The Lifestyle Factors of Physical Activity and Diet Balance associated with HPV Infection

Yantao Li\(^{1}\), Mengping Liu\(^{2}\), Peng Huang\(^{1}\), Zhongzhou Yang\(^{3}\&\), Anli Wang\(^{1}\&\)

1. College of Sports Medicine and Rehabilitation, Faculty of Sport Rehabilitation, Beijing Sport University, Beijing, China
2. School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China
3. Key Laboratory of Marine Drugs, School of Medicine and Pharmacy, Ocean University of China, 5 Yushan Road, Qingdao 266003, China

& Corresponding author
E-mail: jacriyang5872@hotmail.com

¶These authors contributed equally to this work.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background:

Human wellbeing has been linked with lifestyle factors such as physical activity, diet balance, sleep quality, depression, and anxiety. However, few studies illustrate the relationship between such lifestyle factors and HPV infection. In this study, we demonstrate that lifestyle factors might be crucial for reducing the burden of cervical cancer or HPV-related cancer.

Participants and Methods:

Participants were recruited through a digital eHealth platform without vaccination from May 2020 to August 2021. Both lifestyle factors and cervicovaginal mucus (CVM) samples to test for HPV outcomes were collected from each participant. In addition, the eHealth platform recorded age and gynecological diseases, which were adjusted to apply for both univariable and multivariable logistic regression. Furthermore, lifestyle factors and HPV serotype were categorized as low, intermediate, and high risk in order to conduct stratification analysis. Finally, lifestyle factors were studied in connection with sole and multiple HPV serotypes.

Results:

We recruited 149 HPV positive and 346 HPV negative through HPV detection. Physical activity and diet balance were significantly associated with HPV infection in lifestyle factors (P-values < 0.001) after adjusting for age and gynecological diseases. However, stratified analysis showed three factors were insignificant for HPV infection – namely, sleep quality, depression, and anxiety. Most HPV infections involved a sole HPV serotype (83%), and diet balance was the most significant difference between sole and multiple HPV infections.
Conclusions:

Among lifestyle factors, low physical activity or low diet balance can significantly increase HPV infection. In particular, diet balance might be related to the number of HPV serotypes. Our results suggest that exercising and regulating one’s diet may reduce the burden of HPV-related cancer.

Keywords:

Diet balance, physical activity, eHealth Platform, HPV infection
Introduction

Lifestyle behavior factors such as physical activity, diet balance, sleep quality, depression and anxiety impact personal wellness [1]. The current prevalence of physical inactivity is estimated to be 23% [2, 3]. Hence, the World Health Organization has recommended a 10% increase in physical activity by 2020 [Refs].

Though epidemiological research has shown that a diet balance is associated with decreasing cancer mortality [4, 5], related randomized controlled trials (RCT) for diet balance have not unveiled any demonstrative associations [1]. This obvious paradox has led to increased targeting of diet balance patterns as a risk factor. In addition, three other risk factors that indicate a plausibly precancerous condition are sleep quality and psychosocial variables such as depression and anxiety [6]. However, relatively few studies have focused on connecting these lifestyle behaviors with HPV infection or HPV persistent infections.

HPV infection has been shown to be influenced by age, sexually transmitted diseases, and tumors. Existing studies have explored the link between HPV and four diseases including anemia [7], endocrine disease [8], metabolic disease [9, 10] and mental disorders [11]. They also cover HPV and three diseases in medical history – namely, reproductive tract infection [12, 13], tumors, and consanguinity tumors. Our study elucidates the relationship between lifestyle behavior factors and HPV infection.

A digital eHealth platform tool was developed on May 2020 to record age, lifestyle factors through the standard questionnaires and strategies, and disease status. Physical activity was monitored through the International Physical Activity Questionnaire (IPAQ) and could be beneficial for cancer control. Diet balance, sleep quality, depression and anxiety were recorded by the Diet Balance Index (DBI), Pittsburgh Sleep Quality Index (PSQI), Patient Depression Questionnaire-9 (PDQ-9), and Generalized Anxiety Disorder 7 (GAD-7) respectively. In addition, the digital
eHealth platform provided information on HPV prevention courses via weekly live streaming while regularly providing advice for participants.

Methods

Participant Recruitment and Sample Collection

This study was approved by the ethics committee of the Institutional Review Board at Beijing Sports University (2021173H). From November 2021 to January 2022, study population were registered through the digital eHealth platform. Our team explained the study objectives and obtained signed informed consent from participants to collect questionnaires. Then, participants received a mailed package consisting of a nylon conical brush and graphic user instructions (SeqHPV; Beijing Genomics Institute, China). Participants were required to collect cervicovaginal mucus (CVM) samples by rotating brush and smear samples onto a Flinders Technology Associates (FTA) card. Participants were provided with a package and free express delivery to send their FTA cards to the laboratory. Personalized online guidance was offered if participants had any queries about sampling or delivery.

Outcome Measures

Once the samples were returned to the laboratory, each FTA card was suspended in 5ml PreservCyt media in order for each self-sample to be sequenced for HPV (seqHPV) testing. These tests independently reported 16 serotypes in total, with 12 high-risk HPV serotypes (hrHPV) as HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 [14]; 2 intermediate risks as HPV 66 and 68 [15]; and 2 low risk HPV (lrHPV) as HPV 6 and 11 [16]. Participants received a test report online. Doctors also provided further advice to resolve any queries they had about their test.
Each lifestyle behavior had three categories of low, middle, and high risk, except for diet balance, which had an additional fourth category of extremely high risk. Physical activity was scored in three groups as low-risk (high physical activity), middle-risk (moderate physical activity), and high-risk (low physical activity) from IPAQ in Appendix 1 [17]. Diet balance was scored for Low-Bound Score (LBS) and High-Bound Score (HBS). LBS had four risk groups as low-risk -14 ~ -1, middle-risk -29 ~ -14, high-risk -43 ~ -29, and extremely high-risk -72 ~ -43. HBS also had four risk groups as low-risk 1 ~ 9, middle-risk 10 ~ 18, high-risk 19 ~ 27, and extremely high-risk 27 ~ 44 (DBI in Appendix 2). Sleep quality was scored in three groups as low-risk 0 ~ 7, middle risk 7 ~ 14, and high risk 14 ~ 21 (PSQI in Appendix 3).

Depression was scored in three groups as low-risk 0 ~ 9, middle risk 9 ~ 19, and high risk 19 ~ 27 (PHQ-9, 0, 9, 19, 27 in in Appendix 4). Anxiety was scored in three groups as low-risk 0 ~ 9, middle risk 9 ~ 14, and high risk 14 ~ 21 (GAD-7 in Appendix 5).

**Statistical Analysis**

Baseline demographics were reported for two groups – one HPV-positive and the other HPV-negative. For the two groups, we summarized the continuous variables of age and lifestyle scores with both mean and standard deviation. Disease statuses were summarized as categorical variables with numbers and percentages. Differences were compared using the Wilcoxon rank sum test for continuous variables [18] and the Pearson $\chi^2$ test for categorical variables [19]. In addition, the statistical power was also calculated for lifestyle behavior factors.

In order to study the association between lifestyle behavior factors and HPV outcomes, we applied both univariable and multivariable logistic regression [20]. Regression was adjusted for age and disease status. We conducted further stratified
analysis for different risk groups among lifestyle behavior factors. The reference group was low-risk for each lifestyle behavior factor.

To compare and assess the effect of HPV serotype, we differentiated the three risk groups for HPV infection as hrHPV (high-risk HPV), irHPV (intermediate-risk HPV) and lrHPV (low-risk HPV). We also enumerated the number of serotypes 1 ~ 4 from participants. For one serotype (sole) and 2 ~ 4 serotypes (multiple), physical activity and diet balance scores were tested by the Wilcoxon sum rank. All analysis was conducted utilizing R software (4.0.4) for the Macintosh Operating System (Mac OS).
Results

Participant Recruitment

In total, 495 participants were recruited via our digital eHealth tool with 149 HPV positive and 346 HPV negative women (Table 1). Patient features for HPV infection are elaborated in Table 1. The average age was 41.37 (SD: 7.86) for HPV positive and the average age for HPV negative was 41.56 (SD: 8.51), indicating no significant difference. Disease status was not statistically related to HPV infection. 28% of HPV positive participants and 24% of HPV negative participants had a history of reproductive tract infection. Among lifestyle behavior scores, there were statistically significant results for physical activity, diet balance, sleep quality, and depression. For physical activities (14.36 vs 12.83) and diet balance (25.44 vs 17.29), the higher scores indicate higher risk of HPV infection. They have a statistical power of 0.998 and over 0.999 respectively with \( P\)-values < 0.01. Conversely, another three lower scores denote higher risk on sleep scores (4.38 vs 4.94), depression (1.92 vs 2.27) and anxiety (2.75 vs 2.92). However, their statistical power reduced to 0.593 and 0.330 with \( P\)-values = 0.01.

Logistic Regression for Risk Factors

To evaluate the association of risk factors for HPV infection, both univariate and multivariate logistic regression were performed for one demography factor, seven disease status factors, and five lifestyle behavior factors (Table 2). For the left univariate logistic regression, physical activity significantly increased risk by 1.03 (95% CI: 1.02-1.05) and diet balance by 1.02 (95% CI: 1.01-1.02). For the subsequent right multivariate logistic regression, the increase was 1.20 (95% CI: 1.10-1.33) for physical activity and 1.09 (95% CI: 1.06-1.11) for diet balance, after considering the co-relationship with lifestyle behavior.
Lifestyle Behavior Factors

As expected, patients with high risk in physical activities were more likely to be infected with HPV in comparison to participants with low risk. This association was significant for p-value < 0.01. Patients with middle risk had a higher likelihood of being infected when compared with the low-level risk group (Table 3). Participants with middle, high, and extremely high-level diet imbalance had a higher possibility of being infected compared to participants with a low-risk diet balance. The middle level was significantly different. However, high-level and extremely high-level risk demonstrated no significant association between nutrition and HPV infection (Table 3). Another three lifestyle behavior factors were also tested: sleep quality risk, depression risk, and anxiety risk. However, they were not statistically significant.

Although the probability of HPV infection increased with the risk of sleep quality and depression, the high-level risk of anxiety saw fewer HPV infections (Table 3).

Physical Activity and Nutrition among HPV Serotypes

The prevalence of proportions involved 80.66% hrHPVs, 13.81% intermediate HPVs and 5.52% lrHPVs (Fig 1a). HPV 52 had the largest prevalence (19.89%) from the hrHPVs, followed by HPV16 (11.05%), then HPV 51 (9.39%). HPV 18 prevalence accounted for 4.42%. For intermediate HPV, HPV68 was more common than HPV66. However, the proportion was the same for lrHPVs, HPV 11 and HPV 6. More than 80% of HPV infections involved one serotype (Fig 1b). We then plotted a bar chart of two groups with mean and SD for physical activity (Fig 1c). A similar chart was plotted for diet balance and the significance was denoted as P-value < 0.05 (Fig 1d).

Discussion
Most HPV risk factor studies focus on sexual factors or HIV infection in women. However, there are few cross-sectional studies that account for lifestyle risks or other gynecological infections. In our study, we surveyed 495 participants to investigate the connection between HPV infection and lifestyle behaviors as well as their gynecological disease history or current disease status. Apart from anxiety, four lifestyle factors appeared to demonstrate significant association with HPV infection: physical activity, diet balance, sleep quality, and depression. Meanwhile, current disease or disease history are not significantly correlated with HPV infection. After adjusting for age and disease status, multivariable logistic regression further showed significant association for physical activity and diet balance (P-values < 0.0001). In addition, there were 12 high-risk HPV (hrHPV) serotypes, more than the 2 intermediate-risk and 2 low-risk serotypes.

Our statistical analysis shed light on the most significant lifestyle behaviors. To enhance HPV infection through physical activity, stratification analysis reported the many benefits of low-risk physical activity. Potential reasons may include a decrease in monocyte engagement in the preinvasive tumor microenvironment (TME) causing a depletion in tumor-associated macrophages (TAMs) [21, 22]; or explained by the relation between HPV and human immune system like cell immunity such as CD4/CD8 T cell [23]. In other words, we recommend women infected with HPV to undertake higher levels of physical activity. Earlier studies have suggested two possible approaches for increasing physical activity [17]. The first involves vigorous intensive activity to achieve a minimum overall physical activity of no less than 1500 metabolic equivalents of task (MET) minutes for at least 3 days per week. The second involves achieving double the amount over the course of an entire week, or 3000 MET minutes every seven days, through any combination of vigorous or moderate intensity activities (which can even include walking).
To further enhance HPV infection through diet balance, stratification analysis showed that high-risk diet balance created greater difficulties for clearing HPV infection (OR: 3.27). This may be explained by difficulties in HPV infection being related to low levels of vitamin A or lycopene for high-risk diet balance [24, 25]; or also explained with the human immune system like cell immunity system [26]. Hence, we recommend two solutions for improving diet balance. Through our questionnaires, one way is to ensure appropriate consumption of dairy products and animal foods consisting of vitamin A through fruits (e.g., tomatoes) or vegetables in order to achieve a low-risk diet balance [27]. Beyond explanations from existing studies, the number of HPV serotypes is another potential justification. We found that diet balance scores were significantly increased among those with multiple HPV infections. Thus, the other solution implies that diet balance might be more effective for those with multiple HPV genotypes.

After adjusting for confounders, three lifestyle factors (sleep quality, depression, and anxiety) were shown to be insignificant in relation to HPV infection based on univariable and multivariable logistic regression. Furthermore, stratified analysis showed non-significant middle- and high-risk scores on sleep quality (P-value: 0.52 & 0.80); anxiety (P-value: 0.40 & 0.61); and high-risk depression (P-value: 0.84). Moreover, HPV infection does not appear to be associated with 7 gynecological diseases, whether current or in one’s history, as validated by univariable (P-value: 0.160 ~ 0.652) and multivariable logistic regression (P-value: 0.159 ~ 0.768) based on data collected from our eHealth platform.

Beyond lifestyle behaviors, our study also highlighted the benefits of communication via an eHealth platform. On our eHealth platform, participants actively responded to HPV advice. Through the platform, doctors provided weekly livestream training modules with advice about HPV self-sampling and cervical cancer prevention.
courses. During these sessions, participants were informed about the objective of our study and possible health benefits. In our previous pilot study in May 2020, our team established the eHealth platform and continues to serve HPV high-risk cohorts [28]. This regular training led to a high response rate from 2020 to 2021, as reflected by the growing audience size at our weekly livestream. Connecting with a wider audience has made it possible for us to plan further studies. Further ahead, we plan to enlarge the more widespread clinical variables considering the sexual factors. After that, the HPV infection model will be developed to predict the infection period through survival analysis.

Strengths

In our study, one strength was the age range. The age range of HPV positive participants (41.37 ± 7.86) was comparable with HPV negative participants (41.56 ± 8.51). This age range is also the one in which women experience a high risk for HPV infection. Another strength was how the digital eHealth platform enabled participants to fill in their information electronically via their mobile phone. This substantially decreases the scope of human-made error. A third strength is how this study systematically considers the correlation between daily lifestyle behaviors and HPV infection after adjusting for the impact of age and current diseases or disease history.

Weaknesses

There are three limitations to our study. First, other lifestyle behavior factors might affect common HPV infection risk, such as sexual behavior, smoking, or alcohol consumption. However, our study focused on determining HPV infection factors. Other studies have shown that HPV infection factors are different from infection factors. Second, our study locations consisted of more well-off areas such as Beijing.
and Shenzhen, where there is greater opportunity for abundant physical activity and dietary resources. A more comprehensive study would expand to other locations such as second-line cities or even the countryside in order to reduce geographical bias. To address this issue, our next formal study will focus on areas that encompass a broader scale, from less economically developed to more economically developed.

Conclusions

This study suggests that both physical activity and diet balance are significantly beneficial lifestyle factors to reduce HPV infection. However, the three other factors of sleep quality, depression, and anxiety are required for the further evidence to prove any association. Finally, we recommend exercising regularly and regulating one’s diet balance might reduce the burden of HPV-related cancer such as cervical cancer.

Contributors

Yantao Li and Zhongzhou Yang designed the project and carried out this study. Anli Wang found this project. Zhongzhou Yang and Mengping Liu analyzed the data and prepared the manuscript. Yantao Li collected the dataset and Anli Wang helped to obtain approval from the relevant ethics committees. Yantao Li, Mengping Liu and Anli Wang helped to further improve the quality of the manuscript.

Acknowledgements

We acknowledge Tim Yung for English editing and the BGI team for developing the eHealth platform.
References


Figure Legends

Figure 1. HPV Serotypes in Participants.

(a) Each HPV serotype’s prevalence involved 12 high-risk HPV (hrHPV), 2 intermediate-risk HPV (irHPV), and 2 low-risk HPV (lrHPV). All identifiable serotypes were denoted at the x-axis, and their proportions as percentages were designated at the y-axis among all HPV-infected women. The left red columns indicate the prevalence of each serotype for the 12 hrHPV; the middle pink columns the prevalence of each serotype for the 2 irHPV; the right green columns the prevalence of each serotype for the 2 lrHPV.

(b) This panel shows the number of serotypes identified for each patient infected with HPV.

\( P\text{-value} > 0.05 \)

\( P\text{-value} = 0.02 \)
(c) Physical activity risk in two categories of data with sole and multiple serotypes in a bar chart. This plot displays the scores along with mean and SD without illustrating effect size.

(d) Similar bar chart as (c), but with diet balance risk.

**Figure 2. Transmission Cycle from Male to Female and Known/Unknown Factors Influencing Transmission Dynamics**

HIV, human immunodeficiency virus. *Risk factors for the association direction was unclear before the study in this figure.

**Table Legend:**

**Table 1. HPV Statistics for Age, Disease Status and Lifestyle Behavior**
Table 2. Multivariable Logistic Regression for All Factors

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Univariate logistic regression</th>
<th>Multivariable logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio</td>
<td>95% CI</td>
</tr>
<tr>
<td>Demography</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>1.00 - 1.00</td>
</tr>
<tr>
<td>Disease status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Anemia</td>
<td>1.06</td>
<td>0.97 - 1.20</td>
</tr>
<tr>
<td>History of reproductive tract infection</td>
<td>0.95</td>
<td>0.87 - 1.05</td>
</tr>
<tr>
<td>Current endocrine disease</td>
<td>1.05</td>
<td>0.84 - 1.33</td>
</tr>
<tr>
<td>Current Metabolic disease</td>
<td>0.97</td>
<td>0.82 - 1.14</td>
</tr>
<tr>
<td>History of tumor disease</td>
<td>1.05</td>
<td>0.93 - 1.20</td>
</tr>
<tr>
<td>History of consanguinity tumor</td>
<td>0.90</td>
<td>0.60 - 1.36</td>
</tr>
<tr>
<td>Current Mental disorder</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: Statistical differences; ### denotes p-value < 0.01; ## denotes p-value = 0.01, # denotes 0.01 < p-value < 0.05; NA = Not available. 2: IPAQ = International Physical Activity Questionnaires. DBI = Diet Balance Index, nutrition is negative score. PSQI = Pittsburgh Sleep Quality Index. PHQ-9 = Patient Depression Questionnaire-9. GAD 7 = Generalized Anxiety Disorder 7.
Table 3. Univariable Logistic Regression: Risk of HPV Infection in Low, Middle and High-Level Risk Lifestyle Behavior Factors

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>HPV positive (149)</th>
<th>HPV negative (346)</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>38</td>
<td>1.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Middle</td>
<td>13</td>
<td>74</td>
<td>2.14</td>
<td>0.63-10.22</td>
<td>0.22</td>
</tr>
<tr>
<td>High</td>
<td>133</td>
<td>234</td>
<td>6.85</td>
<td>2.41-29.83</td>
<td>&lt;0.01***</td>
</tr>
<tr>
<td>Diet balance risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>15</td>
<td>163</td>
<td>1.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Middle</td>
<td>129</td>
<td>147</td>
<td>0.11</td>
<td>0.06-0.18</td>
<td>&lt;0.01***</td>
</tr>
<tr>
<td>High</td>
<td>4</td>
<td>28</td>
<td>0.63</td>
<td>0.21-2.41</td>
<td>0.46</td>
</tr>
<tr>
<td>Extremely high</td>
<td>1</td>
<td>8</td>
<td>0.66</td>
<td>0.11-17.52</td>
<td>0.78</td>
</tr>
<tr>
<td>Sleep quality risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>126</td>
<td>284</td>
<td>1.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Middle</td>
<td>22</td>
<td>59</td>
<td>1.18</td>
<td>0.70-2.06</td>
<td>0.52</td>
</tr>
<tr>
<td>High</td>
<td>1</td>
<td>3</td>
<td>1.22</td>
<td>0.14-35.31</td>
<td>0.80</td>
</tr>
<tr>
<td>Depression risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>145</td>
<td>338</td>
<td>1.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Middle</td>
<td>3</td>
<td>8</td>
<td>1.11</td>
<td>0.31-5.39</td>
<td>0.84</td>
</tr>
<tr>
<td>High</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Anxiety risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>134</td>
<td>311</td>
<td>1.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Middle</td>
<td>3</td>
<td>12</td>
<td>1.66</td>
<td>0.51-7.72</td>
<td>0.40</td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>23</td>
<td>0.82</td>
<td>0.40-1.76</td>
<td>0.61</td>
</tr>
</tbody>
</table>

IPAQ = international physical activity questionnaires. DBI = diet balance index. PSQI = Pittsburgh sleep quality index. PHQ-9 = Patient Depression Questionnaire-9. GAD 7 = Generalized Anxiety Disorder 7. NA = Not available

Legend: CI, Confidence Interval; NA, Not available; statistically significant difference is marked *** for P < 0.01.