a measure of the strength of an antibody response. MFI were natural log transformed to improve normality.

Table 2.3. Association between antigen specific human polyomavirus (HPyV) antibodies[a] and incident lung cancer, adjusted for matching variables.

Antibody quartile	Mean lnMFI[b]	OR (95%CI)	*P*[c]	Trend Test OR (95%CI)[d]	*P*[c]
MCPyV VP1				0.96 (0.88-1.04)	0.34
1	4.24	Referent			
2	7.66	1.08 (0.62-1.89)	0.78		
3	9.20	0.85 (0.49-1.48)	0.61		
4	10.06	0.79 (0.45-1.37)	0.47		
MCPyV ST-Ag				0.98 (0.89-1.06)	0.66
1	0.04	Referent			
2	1.96	0.59 (0.34-1.04)	0.16		
3	3.85	0.75 (0.43-1.32)	0.39		
4	5.73	0.72 (0.41-1.26)	0.33		
KIV VP1				1.02 (0.89-1.18)	0.78
1	6.66	Referent			
2	8.43	0.92 (0.53-1.61)	0.79		
3	9.03	0.85 (0.48-1.48)	0.61		
4	9.63	1.09 (0.62-1.90)	0.76		
KIV ST-Ag				0.87 (0.69-1.10)	0.27
1	3.33	Referent			
2	4.23	0.79 (0.45-1.37)	0.44		
3	4.74	0.77 (0.44-1.35)	0.43		
4	5.33	0.80 (0.46-1.40)	0.50		
WUV VP1				0.94 (0.74-1.20)	0.63
1	7.78	Referent			
2	8.82	0.85 (0.48-1.49)	0.61		
3	9.19	1.20 (0.69-2.09)	0.48		
4	9.60	0.98 (0.56-1.71)	0.96		
WUV ST-Ag				0.97 (0.83-1.12)	0.70
1	0.69	Referent			
2	2.51	1.01 (0.58-1.75)	0.99		
3	3.17	1.22 (0.70-2.13)	0.45		
4	4.01	0.75 (0.43-1.31)	0.39		

[a] Measured in units of median fluorescence intensity (MFI), a measure of the strength of an antibody response.
[b] lnMFI = natural log transformed MFI
[c] *P*-values are corrected for multiple comparisons using permutation tests.
[d] The trend tests estimate the odds ratio for a one unit increase in natural log transformed MFI, adjusted for matched variables.

Table 2.4. Association between human papillomavirus (HPV) seropositivity, defined as >400 MFI[a], and incident lung cancer, adjusted for matching variables.

Antibody	Cases (n=200) %	Controls (n=200) %	OR (95% CI)	*P*[c]	Trend Test[b] OR (95% CI)	*P*[c]
HPV 16				0.17		0.87
E6					1.01 (0.91-1.12)	0.86
Positive	2.5	1.0	2.54 (0.49-13.3)	0.16		
E7					0.99 (0.92-1.07)	0.81
Positive	8.5	11.0	0.74 (0.38-1.46)	0.44		
L1					0.94 (0.85-1.03)	0.22
Positive	2.0	7.5	0.25 (0.08-0.77)	0.14		
HPV 18				1.00		0.65
E6					0.95 (0.82-1.11)	0.53
Positive	0.5	-	N/A	N/A		
E7					0.94 (0.80-1.09)	0.47
Positive	0.5	0.5	1.00 (0.06-16.7)	1.00		
L1					0.87 (0.71-1.06)	0.20
Positive	-	1.5	N/A	N/A		
Other high-risk HPV				0.57		0.90
31 L1					0.94 (0.86-1.03)	0.20
Positive	7.0	13.5	0.48 (0.24-0.95)	0.11		
33 L1					0.97 (0.89-1.05)	0.50
Positive	4.5	10.5	0.40 (0.18-0.90)	0.11		
52 L1					0.83 (0.68-1.03)	0.10
Positive	0.5	1.5	0.33 (0.03-3.18)	0.49		
58 L1					0.99 (0.87-1.12)	0.88
Positive	0.5	1.0	0.50 (0.04-5.85)	0.78		
Low-risk HPV				0.75		0.62
6 L1					0.93 (0.81-1.06)	0.31
Positive	34.5	40.5	0.77 (0.51-1.16)	0.28		
11 L1					0.97 (0.87-1.08)	0.57
Positive	6.5	6.0	1.09 (0.48-2.47)	0.83		

[a] MFI = median fluorescence intensity, a measure of the strength of an antibody response.
[b] The trend tests estimate the odds ratio for a one unit increase in natural log transformed MFI, adjusted for matched variables.
[c] *P*-values are corrected for multiple comparisons using permutation tests.

Table 2.5. Sensitivity analysis of the association between human papillomavirus (HPV) seropositivity and incident lung cancer, using an alternative cut-point of 200 MFI[a] and adjusted for matching variables.

Antibody	Cases (n=200) %	Controls (n=200) %	OR (95% CI)	*P*[b]
HPV 16				0.38
E6				
Positive	4.5	3.0	1.54 (0.53-4.42)	0.29
E7				
Positive	13.5	13.5	1.00 (0.56-1.77)	1.00
L1				
Positive	6.5	10.0	0.63 (0.30-1.3)	0.33
HPV18				0.58
E6				
Positive	0.5	-	N/A	N/A
E7				
Positive	0.5	1.0	0.50 (0.04-5.72)	0.80
L1				
Positive	1.5	5.5	0.26 (0.07-0.94)	0.18
Other high-risk HPV				0.64
31L1				
Positive	11.5	16.0	0.68 (0.38-1.21)	0.34
33L1				
Positive	11.0	13.5	0.79 (0.43-1.44)	0.41
52L1				
Positive	3.5	10.0	0.32 (0.13-0.78)	0.11
58L1				
Positive	1.0	2.0	0.50 (0.09-2.81)	0.58
Low-risk HPV				0.47
6L1				
Positive	56.5	62.0	0.79 (0.53-1.18)	0.35
11L1				
Positive	8.5	15.5	0.50 (0.27-0.94)	0.12

[a] MFI = median fluorescence intensity, a measure of the strength of an antibody response.

REFERENCES

1. Jones LP, Zheng H, Karron RA, Peret TCT, Tsou C, Anderson LJ. Multiplex assay for detection of strain-specific antibodies against the two variable regions of the G protein of respiratory syncytial virus. Clin Diagn Lab Immunol 2002;9:633–8.

2. Waterboer T, Sehr P, Michael KM, Franceschi S, Nieland JD, Joos TO, et al. Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. Clin Chem 2005;51:1845–53.

3. Clifford GM, Shin H-R, Oh J-K, Waterboer T, Ju Y-H, Vaccarella S, et al. Serologic response to oncogenic human papillomavirus types in male and female university students in Busan, South Korea. Cancer Epidemiol Biomarkers Prev 2007;16:1874–9.

4. Gu A, Xie Y, Mo H, Jia W, Li M-Y, Li M, et al. Antibodies against Epstein-Barr virus gp78 antigen: a novel marker for serological diagnosis of nasopharyngeal carcinoma detected by xMAP technology. J Gen Virol 2008;89:1152–8.

5. Michael KM, Waterboer T, Sehr P, Rother A, Reidel U, Boeing H, et al. Seroprevalence of 34 human papillomavirus types in the German general population. PLoS Pathog 2008;4:e1000091.

6. Carter JJ, Paulson KG, Wipf GC, Miranda D, Madeleine MM, Johnson LG, et al. Association of Merkel cell polyomavirus-specific antibodies with Merkel cell carcinoma. J Natl Cancer Inst 2009;101:1510–22.

7. Rowhani-Rahbar A, Carter JJ, Hawes SE, Hughes JP, Weiss NS, Galloway DA, et al. Antibody responses in oral fluid after administration of prophylactic human papillomavirus vaccines. J Infect Dis 2009;200:1452–5.

8. Burnett-Hartman AN, Newcomb PA, Mandelson MT, Galloway DA, Madeleine MM, Wurscher M a, et al. No evidence for human papillomavirus in the etiology of colorectal polyps. Cancer Epidemiol Biomarkers Prev 2011;20:2288–97.

9. Muñoz N, Bosch FX, de Sanjosé S, Tafur L, Izarzugaza I, Gili M, et al. The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Colombia and Spain. Int J Cancer 1992;52:743–9.

10. Frisch M, Glimelius B, van den Brule AJ, Wohlfahrt J, Meijer CJ, Walboomers JM, et al. Sexually transmitted infection as a cause of anal cancer. N Engl J Med 1997;337:1350–8.

11. Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 2000;92:709–20.

12. Frey A, Di Canzio J, Zurakowski D. A statistically defined endpoint titer determination method for immunoassays. J Immunol Methods 1998;221:35–41.

13. Lagakos SW. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. Stat Med 1988;7:257–74.

14. Selvin S. Two issues concerning the analysis of grouped data. Eur J Epidemiol 1987;3:284–7.

15. Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. Epidemiology 1992;3:434–40.

16. Breslow NE, Day NE. Statistical methods in cancer research. Vol. 1. Stat. methods cancer Res. Vol. 1. Lyon: International Agency for Research on Cancer 1980.

17. Wartenberg D, Northridge M. Defining exposure in case-control studies: a new approach. Am J Epidemiol 1991;133:1058–71.

18. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology 1995;6:356–65.

19. Burnett-Hartman AN, Newcomb PA, Schwartz SM, Bostick RM, Pawlita M, Waterboer T, et al. No association between antibodies to sexually transmitted infections and colorectal hyperplastic polyps in men: Minnesota Cancer Prevention Research Unit Polyp Study. Cancer Epidemiol Biomarkers Prev 2012;21:1599–601.

20. Muñoz N, Bosch FX, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, et al. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. Int J Cancer 2004;111:278–85.

21. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet 2007;370:890–907.

22. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.

23. Buis ML. POSTRCSPLINE: Stata module containing post-estimation commands for models using a restricted cubic spline. 2009;

24. Dahl FA, Grotle M, Saltyte Benth J, Natvig B. Data splitting as a countermeasure against hypothesis fishing: with a case study of predictors for low back pain. Eur J Epidemiol 2008;23:237–42.

25. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2095–128.

26. Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer Nature Publishing Group 2010;10:878–89.

27. Schulz TF. Cancer and viral infections in immunocompromised individuals. Int J Cancer 2009;125:1755–63.

28. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet 2007;370:59–67.

29. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008;319:1096–100.

30. Babakir-Mina M, Ciccozzi M, Lo Presti A, Greco F, Perno CF, Ciotti M. Identification of Merkel cell polyomavirus in the lower respiratory tract of Italian patients. J Med Virol 2010;82:505–9.

31. Allander T, Andreasson K, Gupta S, Bjerkner A, Bogdanovic G, Persson MAA, et al. Identification of a third human polyomavirus. J Virol 2007;81:4130–6.

32. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog 2007;3:e64.

33. Helmbold P, Lahtz C, Herpel E, Schnabel PA, Dammann RH. Frequent hypermethylation of RASSF1A tumour suppressor gene promoter and presence of Merkel cell polyomavirus in small cell lung cancer. Eur J Cancer Elsevier Ltd 2009;45:2207–11.

34. Joh J, Jenson a B, Moore GD, Rezazedeh A, Slone SP, Ghim S, et al. Human papillomavirus (HPV) and Merkel cell polyomavirus (MCPyV) in non small cell lung cancer. Exp Mol Pathol Elsevier B.V. 2010;89:222–6.

35. Gheit T, Muñoz JP, Levican J, González C, Ampuero S, Parra B, et al. Merkel cell polyomavirus in non-small cell lung carcinomas from Chile. Exp Mol Pathol Elsevier Inc. 2012;93:162–6.

36. Hashida Y, Imajoh M, Nemoto Y, Kamioka M, Taniguchi A, Taguchi T, et al. Detection of Merkel cell polyomavirus with a tumour-specific signature in non-small cell lung cancer. Br J Cancer 2013;108:629–37.

37. Babakir-Mina M, Ciccozzi M, Campitelli L, Aquaro S, Lo Coco A, Perno CF, et al. Identification of the novel KI Polyomavirus in paranasal and lung tissues. J Med Virol 2009;81:558–61.

38. Duncavage EJ, Le B, Wang D, Pfeifer JD. Merkel cell polyomavirus: a specific marker for Merkel cell carcinoma in histologically similar tumors. Am J Surg Pathol 2009;33:1771–7.

39. Teramoto S, Kaiho M, Takano Y, Endo R, Kikuta H, Sawa H, et al. Detection of KI polyomavirus and WU polyomavirus DNA by real-time polymerase chain reaction in nasopharyngeal swabs and in normal lung and lung adenocarcinoma tissues. Microbiol Immunol 2011;55:525–30.

40. Houben R, Grimm J, Willmes C, Weinkam R, Becker JC, Schrama D. Merkel cell carcinoma and Merkel cell polyomavirus: evidence for hit-and-run oncogenesis. J Invest Dermatol 2012;132:254–6.

41. Grulich AE, Jin F, Conway EL, Stein AN, Hocking J. Cancers attributable to human papillomavirus infection. Sex Health 2010;7:244–52.

42. Yuan H, Myers S, Wang J, Zhou D, Woo JA, Kallakury B, et al. Use of Reprogrammed Cells to Identify Therapy for Respiratory Papillomatosis. N Engl J Med 2012;367:1220–7.

43. DiLorenzo TP, Tamsen A, Abramson AL, Steinberg BM. Human papillomavirus type 6a DNA in the lung carcinoma of a patient with recurrent laryngeal papillomatosis is characterized by a partial duplication. J Gen Virol 1992;73 (Pt 2):423–8.

44. Srinivasan M, Taioli E, Ragin CC. Human papillomavirus type 16 and 18 in primary lung cancers--a meta-analysis. Carcinogenesis 2009;30:1722–8.

45. Klein F, Amin Kotb WFM, Petersen I. Incidence of human papilloma virus in lung cancer. Lung Cancer 2009;65:13–8.

46. Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, et al. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. J Natl Cancer Inst 1996;88:1550–9.

47. Omenn GS, Goodman G, Thornquist M, Grizzle J, Rosenstock L, Barnhart S, et al. The beta-carotene and retinol efficacy trial (CARET) for chemoprevention of lung cancer in high risk populations: smokers and asbestos-exposed workers. Cancer Res 1994;54:2038s–2043s.

48. Goodman GE, Thornquist MD, Balmes J, Cullen MR, Meyskens FL, Omenn GS, et al. The Beta-Carotene and Retinol Efficacy Trial: incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. J Natl Cancer Inst 2004;96:1743–50.

49. Waterboer T, Sehr P, Pawlita M. Suppression of non-specific binding in serological Luminex assays. J Immunol Methods 2006;309:200–4.

50. Antonsson A, Green AC, Mallitt K-A, O'Rourke PK, Pawlita M, Waterboer T, et al. Prevalence and stability of antibodies to the BK and JC polyomaviruses: a long-term longitudinal study of Australians. J Gen Virol 2010;91:1849–53.

51. Sehr P, Zumbach K, Pawlita M. A generic capture ELISA for recombinant proteins fused to glutathione S-transferase: validation for HPV serology. J Immunol Methods 2001;253:153–62.

52. Tolstov YL, Pastrana D V, Feng H, Becker JC, Jenkins FJ, Moschos S, et al. Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays. Int J Cancer 2009;125:1250–6.

53. Syrjänen S, Waterboer T, Sarkola M, Michael K, Rintala M, Syrjänen K, et al. Dynamics of human papillomavirus serology in women followed up for 36 months after pregnancy. J Gen Virol 2009;90:1515–26.

54. Simen-Kapeu A, Surcel H-M, Koskela P, Pukkala E, Lehtinen M. Lack of association between human papillomavirus type 16 and 18 infections and female lung cancer. Cancer Epidemiol Biomarkers Prev 2010;19:1879–81.

55. Anantharaman D, Gheit T, Waterboer T, Halec G, Carreira C, Abedi-Ardekani B, et al. No causal association identified for human papillomavirus infections in lung cancer. Cancer Res 2014;

56. Koshiol J, Rotunno M, Gillison ML, Van Doorn L-J, Chaturvedi AK, Tarantini L, et al. Assessment of human papillomavirus in lung tumor tissue. J Natl Cancer Inst 2011;103:501–7.

57. Syrjänen K. Detection of human papillomavirus in lung cancer: systematic review and meta-analysis. Anticancer Res 2012;32:3235–50.

58. Zhang X, Zhu Y, Li L. [Point mutation of p53 and detection of human papillomavirus DNA in bronchogenic carcinoma]. Zhonghua Nei Ke Za Zhi 1995;34:673–5.

59. Da J, Chen L, Hu Y. [Human papillomavirus infection and p53 gene mutation in primary lung cancer]. Zhonghua Zhong Liu Za Zhi 1996;18:27–9.

60. Szabó I, Sepp R, Nakamoto K, Maeda M, Sakamoto H, Uda H. Human papillomavirus not found in squamous and large cell lung carcinomas by polymerase chain reaction. Cancer 1994;73:2740–4.

61. Tsuhako K, Nakazato I, Hirayasu T, Sunakawa H, Iwamasa T. Human papillomavirus DNA in adenosquamous carcinoma of the lung. J Clin Pathol 1998;51:741–9.

62. Mogha A, Fautrel A, Mouchet N, Guo N, Corre S, Adamski H, et al. Merkel cell polyomavirus small T antigen mRNA level is increased following in vivo UV-radiation. PLoS One 2010;5:e11423.

63. Bohlmeyer T, Le TN, Shroyer AL, Markham N, Shroyer KR. Detection of human papillomavirus in squamous cell carcinomas of the lung by polymerase chain reaction. Am J Respir Cell Mol Biol 1998;18:265–9.

64. Smith EM, Rubenstein LM, Haugen TH, Pawlita M, Turek LP. Complex etiology underlies risk and survival in head and neck cancer human papillomavirus, tobacco, and alcohol: a case for multifactor disease. J Oncol 2012;2012:571862.

VITA

Danny V. Colombara was born in New York City and currently resides in Seattle, WA. He earned a Bachelor of Science degree in Biology from the Massachusetts Institute of Technology and a Master of Divinity degree from Gordon-Conwell Theological Seminary. At the University of Washington, he earned a Master of Public Health degree in the Epidemiology Global Health Track in 2010 and a Doctor of Philosophy in Epidemiology in 2014.