

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Geospatial patterns of human papillomavirus vaccine uptake in Minnesota
AUTHORS	Nelson, Erik; Hughes, John; Oakes, J.; Pankow, James; Kulasingam, Shalini

VERSION 1 - REVIEW

REVIEWER	Bednarczyk, Robert A. Emory University, Hubert Department of Global Health
REVIEW RETURNED	15-Jun-2015

GENERAL COMMENTS	This is a very interesting paper and one that aims to answer a key question regarding HPV vaccine uptake, namely how smaller-level geography impacts HPV vaccine uptake. However, there is one major issue with the paper that needs to be addressed, along with some smaller outstanding questions. These are detailed below. First, the authors describe that the "study aimed to determine geographic variation in vaccine uptake using ZIP code level data". However, the analysis presented essentially uses these spatial correlations as an adjustment factor, and doesn't actually present geographic-specific vaccine uptake levels and variation across these geographies. While I believe that this study does add to the literature in showing how traditional analysis, not account for these smaller scale geographic differences, can produce overly narrow confidence intervals around the point estimates. Throughout the manuscript I expected to see differences presented by geography, including potential differences in access to, or geographic spread of health care providers, which was never presented. This portion of the abstract and methods needs to be clarified to describe exactly what was being evaluated. Without evaluating where providers are located, it is not possible to evaluate how distance from a provider can impact access to vaccines. Additionally, there are some other issues that need to be addressed: 1. Introduction, first paragraph - you mention HPV vaccine uptake data as of 2012, but there has been more updated adolescent coverage published in 2013. 2. Introduction, first paragraph, last sentence - There are a number of different topics addressed here, that are contained in the references [10-14]. It seems that these references can be split up across the sentence, as all references are not consistently relevant over the full sentence. 3. Methods, first paragraph - the age range for this study is later than the recommended age of vaccination of 11-12 years. However, there does not seem to be an assessment of age at vaccination, and how that could have impacted vaccine uptake at earlier ages (did
-------------------------	---

	individuals move from where they lived when they were vaccinated?). Were these data collected? If so, please include them in the analysis; if not, please address this in the limitations. 4. Results - Given the differences in vaccine uptake by gender, was there any attempt to stratify the results by gender? Admittedly, this will probably be underpowered, but it would be interesting to see if there were gender differences in vaccine related to geography (which may be related to access to care, location of primary care vs. gynecologic care, etc.). 5. Results - was there any assessment of poverty status or income of participants? As documented in our HPV vaccine disparities manuscript that you referenced (reference # 18), there are differences, at least on a national level, in vaccine uptake by poverty status within racial/ethnic categories. 6. Table 1 - The orientation of the table data was difficult to follow. For the Vaccinated and Not Vaccinated columns, it appears that the proportions are read across rows to total to 100%, whereas the Total proportions are read down the column. 7. Discussion, paragraph 1 - You state that younger age was associated with HPV vaccination, but this does not account for age at vaccination, but age at survey. 8. Discussion, paragraph 2 - Starting with "Previous studies that have attempted...", this portion of the paragraph should include references for context. 9. Discussion, paragraph 3 - When discussing HPV vaccine policies, cost, etc., you do not address poverty as a potential issue in vaccine uptake. 10. Discussion, paragraph 5 - In the limitations, you describe how within-zip code differences may be masked. However, this doesn't account for vaccine practices relative to zip codes. For example, if individuals living in one zip code have providers that practice in multiple different zip codes across a variety of geographic spread, there may be access issues that are not addressed by not accounting for provider location.
--	---

REVIEWER	Broughton, Edward University Research Co., LLC, Research and Evaluation
REVIEW RETURNED	25-Jun-2015

GENERAL COMMENTS	This is an interesting manuscript but I disagree with the way the authors have presented the findings. I don't see any biologic (or other) plausibility for the association between where a person resides and their uptake of HPV. I think it's clear that it is a good proxy for a combination of factors that most likely are associated with HVP vax, notably income, health provider, distance to clinic, school of attendance if any, political view (accounted for in the model) etc. The way the authors state it makes it seem like it is actually where people live that is a causal factor in HPV vax uptake. I think it should be stressed in the paper that post code is a very good proxy for unmeasured factors. They do state "...availability, costs, financial assistance, and availability of education materials to promote uptake collectively contribute to this variability..." but it seems they are suggesting geography determines vax uptake. This is a key clarification that needs to be spelled out. Other points: Selection bias should be noted in the limitations The analysis section in methods can be made more clear
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Specific Comments – Reviewer #1

First, the authors describe that the "study aimed to determine geographic variation in vaccine uptake using ZIP code level data". However, the analysis presented essentially uses these spatial correlations as an adjustment factor, and doesn't actually present geographic-specific vaccine uptake levels and variation across these geographies. While I believe that this study does add to the literature in showing how traditional analysis, not account for these smaller scale geographic differences, can produce overly narrow confidence intervals around the point estimates. Throughout the manuscript I expected to see differences presented by geography, including potential differences in access to, or geographic spread of health care providers, which was never presented. This portion of the abstract and methods needs to be clarified to describe exactly what was being evaluated. Without evaluating where providers are located, it is not possible to evaluate how distance from a provider can impact access to vaccines.

We appreciate this comment. In order to show differences presented by geography, we have included a map of the posterior means of the spatial analysis. This map shows how the distribution of HPV vaccine uptake varies across ZIP codes and will help readers visualize the geographic distribution of HPV vaccination in Minneapolis, Minnesota.

The SMASH study was one of the first health studies to recruit participants via Facebook. When the study was launched it was unclear if people would participate and provide accurate information. Since this recruitment methodology was in its infancy and was untested at that time, we shortened our survey and did not collect data on insurance status, provider information, income, and several other factors that we wanted to collect. Therefore we do not have this information at the individual level. We considered building a multilevel model where we could adjust for factors at the ZIP code level (e.g., insurance, number of providers, STD rates, and poverty indicators), however we purposefully wanted to explore variation in personal level factors. We agree that adjusting for access to providers, poverty and other factors would be very beneficial, however these data are not available. However, the study data utilized here are the only data that have been collected to represent a generalizeable population with a focus on HPV-related information at such a granular level.

1. Introduction, first paragraph - you mention HPV vaccine uptake data as of 2012, but there has been more updated adolescent coverage published in 2013.

Thank you for bringing this to our attention.. We have updated the estimates to reflect the most recently published data (2013).

2. Introduction, first paragraph, last sentence - There are a number of different topics addressed here, that are contained in the references [10-14]. It seems that these references can be split up across the sentence, as all references are not consistently relevant over the full sentence.

We agree, and have split up references as suggested.

3. Methods, first paragraph - the age range for this study is later than the recommended age of vaccination of 11-12 years. However, there does not seem to be an assessment of age at vaccination, and how that could have impacted vaccine uptake at earlier ages (did individuals move from where they lived when they were vaccinated?). Were these data collected? If so, please include them in the analysis; if not, please address this in the limitations.

Unfortunately, age at vaccination was not collected in this data set. We have addressed this in the limitations section of the paper.

4. Results - Given the differences in vaccine uptake by gender, was there any attempt to stratify the

results by gender? Admittedly, this will probably be underpowered, but it would be interesting to see if there were gender differences in vaccine related to geography (which may be related to access to care, location of primary care vs. gynecologic care, etc.).

We attempted to do this when we were originally analyzing the data, however the analysis was terribly underpowered (as you speculated). These type of analyses would be very interesting though.

5. Results - was there any assessment of poverty status or income of participants? As documented in our HPV vaccine disparities manuscript that you referenced (reference # 18), there are differences, at least on a national level, in vaccine uptake by poverty status within racial/ethnic categories.

As was mentioned above, we did not collect this information due to the novelty of the recruitment and data collection methods. However, it has been shown by Paciorek (2010) that controlling for spatial variation reduces bias from some unmeasured confounders. In the limitations paragraph, we make mention of our inability to adjust for income and also cite Paciorek's work to support our claim that the spatial CAR model likely accounts for at least some of the association, albeit indirectly.

6. Table 1 - The orientation of the table data was difficult to follow. For the Vaccinated and Not Vaccinated columns, it appears that the proportions are read across rows to total to 100%, whereas the Total proportions are read down the column.

We agree. This was a grievous error on our part. We have calculated the percentages to be read down the column throughout the table.

7. Discussion, paragraph 1 - You state that younger age was associated with HPV vaccination, but this does not account for age at vaccination, but age at survey.

We have added language to the limitations paragraph that explicitly states this limitation.

8. Discussion, paragraph 2 - Starting with "Previous studies that have attempted...", this portion of the paragraph should include references for context.

These references were cited earlier in the manuscript, but we have added these references to this particular sentence. We agree that this is more a transparent and direct citation and sincerely appreciate this suggestion.

9. Discussion, paragraph 3 - When discussing HPV vaccine policies, cost, etc., you do not address poverty as a potential issue in vaccine uptake.

We have addressed this comment by mentioning poverty explicitly in the introduction (2nd paragraph, last sentence), and in the limitations paragraph of the manuscript. As mentioned above, we could have used a multilevel modeling approach, however we wanted to emphasize solely the individual level data in this analysis.

10. Discussion, paragraph 5 - In the limitations, you describe how within-zip code differences may be masked. However, this doesn't account for vaccine practices relative to zip codes. For example, if individuals living in one zip code have providers that practice in multiple different zip codes across a variety of geographic spread, there may be access issues that are not addressed by not accounting for provider location.

The issue that this statement addresses is the "modifiable areal unit problem" or MAUP. The MAUP is a bias that can occur when aggregating individual observations to an areal unit. It is akin to the ecologic fallacy. In other words, our model assumes that the spatial dependence is the same for each individual within a given ZIP code and should not be inferred for each individual. It is also true that we do not account for access to providers, as we do not have information as to where each individual sought and obtained healthcare. We agree that future analysis should consider provider locations and access to care. However, that information could not be incorporated here.

Specific Comments – Reviewer #2

This is an interesting manuscript but I disagree with the way the authors have presented the findings.

I don't see any biologic (or other) plausibility for the association between where a person resides and their uptake of HPV. I think it's clear that it is a good proxy for a combination of factors that most likely are associated with HVP vax, notably income, health provider, distance to clinic, school of attendance if any, political view (accounted for in the model) etc. The way the authors state it makes it seem like it is actually where people live that is a causal factor in HPV vax uptake. I think it should be stressed in the paper that post code is a very good proxy for unmeasured factors. They do state "...availability, costs, financial assistance, and availability of education materials to promote uptake collectively contribute to this variability...." but it seems they are suggesting geography determines vax uptake. This is a key clarification that needs to be spelled out.

We appreciate this comment very much. Our intention with this manuscript is to demonstrate that failing to account for geography will lead to incorrect inference and potentially spurious associations. The spatial models used in this manuscript account for correlation between geographic units because the assumption of independence has been violated. Thus, these models provide more robust estimates for the effect of independent variables on the dependent variable. As the manuscript is currently written, this is the message that is conveyed. The only time we discuss where people live is in the 2nd paragraph of the discussion section where we point out that neighborhood context (or where people live) is more likely to determine their health behaviors than is their biological race. Indeed, we believe that this type of analysis should be used to identify geographic units where targeted interventions should be implemented. We have added a few clarifying statements throughout the text and have added language in the limitations paragraph to address this concern. We have also added a map to show the geographic variability of HPV vaccine uptake.

1. Selection bias should be noted in the limitations

We agree and have added a statement in the limitations.

2. The analysis section in methods can be made more clear

In response to this comment, we have reviewed several (approximately 20) other peer-reviewed articles (from various disciplines) to compare our wording, disclosure/description of methods, and the opportunity for reproducing the model. In all cases, we have met or exceeded what has been traditionally reported and believe that we have fully explained our methodology, as well as the interpretation of each component in our spatial model. We refer you to the following articles as a reference point.

1. Kazembe, L.N. and J.J. Namangale, A Bayesian multinomial model to analyse spatial patterns of childhood co-morbidity in Malawi. Eur J Epidemiol, 2007. 22(8): p. 545-56. PMID: 17565446.
2. Lee, J., M. Abdel-Aty, and X. Jiang, Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid Anal Prev, 2015. 78: p. 146-54. PMID: 25790973.
3. Ji, C., N.B. Kandala, and F.P. Cappuccio, Spatial variation of salt intake in Britain and association with socioeconomic status. BMJ Open, 2013. 3(1). PMID: 23295624.

VERSION 2 – REVIEW

REVIEWER	Bednarczyk, Robert A. Emory University, Hubert Department of Global Health
REVIEW RETURNED	06-Aug-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.
-------------------------	--

REVIEWER	Broughton, Edward University Research Co., LLC, Research and Evaluation
REVIEW RETURNED	03-Aug-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.
------------------	--